Revista Matemática Iberoamericana
Vol. 12, N.º 1, 1996

Wavelets obtained by continuous deformations of the Haar wavelet

Aline Bonami, Sylvain Durand and Guido Weiss

Abstract. One might obtain the impression, from the wavelet literature, that the class of orthogonal wavelets is divided into subclasses, like compactly supported ones on one side, band-limited ones on the other side. The main purpose of this work is to show that, in fact, the class of low-pass filters associated with "reasonable" (in the localization sense, not necessarily in the smooth sense) wavelets can be considered to be an infinite dimensional manifold that is arcwise connected. In particular, we show that any such wavelet can be connected in this way to the Haar wavelet.

0. Introduction.

The aim of this paper is to show that, in some sense, any "localized", or of "polynomial decrease" (see below) wavelet may be obtained by a continuous deformation from the Haar function. The case of compactly supported wavelets is due to P. G. Lemarié-Rieusset and G. Malgouyres [6]. More precisely, we shall consider those wavelets which are obtained from a multiresolution analysis (MRA). Let us recall that an MRA is given by an increasing sequence $\{V_j\}_{j\in\mathbb{Z}}$ of closed subspaces of $L^2(\mathbb{R})$, whose union is dense in $L^2(\mathbb{R})$. The space V_{j+1} is obtained from V_j by a dilation by 2; that is, $f \in V_{j+1}$ if and only if $f(2^{-1}x) \in V_j$. One also assumes that there exists a function φ such that $\{\varphi(x-k) : k \in \mathbb{Z}\}$ is an orthonormal basis of V_0. This function φ is generaly called a *scaling function* or a *father wavelet.*

From $V_0 \subset V_1$, we have

$$\varphi(x) = 2\sum_{k\in\mathbb{Z}} c_k\,\varphi(2x-k)\,. \tag{1}$$

That is to say, in terms of the Fourier transform,

$$\hat{\varphi}(\xi) = m_0\big(\frac{\xi}{2}\big)\,\hat{\varphi}\big(\frac{\xi}{2}\big)\,, \tag{2}$$

with

$$m_0(\xi) = \sum_{k\in\mathbb{Z}} c_k\, e^{-ik\xi}\,. \tag{3}$$

This 2π-periodic function m_0 is called the *low-pass filter* associated with this MRA and satisfies the basic properties $m_0(0) = 1$ and

$$|m_0(\xi)|^2 + |m_0(\xi+\pi)|^2 = 1\,. \tag{4}$$

It is then easy to see that one can construct a 2π-periodic function m_1 such that

$$U(\xi) = \begin{pmatrix} m_0(\xi) & m_1(\xi) \\ m_0(\xi+\pi) & m_1(\xi+\pi) \end{pmatrix} \tag{5}$$

is an unitary matrix. The choice of m_1 is closely related to the construction of an orthonormal (or mother) wavelet: one can define $\psi \in L^2(\mathbb{R})$ by letting

$$\hat{\psi}(\xi) = m_1\big(\frac{\xi}{2}\big)\,\hat{\varphi}\big(\frac{\xi}{2}\big)\,, \tag{6}$$

in such a way that $\psi_{jk}(x) = 2^{j/2}\psi(2^j x + k)$ is an orthonormal basis of $L^2(\mathbb{R})$.

Let us remark that the choice of m_1 is not unique. The fact that $U(\xi)$ is unitary implies that any other $\tilde{m}_1$ is given by $\tilde{m}_1(\xi) = a(\xi)\,m_1(\xi)$, where a is π-periodic with values of modulus 1. For example, we can take $m_1(\xi) = e^{i\xi}\,\bar{m}_0(\xi+\pi)$.

The construction of an MRA, and the associated orthonormal wavelet basis, can also be done in terms of m_0 (if it satisfies appropriate properties). We define φ from m_0 by

$$\hat{\varphi}(\xi) = \prod_{j=1}^{\infty} m_0(2^{-j}\xi). \tag{7}$$

I. Daubechies used this equality to construct her compactly supported wavelets. A characterization of the filters m_0 which generate an MRA has been obtained by A. Cohen in [1] (see Theorem 1.2 below). The basic question is: what property, in addition to (4), must m_0 satisfy to give us an MRA.

In order to understand such characterizations, we make the following remarks. Perhaps the "simplest" low-pass filter is the one associated with the Haar system: $m_0(\xi) = (1 + e^{i\xi})/2$. Clearly, $m_0(0) = 1$ and (4) is satisfied. Another simple function satisfying these properties is $m_0(\xi) = (1 + e^{i3\xi})/2$; but a simple calculation shows that (7), in this case, gives us, $\varphi = (1/3)\chi_{[-3,0]}$ for which $\{\varphi(x-k) : k \in \mathbb{Z}\}$ is not an orthonormal system. It is known, for example that if $m_0(\xi)$ is never 0 in $[-\pi/3, \pi/3]$, then (3) does give us a scaling function that generates a localized MRA. A. Cohen, in his thesis, shows that this is included in a characterization of these low-pass filter that we announce in condition 2.b) in Theorem 1.2 below. One of our aims is to show that if we rephrase this condition, then the set of functions m_0 may be seen as consisting of a "manifold".

More precisely, in this paper, we show that the set $\mathcal{E}$ of the C^∞ filters m_0, is a "connected manifold" in the Frechet space $C^\infty(\mathbb{T})$ of 2π-periodic functions, defined by the family of semi-norms $\|D^\alpha f\|_\infty$ ($\alpha = 0, 1, \dots$). In particular, we construct a continuous path in $\mathcal{E}$, connecting any element of $\mathcal{E}$ to the Haar filter $(1+e^{i\xi})/2$. This gives us a continuous "deformation" between any mother wavelet ψ with polynomial decay, and the Haar wavelet $h = \chi_{[0,1/2]} - \chi_{[1/2,1]}$. That is to say, we obtain a continuous function $t \mapsto \psi_t$, from $[0,1]$ to $L^2((1+|x|)^n dx)$ for any n, such that ψ_t is a wavelet, $\psi_0 = h$ and $\psi_1 = \psi$.

1. Characterization of the low-pass filter.

We start with the following definitions:

Definition 1.1. *We say that φ has polynomial decay in $L^2(\mathbb{R})$ if $|x|^N\varphi \in L^2(\mathbb{R})$ for all $N \in \mathbb{N}$, and that φ has exponential decay of order λ in $L^2(\mathbb{R})$ if there exists $\lambda > 0$ so that $e^{\lambda|x|}\varphi \in L^2(\mathbb{R})$.*

Our main interest is to study the existence of a scaling function φ, associated with a given low-pass filter m_0, that generates a multiresolution analysis (MRA). We quote a result in [1]:

Theorem 1.2. *Suppose $\varphi \in L^2(\mathbb{R})$ and m_0, 2π-periodic, are related as in equalities (2) and (7). Then the following properties are equivalent:*

1) *The function φ is the scaling function of an* MRA *and has polynomial decay in $L^2(\mathbb{R})$.*

2) *The function m_0 belongs to $C^\infty(\mathbb{T})$ and satisfies*

a) $m_0(0) = 1$ *and* $|m_0(\xi)|^2 + |m_0(\xi+\pi)|^2 = 1$,

b) *There exists a finite union of closed bounded intervals K such that $0 \in K^\circ$, $\sum_{k\in\mathbb{Z}} \chi_K(\xi + 2k\pi) = 1$ almost everywhere and, for all $j \in \mathbb{N}$, $\xi \in K$, we have $m_0(2^{-j}\xi) \neq 0$.*

Using (7), $m_0(0) = 1$ and $m_0 \in C^\infty(\mathbb{T})$, we observe that this last inequality is equivalent to $\hat{\varphi}(\xi) \neq 0$ for all $\xi \in K$. Thus we can restate this theorem by changing condition b) to

b') *There exists a finite union of closed bounded intervals K such that $0 \in K^\circ$, $\sum_{k\in\mathbb{Z}} \chi_K(\xi + 2k\pi) = 1$ almost everywhere and, for all $\xi \in K$, $\hat{\varphi}(\xi) \neq 0$.*

Moreover, we add two further conclusions:

Theorem 1.3. *With φ and m_0 related as above and satisfying one of the equivalent properties of Theorem 1.2, we have:*

i) *The support of φ is in $[0, N]$ if and only if m_0 is a trigonometric polynomial of degree $\leq N$.*

ii) *The function φ has exponential decay in $L^2(\mathbb{R})$ if and only if m_0, regarded as a function of $e^{i\xi}$ (on the unit circle), extends to a holomorphic function on an annulus (i.e. a region lying between two concentric circles centered at 0, including the unit circle).*

Property i) was proved by I. Daubechies in [2].

Property ii) is implicit in the theory of wavelets. We shall present an argument for completeness. Let us first assume that m_0 extends to a holomorphic function on an annulus. By the Cauchy formula, we have $\|m_0^{(l)}\|_\infty \leq l!\,\tilde{M}^l$ for some constant $\tilde{M}$ depending on the size of the annulus (see [8]). From this we can deduce $\|\hat{\varphi}^{(n)}\|_2 \leq n!\,M^n$. To see this, we write

$$\varphi^{(n)} = \sum_{J \in \mathbb{N}^n} 2^{-J} h_J \,,$$

where $J = (j_1, \ldots, j_n)$, $2^{-J} = 2^{-j_1} \cdots 2^{-j_n}$ and the h_J are obtained by differenting the identity $\hat{\varphi}(\xi) = \prod_{j=1}^{\infty} m_0(\xi/2^j)$ (and replacing $m_0(\xi/2^j)$ by $m_0^{(l)}(\xi/2^j)$ when j occurs l times in the sequence $j_1, \ldots, j_n$). Then, for $0 \leq t < 1/\sqrt{M}$, we have

$$\begin{aligned} C \int \Big(\sum_{n=1}^{\infty} \frac{t^{2n}\,|x|^{2n}}{(2n)!} \Big)\, |\varphi(x)|^2 \, dx &= \int \Big(\sum_{n=1}^{\infty} \frac{t^{2n}}{(2n)!} \Big)\, |\hat{\varphi}^{(n)}(\xi)|^2 \, d\xi \\ &\leq \sum_{n=1}^{\infty} \frac{(n!)^2}{(2n)!}\, t^{2n}\, M^n < \infty \,. \end{aligned}$$

From this it is then easy to check that

$$\begin{aligned} \int e^{t|x|}\, |\varphi(x)|^2 \, dx &= \int \Big(\sum_{k=0}^{\infty} \frac{(t\,|x|)^k}{k!} \Big)\, |\varphi(x)|^2 \, dx \\ &\leq C \Big(\int \Big(\sum_{n=1}^{\infty} \frac{(t\,|x|)^{2n}}{(2n)!} \Big)\, |\varphi(x)|^2 \, dx + \|\varphi\|_2^2 \Big) < \infty \,. \end{aligned}$$

We thus obtain an exponential decay of order λ in $L^2(\mathbb{R})$, for φ, whenever $0 < \lambda < 1/\sqrt{M}$.

On the other hand, when φ has such an exponential decay, we have (from the definition of m_0), for $k \geq 0$

$$\begin{aligned} |e^{\lambda k}\, \hat{m}_0(k)| &= \Big| e^{\lambda k}\, \frac{1}{2} \int \varphi\big(\frac{x}{2}\big)\, \bar{\varphi}(x-k)\, dx \Big| \\ &= \Big| \frac{1}{2} \int e^{-\lambda x}\, \varphi\big(\frac{x}{2}\big)\, e^{\lambda(x+k)}\, \bar{\varphi}(x+k)\, dx \Big| \,, \end{aligned}$$

as well as a similar equality, for $k < 0$. Thus, the sequence $\hat{m}_0(k)$ has exponential decay $|e^{\lambda|k|}\, \hat{m}_0(k)| \leq A < \infty$ since $\int e^{\lambda|x|}\, |\varphi(x)|^2 \, dx < 0$.

This implies that m_0 is the restriction to the unit circle of a holomorphic function on an annulus about the origin.

2. A geometric interpretation.

We shall consider the Frechet space $C^\infty(\mathbb{T})$ of 2π-periodic functions endowed with the topology generated by the semi-norms $\|F^{(n)}\|_\infty$. We also consider the space $C^\infty(\mathbb{T}/2)$ of all π-periodic functions endowed with the same semi-norms. Let us define $\mathcal{E}$ by

$$\mathcal{E} = \{F \in C^\infty(\mathbb{T}) : \ F \text{ satisfies a) and b)}\}\,.$$

We shall show the following.

Theorem 2.1. *The set $\mathcal{E}$ is a Frechet manifold in the sense that each $m_0 \in \mathcal{E}$ has a neighborhood that is homeomorphic to a neighborhood of 0 in $C^\infty(\mathbb{T})$, where here 0 is the constant zero function.*

To prove the theorem we define the set

$$\mathcal{F} = \{F \in C^\infty(\mathbb{T}) : \ F \text{ satisfies a)}\}\,.$$

We shall show that $\mathcal{F}$ is a manifold in this sense and that $\mathcal{E}$ is an open set in $\mathcal{F}$.

Let $m_0 \in \mathcal{E}$, and let $m_1 \in C^\infty(\mathbb{T})$ be such that

$$U(\xi) = \begin{pmatrix} m_0(\xi) & m_1(\xi) \\ m_0(\xi+\pi) & m_1(\xi+\pi) \end{pmatrix}$$

is a unitary matrix. We shall use the elementary lemma:

Lemma 2.2. *Any C^∞ and 2π-periodic function F may be written uniquely*

$$F = G\,m_0 + H\,m_1\ ,$$

where G and H are C^∞ and π-periodic. Moreover $F \mapsto (G,H)$ is an isomorphism between $C^\infty(\mathbb{T})$ and $C^\infty(\mathbb{T}/2) \times C^\infty(\mathbb{T}/2)$.

To prove the lemma, it is sufficient to remark that G and H are given by

$$\begin{pmatrix} G(\xi) \\ H(\xi) \end{pmatrix} = U(\xi)^{-1} \begin{pmatrix} F(\xi) \\ F(\xi+\pi) \end{pmatrix}\,, \tag{8}$$

since U^{-1} has C^∞ coefficients.

Moreover, U being unitary,

$$|F(\xi)|^2 + |F(\xi+\pi)|^2 = |G(\xi)|^2 + |H(\xi)|^2 ,$$

so that condition a) for F becomes

a') $G(0) = 1$, $H(0) = 0$, $|G|^2 + |H|^2 = 1$.

From (4), we see that

$$\begin{pmatrix} G(\xi) - 1 \\ H(\xi) - 0 \end{pmatrix} = U(\xi)^{-1} \begin{pmatrix} F(\xi) - m_0(\xi) \\ F(\xi+\pi) - m_0(\xi+\pi) \end{pmatrix} .$$

Hence, if F is close to m_0 in $C^\infty(\mathbb{T})$, then G is close to 1 and H is close to 0 in the topology we introduced. Then, to show that $\mathcal{F}$ is a manifold, it is sufficient to show that a neighborhood of $(1,0)$ in $\mathcal{F}' = \{(G,H) \in C^\infty(\mathbb{T}/2) \times C^\infty(\mathbb{T}/2) : (G,H)$ satisfies a')$\}$ is homeomorphic to a neighborhood of $(0,0)$ in $C^\infty(\mathbb{T}/2)\times C^\infty(\mathbb{T}/2)$ (that is homeomorphic to a neighborhood of 0 in $C^\infty(\mathbb{T})$, by Lemma 2.2. We can take the neighborhood given by $\|H\|_\infty < 1/2$ and $\|1-G\|_\infty \leq 1/2$. Then, clearly, $G = e^{iA}\sqrt{1-|H|^2}$, where A is C^∞, π-periodic with values in $[-\pi/2, \pi/2]$, $A(0) = 0$, and the application $(F,G) \mapsto (A,H)$ is a continuous bijection. This shows that $\mathcal{F}$ is a manifold in the sense we described above.

Let us prove that $\mathcal{E}$ is an open set in $\mathcal{F}$. We have to prove that if $m_0 \in \mathcal{E}$ then for F close to m_0, the scaling function $\hat\varphi_F$ which corresponds to F satisfies a condition equivalent to b).

We shall use the following lemma that will be proved later.

Lemma 2.3. *If F tends to m_0 in $\mathcal{F}$, then $\hat\varphi_F$ tends to $\hat\varphi$ uniformly on compacts.*

From b'), we have $|\hat\varphi(\xi)| \geq C > 0$ when $\xi \in K$, since the latter is compact and $\hat\varphi$ is continuous. The fact that φ_F tends to φ uniformly on K permits us to conclude that $|\hat\varphi_F(\xi)| \geq C/2$ on the same compact set, for F close enough to m_0. This shows that $\mathcal{E}$ is open in $\mathcal{F}$.

Let us prove the last lemma now. In fact, we are going to prove a more powerful property that we shall use later.

Proposition 2.4. *If F tends to m_0 in $\mathcal{F}$, then $\hat{\varphi}_F$ and its derivatives tend uniformly to $\hat{\varphi}$ and its derivatives on compacts. Moreover if $m_0 \in \mathcal{E}$, the convergence of F towards m_0 in $\mathcal{E}$ is equivalent to the convergence of $x^n\varphi_F$ toward $x^n\varphi$ in $L^2(\mathbb{R})$ for each n.*

A version of this proposition has been obtained, by a different method, by P.G. Lemarié-Rieusset.

First, to prove the uniform convergence on compacts, it is sufficient to prove it on an interval $[-a,a]$ on which $|m_0(\xi) - 1| < 1/2$. We see this since it gives us the convergence on $[-2^N a, 2^N a]$ for each N, by using $\hat{\varphi}_F(\xi) = \prod_{j=1}^N F(\xi/2^j)\,\hat{\varphi}_F(\xi/2^N)$. We can also assume that $\|F - m_0\|_\infty < 1/4$, so that the logarithms in the sequel are well defined and belong to C^∞.

Let us first prove that $\hat{\varphi}_F$ tends to $\hat{\varphi}$ uniformly on $[-a,a]$. We prove this by showing $\hat{\varphi}_F/\hat{\varphi} \to 1$ or $\log(\hat{\varphi}_F/\hat{\varphi}) \to 0$ uniformly on $[-a,a]$; that is, using (7), we show

$$\log\frac{\hat{\varphi}_F(\xi)}{\hat{\varphi}(\xi)} = \sum_{j=1}^{\infty} \log\frac{F(\xi/2^j)}{m_0(\xi/2^j)} \to 0$$

as $F \to m_0$ when $\xi \in [-a,a]$. But, by mean value theorem,

$$\left|\log\frac{F(\xi/2^j)}{m_0(\xi/2^j)}\right| \le \frac{|\xi|}{2^j} \sup_{[-a,a]} \left|\frac{F'}{F} - \frac{m_0'}{m_0}\right|.$$

Hence, we have

$$\left|\sum_{j=1}^{\infty} \log\frac{F(\xi/2^j)}{m_0(\xi/2^j)}\right| \le a \sup_{[-a,a]} \left|\frac{F'}{F} - \frac{m_0'}{m_0}\right|,$$

which tends to zero as $\|F - m_0\|_\infty$ and $\|F' - m_0'\|_\infty$ tend to zero.

Let us now prove that $\hat{\varphi}_F^{(n)} \to \hat{\varphi}^{(n)}$ uniformly on $[-a,a]$, for $n > 0$. It is equivalent to show that $(d/d\xi)^n(\hat{\varphi}_F/\hat{\varphi})$ tends uniformly to 0 on $[-a,a]$ since $\hat{\varphi}$ is bounded away from zero on $[-a,a]$. Thus, we consider

$$\sum_{j=1}^{\infty} \frac{1}{2^{nj}} \left((\log F)^{(n)}\left(\frac{\xi}{2^j}\right) - (\log m_0)^{(n)}\left(\frac{\xi}{2^j}\right)\right);$$

and, it is elementary to deduce that $(\log F)^{(n)} - (\log m_0)^{(n)}$ is uniformly small on $[-a,a]$ when $\|F^{(k)} - m_0^{(k)}\|_\infty$ is small for each $k \ge 0$, since our assumptions imply that the values assumed by F and m_0, and their derivatives, are appropriately restricted.

Let us show that if $m_0 \in \mathcal{E}$ then the convergence of φ_F to φ is in $L^2(\mathbb{R})$ (remember that $\int |\hat{\varphi}_F|^2 = \int |\hat{\varphi}|^2 = 2\pi$). Let K be a compact such that $\int_K |\hat{\varphi}|^2/2\pi > (1-\varepsilon)^2$. Then, if F is close enough to m_0, so that $\int_K |\hat{\varphi}_F - \hat{\varphi}|^2 < 2\pi\varepsilon^2$, then $\int_K |\hat{\varphi}_F|^2 > 2\pi(1-2\,\varepsilon)^2$, and $\int_{CK} |\hat{\varphi}_F|^2 < 2\pi(2\,\varepsilon)^2$. Finally, $\int |\hat{\varphi}_F - \hat{\varphi}|^2 < 2\pi(4\,\varepsilon)^2$.

Let us now prove the convergence in L^2 of the derivatives. We claim that if $m_0 \in \mathcal{F}$, then $\hat{\varphi}_F^{(n)} \to \hat{\varphi}^{(n)}$ weakly (in $L^2(\mathbb{R})$) for any n. Observe that it follows from the proof of Theorem 1.3.ii) that $\|\hat{\varphi}_F^{(n)}\|_2$ is bounded uniformly when F lies in a neighborhood of m_0. Hence, there exists a subsequence that converges weakly. Clearly the only possible limit is $\hat{\varphi}^{(n)}$. Since any sequence of F's converging to m_0 has a subsequence such that $\hat{\varphi}_F^{(n)}$ converge weakly to $\hat{\varphi}^{(n)}$, our claim follows.

As $\hat{\varphi}_F^{(2n)} \to \hat{\varphi}^{(2n)}$ weakly and $\hat{\varphi}_F \to \hat{\varphi}$ in $L^2(\mathbb{R})$, we have

$$\int \hat{\varphi}_F^{(2n)}\, \bar{\hat{\varphi}}_F \to \int \hat{\varphi}^{(2n)}\, \bar{\hat{\varphi}}\,.$$

Thus,

$$\int x^{2n}\, |\varphi_F(x)|^2\, dx \to \int x^{2n}\, |\varphi(x)|^2\, dx\,,$$

and

$$\int |\hat{\varphi}_F^{(n)}|^2 \to \int |\hat{\varphi}^{(n)}|^2\,.$$

Then let J be a compact set such that

$$\int_J |\hat{\varphi}^{(n)}|^2 > (1-\varepsilon)^2\, \|\hat{\varphi}^{(n)}\|^2\,.$$

If F is close enough to m_0 so that

$$\int_J |\hat{\varphi}_F^{(n)} - \hat{\varphi}^{(n)}|^2 < \varepsilon^2\, \|\hat{\varphi}^{(n)}\|_2$$

(since $\hat{\varphi}_F^{(n)}$ converges uniformly to $\hat{\varphi}^{(n)}$ on the compact set J) and

$$\big|\|\hat{\varphi}_F^{(n)}\|^2 - \|\hat{\varphi}^{(n)}\|^2\big| < \varepsilon^2 \|\hat{\varphi}^{(n)}\|^2\,,$$

then

$$\int_J |\hat{\varphi}_F^{(n)}|^2 \geq (1-2\,\varepsilon)^2 \|\hat{\varphi}^{(n)}\|^2\,, \qquad \int_{J^c} |\hat{\varphi}_F^{(n)}|^2 \leq (3\,\varepsilon)^2\, \|\hat{\varphi}^{(n)}\|^2\,,$$

and

$$\int |\hat{\varphi}_F^{(n)} - \hat{\varphi}^{(n)}|^2 \leq (5\,\varepsilon)^2\, \|\hat{\varphi}^{(n)}\|^2 .$$

Conversely, we now show that $m_0 \in \mathcal{E}$ and $x^n \varphi_F \to x^n \varphi$ in $L^2(\mathbb{R})$, for each n, implies $F \to m_0$ in C^∞. We have

$$\hat{m}_0(k) = \frac{1}{2} \int \varphi\big(\frac{x}{2}\big)\, \bar{\varphi}(x+k)\, dx . \tag{9}$$

Since $\hat{\varphi}_F$ tends to $\hat{\varphi}$ uniformly on compacts, we can assume that $\hat{\varphi}_F$ also satisfies condition b'). Thus, by Theorem 1.2, equality (9) is also true for F, and we have

$$\begin{aligned} |k|^n(\hat{m}_0(k) - \hat{F}(k)) &= \frac{|k|^n}{2} \int \Big(\varphi\big(\frac{x}{2}\big)\, \bar{\varphi}(x+k) - \varphi_F\big(\frac{x}{2}\big)\, \bar{\varphi}_F(x+k)\Big)\, dx \\ &= \frac{|k|^n}{2} \int \bar{\varphi}(x+k) \Big(\varphi\big(\frac{x}{2}\big) - \varphi_F\big(\frac{x}{2}\big)\Big)\, dx \\ &\quad + \frac{|k|^n}{2} \int \varphi_F\big(\frac{x}{2}\big)\, \big(\bar{\varphi}(x+k) - \bar{\varphi}_F(x+k)\big)\, dx . \end{aligned}$$

Let us majorize

$$\frac{|k|^n}{2} \int \varphi_F\big(\frac{x}{2}\big)\, \big(\bar{\varphi}(x+k) - \bar{\varphi}_F(x+k)\big)\, dx .$$

For any $x \in \mathbb{R}$,

$$|k|^n \leq 2^n\, (|x|^n + |k+x|^n) .$$

But

$$\int |x|^{2n}\, \big|\varphi_F\big(\frac{x}{2}\big)\big|^2\, dx \leq C < \infty$$

and

$$\int |\varphi(x+k) - \varphi_F(x+k)|^2\, dx \longrightarrow 0 .$$

Thus,

$$\int |x|^n\, \varphi_F\big(\frac{x}{2}\big)\, (\bar{\varphi}(x+k) - \bar{\varphi}_F(x+k))\, dx \longrightarrow 0 .$$

Moreover,

$$\int \big|\varphi_F\big(\frac{x}{2}\big)\big|^2\, dx \leq C < \infty$$

and

$$\int |\varphi(x) - \varphi_F(x)|^2\, |x|^{2n}\, dx \longrightarrow 0\,.$$

Thus,

$$\int |k+x|^n\, \varphi_F\big(\frac{x}{2}\big)\, (\bar\varphi(x+k) - \bar\varphi_F(x+k))\, dx \longrightarrow 0\,.$$

We majorize

$$\frac{|k|^n}{2} \int \bar\varphi(x+k) \Big(\varphi_F\big(\frac{x}{2}\big) - \varphi\big(\frac{x}{2}\big)\Big)\, dx$$

in the same way and we obtain

$$\|\, |k|^n \hat m_0(k) - |k|^n \hat F(k)\|_\infty \longrightarrow 0\,.$$

In particular, if $n \geq 2$, we obtain

$$|\, |k|^{n-2} \hat m_0(k) - |k|^{n-2} \hat F(k)\,| \leq |k|^{-2}\, o(1)\,.$$

Therefore,

$$\|m_0^{(n-2)} - F^{(n-2)}\|_\infty \leq \Big(\sum_{k\in\mathbb{Z}} |k|^{-2}\Big)\, o(1)\,.$$

EXAMPLE. let $\mathcal{F}_N$ be the set of polynomials with real coefficients of degree N that belong to $\mathcal{F}$; that is, $m \in \mathcal{F}_N$ if and only if $m(\xi) = a_0 + a_1 e^{i\xi} + \cdots + a_N e^{iN\xi}$, the a_j's are real and m satisfies a). Let $\mathcal{E}_N = \mathcal{E} \cap \mathcal{F}_N$.

Let us examine the example given in the end of the introduction in these terms. For $N = 3$, $\mathcal{F}_N$ consists of those m satisfying

$$m(\xi) = \frac{1+e^{i\xi}}{2}\, (a + b\, e^{i\xi} + c\, e^{2i\xi})\,,$$

with $b = 1 - a - c$, and $a^2 + c^2 = a + c$. $\mathcal{E}_3$ corresponds to the circle

$$\{(a,c):\ a^2 + c^2 = a + c\} \setminus \{(1,1)\}\,.$$

If $(1,1)$ were a point of this circle, the corresponding filter would be $m(\xi) = (1 + e^{3i\xi})/2$ and the corresponding scaling function would be $\varphi = (1/3)\,\chi_{[-3,0]}$. but the latter has L^2-norm $1/\sqrt{3}$ $(\neq 1)$, thus, as observed before, we would not obtain an MRA.

3. The deformation of wavelets associated with the class $\mathcal{E}$.

Let us introduce a dense subset in $\mathcal{F}$ that will be useful to us: let $\mathcal{F}_{\exp}$ (respectively $\mathcal{E}_{\exp}$) be the set for which $\hat m_0(k)$ has exponential decay. Then we have the following:

Proposition 3.1. *$\mathcal{F}_{\exp}$ is dense in $\mathcal{F}$.*

To prove the proposition, we take m_0 in $\mathcal{F}$ and select a sequence of trigonometric polynomials P_n which tends to m_0 in $C^\infty(\mathbb{T})$.

We can assume that, for all integer n, $P_n(0) = 1$ and $P_n(\pi) = 0$. Indeed, let $\tilde{m}_0(\xi) = ((1 + e^{i\xi})/2)^{-1} m_0(\xi)$. Since $m_0(\pi) = 0$, $\tilde{m}_0$ is well defined and C^∞. Build a sequence of trigonometric polynomials $\tilde{P}_n$ which tends to $\tilde{m}_0$. The sequence $P_n(\xi) = (1+e^{i\xi})(\tilde{P}_n(\xi)+1-\tilde{P}_n(0))/2$ tends to m_0 and satisfies the required properties.

Now $|P_n(\xi)|^2 + |P_n(\xi + \pi)|^2$ tends to 1 in $C^\infty(\mathbb{T})$. So $(|P_n(\xi)|^2 + |P_n(\xi + \pi)|^2)^{-1/2}$ is well defined for n big enough, and tends to 1 in $C^\infty(\mathbb{T})$. Finally we take

$$F_n(\xi) = \frac{P_n(\xi)}{(|P_n(\xi)|^2 + |P_n(\xi + \pi)|^2)^{1/2}} ,$$

which belongs to $\mathcal{F}$, and tends to m_0.

The only thing to prove is that $F_n \in \mathcal{F}_{\exp}$. By the argument that establishes Theorem 1.3, part ii), it suffices to show that $(|P_n(\xi)|^2 + |P_n(\xi+\pi)|^2)^{-1/2}$ extends to a holomorphic function on an annulus. But $|P_n(\xi)|^2 + |P_n(\xi + \pi)|^2$ is a trigonometric polynomial. There exists an integer m such that $e^{im\xi}(|P_n(\xi)|^2 + |P_n(\xi + \pi)|^2)$ is a polynomial in $e^{i\xi}$ which extends to a holomorphic function on $\mathbb{C}$. We can then extend $|P_n(\xi)|^2 + |P_n(\xi + \pi)|^2$ to a holomorphic function f_n on $\mathbb{C} \setminus \{0\}$. And, since there exists a neighborhood of the unit circle on which $\operatorname{Re} f_n(z) > 1/2$ for n big enough, $f_n^{-1/2}$ is also holomorphic on that neighborhood.

Finally, let us prove the following:

Theorem 3.2. *$\mathcal{F}$ and $\mathcal{E}$ are connected.*

We shall prove that any m_0 in $\mathcal{F}$ can be joined to $(1 + e^{i\xi})/2$ by a continuous path, and that this path can be chosen within $\mathcal{E}$ if m_0 is in $\mathcal{E}$.

Let us first remark that, if we go back to the example discussed in Section 2, we can clearly see that there exists such a continuous path between $e^{i\xi}(1 + e^{i\xi})/2$ and $(1 + e^{i\xi})/2$ (just follow the circle indicated). Then, for any integer k, there exists a continuous path between $e^{i(k+1)\xi}(1 + e^{i\xi})/2$ and $e^{ik\xi}(1 + e^{i\xi})/2$, and consequently, between $e^{ik\xi}(1 + e^{i\xi})/2$ and $(1 + e^{i\xi})/2$. Thus, it suffices to join m_0 to $e^{ik\xi}(1 + e^{i\xi})/2$ for an appropriate integer k.

We begin by constructing a path made up of trigonometric polynomials (not necessarily in $\mathcal{F}$).

Lemma 3.3. *Let F be a trigonometric polynomial of degree less than or equal to N such that $F(0) = 1$ and $F(\pi) = 0$. There exists a continuous map $t \mapsto F_t$ from $[0,1]$ to the space of trigonometric polynomials of degree $\leq N$ such that*

1) $F_t(0) = 1,\ F_t(\pi) = 0,$ *if* $0 \leq t \leq 1$,

2) $F_1(\xi) = F(\xi)$,

3) $|F_t(\xi)|^2 = (1-t)\left(\frac{1+e^{i\xi}}{2}\right)\left(\frac{1+e^{-i\xi}}{2}\right) + t\,|F(\xi)|^2 .$

Moreover, $F_0(\xi) = e^{ik\xi}\,(1+e^{i\xi})/2$ *for an integer* k.

We postpone the proof of this lemma and, using it, pass to the proof of Theorem 3.2. In the general case, since $\mathcal{F}$ is a manifold, one can join m_0 to a neighboring m belonging to the dense subset $\mathcal{F}_{\exp} \subset \mathcal{F}$; thus, we can assume that $m \in \mathcal{F}_{\exp}$. If we examine the proof of Proposition 3.1, in fact, we observe that we can assume

$$m(\xi) = \frac{F(\xi)}{(|F(\xi)|^2 + |F(\xi+\pi)|^2)^{1/2}} ,$$

where F is a trigonometric polynomial which satisfies $F(0) = 1$, $F(\pi) = 0$ and $|1 - |F(\xi)|^2 - |F(\xi+\pi)|^2| < 1/2$. We can then apply Lemma 3.3 to F. We obtain a continuous function $t \mapsto F_t$ with $F_0(\xi) = e^{ik\xi}(1+e^{i\xi})/2$ and $F_1(\xi) = m_0(\xi)$.

So let

$$G_t(\xi) = |F_t(\xi)|^2 + |F_t(\xi+\pi)|^2 .$$

We also have $G_t(\xi) = 1 - t + t\,G(\xi)$, where $G(\xi) = |F(\xi)|^2 + |F(\xi+\pi)|^2$.

The path will join $e^{ik\xi}(1+e^{i\xi})/2$ to m_0, via

$$t \mapsto F_t(\xi)\,(G_t(\xi))^{-1/2} = \Phi_t(\xi) .$$

It remains for us to show that $t \mapsto \Phi_t$ is continuous from $[0,1]$ to $C^\infty(\mathbb{T})$, and that, for all t, Φ_t belongs to $\mathcal{F}$, or to $\mathcal{E}$ if $m_0 \in \mathcal{E}$.

It is clear that Φ_t is well defined and satisfies a) (by the arguments in the proof of Proposition 3.1, we see that Φ_t is the restriction to the unit circle of a holomorphic function on an annulus). Since $t \mapsto G_t^{-1/2}$

is continuous from $[0,1]$ to the space of holomorphic functions on the annulus (in the L^∞ norm), it is continuous from $[0,1]$ to $C^\infty(\mathbb{T})$. So $t \mapsto \Phi_t$ is continuous from $[0,1]$ to $C^\infty(\mathbb{T})$ (and even maps into the class of functions which extend to a holomorphic function on an annulus).

Finally, Φ_t belongs to $\mathcal{E}_{\exp}$ for all $0 \le t < 1$ as Φ_t has no other zero on the unit disc than the one at -1. So the path is in $\mathcal{E}$, except, perhaps, for its endpoint Φ_1 which is m_0.

We have also proved

Proposition 3.4. *Let ψ a wavelet that arises from an* MRA *with a scaling function that has polynomial decay. Then there exists a continuous family of such wavelets, $t \mapsto \psi_t$, $t \in [0,1]$, such that $\psi_0 = h$, $\psi_1 = \psi$, where h is the Haar wavelet.*

Proposition 3.5. (P.G. Lemarié-Rieusset and G. Malgouyres [6]). *Let ψ a compactly supported wavelet that arises from an* MRA. *Then there exists a continuous family of such wavelets, $t \mapsto \psi_t$, $t \in [0,1]$, such that $\psi_0 = h$, $\psi_1 = \psi$, where h is the Haar wavelet.*

Observe that if the scaling function has polynomial decay, so does the wavelet Φ. "Continuous" means that $t \mapsto \psi_t$ is continuous from $[0,1]$ to $L^2((1+|x|)^n\,dx)$ for any n.

Finally, let us prove Lemma 3.3. In fact, we are going to show a version of the lemma, where trigonometric polynomials have been replaced by polynomials in z. It will be clear that Lemma 3.3 follows from Lemma 3.6 using $F_t(\xi) = e^{-il\xi}(1+e^{i\xi})P_t(e^{i\xi})/2$ for an appropriate positive integer l.

Lemma 3.6. *Let P be a polynomial in $z \in \mathbb{C}$ of degree less than or equal to N, such that $P(2) = 1$. There exists a continuous map $t \to P_t$, from $[0,1]$ to the space of polynomials of degree $\le N$ such that*

1) $P_t(2) = 1$, *if* $0 \le t \le 1$,

2) $P_1(z) = P(z)$,

3) $|P_t(z)|^2 = (1-t) + t\,|P(z)|^2$, *if* $|z| = 1$.

Moreover, there exists an integer k such that $P_0(z) = z^k$.

We can assume that $P(0) \neq 0$, otherwise $P(z) = z^j \tilde{P}(z)$ with $\tilde{P}(0) \neq 0$, and we can take $P_t(z) = z^j \tilde{P}_t(z)$. Let

$$Q(z) = z^N P(z) \bar{P}(\frac{1}{z}),$$

and

$$Q_t(z) = (1-t) z^N + t Q(z).$$

The map $t \mapsto Q_t$ is obviously continuous, $Q_t(2) = 1$, $Q_1 = Q$ and $Q_0(z) = z^N$. We shall introduce polynomials P_t such that $Q_t(z) = z^N P_t(z) \bar{P}_t(1/z)$. These polynomials are constructed with the aid of the zeros of the polynomials $Q_t(z)$ by an argument very similar to that used to establish the Lemma of Fejér-Riesz (see [3, p. 117]).

Lemma 3.7. *Let $z_1, z_2, \dots, z_N$ be the zeros of P (possibly repeated with their multiplicity) chosen so that $z_1, \dots, z_k$ are the only zeros inside the unit disc. Let $z_j = 1/\bar{z}_{j-N}$ for $j = N+1, \dots, 2N$. Then there exist $2N$ continuous functions on $(0,1]$ such that $z_1(t), z_2(t), \dots, z_{2N}(t)$ are the $2N$ zeros of Q_t, $z_j(t) = 1/\bar{z}_{j-N}(t)$ for $j = N+1, \dots, 2N$, and $z_1(t), \dots, z_k(t)$ are inside the unit disc while $z_{k+1}(t), \dots, z_N(t)$ are outside the unit disc.*

Let us remark that, since $Q_t(2) = 1$ we must have $z_j(t) \neq 1$. Assuming that Lemma 3.7 is proved, we can then define

$$P_t(z) = \prod_{j=1}^{N} \frac{z - z_j(t)}{1 - z_j(t)}, \qquad 0 < t \leq 1.$$

It is clear that $Q_t(z) = z^N P_t(z) \bar{P}_t(1/z)$ and $t \mapsto P_t$ is continuous. In order to obtain Lemma 3.6, it suffices to prove that $P_t \to z^k$ when $t \to 0$. But $Q_t \to z^N$ as $t \to 0$; then, for $t < \eta$, N of the $z_j(t)$'s are inside a small disc $\{|z| < \varepsilon\}$,and the other N (which are their reciprocals) are outside the disc $\{|z| < 1/\varepsilon\}$. That is to say, $z_1(t), \dots, z_k(t)$ tend to 0 while $|z_{k+1}(t)|, \dots, |z_N(t)|$ tend to ∞. Thus the polynomials $P_t(z)$ (each of degree $\leq N$) tend uniformly with the unit circle to z^k. This shows Lemma 3.6.

Hence, we just have to prove Lemma 3.7. Let us start by defining $z_j(t)$ for t close to 1. Let z_0 be a zero of P with multiplicity k_0.

First case: $|z_0| < 1$. We can assume that $z_1 = \cdots = z_{k_0} = z_0$. $1/\bar{z}_0$ may or may not be a zero of P. In the first case, let $k_0' \geq 1$ be its multiplicity, and let $z_{k+1} = \cdots = z_{k+k'_0} = 1/\bar{z}_0$. We shall define $z_j(t)$ for $j \in J_0 = \{1, \ldots, k_0, N+k+1, \ldots, N+k+k'_0\}$. We know that z_0 is a zero of Q with multiplicity $k_0 + k'_0$. In a neighborhood of z_0, $Q(z)/z^N$ can also be written $(z-z_0)^{k_0+k'_0}F(z)$, with $F(z_0) \neq 0$, and $Q_t(z)$ has $k_0 + k'_0$ distinct zeros which are solutions of

$$\frac{Q(z)}{z^N} = -\frac{1-t}{t} = (z-z_0)^{k_0+k'_0}\, F(z)\,. \tag{10}$$

From this, we obtain

$$e^{2\pi i l(j)/(k_0+k'_0)}\, s = (z-z_0)\, \alpha(z)^{-1}\,,$$

where $\alpha(z)^{k_0+k'_0} = (-F(z))^{-1}$, $s = ((1-t)/t)^{1/(k_0+k'_0)}$ and $j \mapsto l(j)$ is a bijection between J_0 and $\{1, \ldots, k_0 + k'_0\}$. It is then easy to see that we can define $z_j(t)$ so that $t \mapsto z_j(t)$ is tangent at z_0 to the half-lines $s \mapsto z_0 + \alpha\, e^{2\pi i l(j)/(k_0+k'_0)} s$, where $\alpha = \alpha(z_0)$.

When $1/\bar{z}_0$ is not a zero of P, we obtain the same result with k_0 instead of $k_0 + k'_0$.

Second case: $|z_0| > 1$. The previous reasoning applies and allows us to define $z_j(t)$ for t close to 1 and j such that $|z_j| \neq 1$.

Third case: $|z_0| = 1$. This time, z_0 is a zero of Q with multiplicity $2k_0$, and we can assume that $z_0 = z_{k+1} = \cdots = z_{k+k_0} = z_{N+k+1} = \cdots = z_{N+k+k_0}$. Once again, the question is to define $z_j(t)$ for $j \in J_0 = \{k+1, \ldots, k+k_0, N+k+1, \ldots, N+k+k_0\}$ so that $|z_j(t)| > 1$ for $j \leq N$ and $|z_j(t)| < 1$ for $j \geq N$. Again, the z_j's can be chosen tangent to the half-lines $s \mapsto z_0 + \alpha\, e^{2\pi i l(j)/2k_0} s$, where $\alpha^{2k_0} = -1/F(z_0)$ and l is a bijection between J_0 and $\{1, \ldots, 2k_0\}$.

But the positivity of $Q(z)/z^N$ on the unit circle implies that $(-1)^{k_0}$ $z_0^{-2k_0}\, F(z_0)$ is positive; hence, we can take $\alpha = z_0\, \beta$, where $\beta > 0$, if k_0 is odd, and $\alpha = z_0\, e^{i\pi/2k_0} \beta$, where $\beta > 0$, if k_0 is even. In both cases half of the half-lines are outside the unit circle, the rest are inside. We choose $l(j)$ so that the half-line lies outside the unit circle if $j \leq N$, while, for $j > N$, the half-line crosses the circle.

We can now finish the proof of Lemma 3.7. By continuity, $z_j(t)$ is well defined as long as it is distinct from $z_l(t)$ for $l \neq j$. Otherwise let ε be such that $z_{j_1}(t) \to z_0$, $z_{j_2}(t) \to z_0$, ..., $z_{j_l}(t) \to z_0$ when $t \to \varepsilon$, $\varepsilon > 0$, while the other zeros stay outside a neighborhood of z_0. That is to say, Q_ε has at z_0 a zero with multiplicity l. Moreover $z_0 \neq 0$ since $Q_\varepsilon(0) = \varepsilon Q(0)$ of z_0, we have, once again, $Q_\varepsilon(z)/z^N = (z-z_0)^l\, F(z)$, $F(z_0) \neq 0$; thus, for $t < \varepsilon$, $Q_t(z) = 0$ if and only if $(z-z_0)^l\, F(z) =$

$-(\varepsilon - t)/t$. As before we obtain continuous functions $z_{j_1}(t)(t), \dots, z_{j_l}(t)$ defined for $t \leq \varepsilon$, t close to ε, and equal to z_0 at ε.

REMARK. In the third case, we could just as well have chosen to reverse the property of the half-lines. That is to say that, among $z_1(t), \dots, z_N(t)$, we can choose l of them in the unit circle, with $k \leq l \leq k'$, where k is the number of zeros of P in the open unit circle, and k' the number of zeros of P in the closed unit circle. This is a way of interchanging z^k and z^l.

REMARK. In this paper, we considered only the case of 0-regular MRA's as they have been defined by Y. Meyer in [7]. The question whether r-regular MRA's have the same property remains open.

Acknowledgements. We wish to thank Professor John Hubbard who brought to our attention the problem of examining the connectivity notions we are investigating. Also, we wish to thank Professor P. G. Lemarié-Rieusset for several very useful discussions about the material we are presenting. The first author remembers quite well her informal discussion in Orsay with J. Hubbard and P. G. Lemarié-Rieusset; J. Hubbard, as a geometer, asked wether the Haar wavelet was isolated, and P. G. Lemarié-Rieusset showed immediately how to join it to any compactly supported wavelet. This paper is an attempt to give a more complete answer to the question of J. Hubbard, and relies deeply on the first construction given by P. G. Lemarié-Rieusset.

References.

[1] Cohen, A., *Ondelettes, analyses multirésolutions et traitement numérique du signal.* Rech. en Math. App. **27**, Masson, 1992.

[2] Daubechies, I., *Ten lectures on wavelets.* SIAM, 1992.

[3] Riesz, F., Sz.-Nagy, B., *Leçons d'analyses fonctionelle.* Akadémiai kiadó, 1952.

[4] Lemarié-Rieusset, P. G., Polynômes de Bernstein en théorie des ondelettes. *C. R. Acad. Sci. Paris* **319** (1994), 21-24.

[5] Lemarié-Rieusset, P. G., Fonctions d'échelle interpolantes, polynômes de Bernstein et ondelettes non stationnaires. Preprint, 1994.

[6] Lemarié-Rieusset, P. G., Malgouyres, G., personal communication.

[7] Meyer, Y., *Ondelettes et opérateurs.* Hermann, 1990.
[8] Rudin, W., *Real and complex analysis.* McGraw-Hill, 1987.

Recibido: 28 de octubre de 1.994
Revisado: 19 de febrero de 1.995

Aline Bonami and Sylvain Durand
URA 1803
Université d'Orléans
45067 Orléans Cedex 2, FRANCE
bonami@math.u-psud.fr
durand@ceremade.dauphine.fr

Guido Weiss*
Washington University in Saint Louis
Saint Louis, MO 63130, USA.
guido@math.wustl.edu

* This research was partially supported by a grant from the Southwestern Bell Company.

Revista Matemática Iberoamericana
Vol. 12, N.º 1, 1996

The spin of the ground state of an atom

Charles L. Fefferman and Luis A. Seco

In this paper we address a question posed by M. and T. Hoffmann-Ostenhof, which concerns the total spin of the ground state of an atom or molecule. Each electron is given a value for spin, $\pm 1/2$. The total spin is the sum of the individual spins.

For a neutral atom, say, of nuclear charge Z, if all Z electrons have the same spin, then the total spin would be $\pm Z/2$. There is a result of Lieb and Mattis [LM] where they show that in one dimension, ground states have lowest possible total spin. Their result also holds for a class of 3-dimensional systems which does not include the quantum atom. A related result [AL] extends this result to positive temperature, and also shows that for systems with certain parity constraints, spin alignment is in fact favored at all temperatures. It is expected that, for the atom, this is not the case, and there is a lot of spin-cancellation among the different electrons. In rigorous mathematical terms, this can be expressed in the form

$$\text{total spin } \leq C\, Z^{\gamma}, \qquad \gamma < 1\,.$$

The goal of this paper is to prove such a bound. Unfortunately, we do not have control over the constant C, which we only know to be independent of Z.

For a solid, or a molecule with many nuclei, it is expected that the total spin may get as large as the order of magnitude of the number of particles (or perhaps nuclei), which would account for ferro-magnetism.

It is also conjectured that for an atom, the order of magnitude of the spin can be as large as $Z^{1/3}$. For non-interacting radial systems, with degeneracy of the order of $Z^{1/3}$, this is certainly possible. In fact, Hund's rule, well known in chemistry, states that this degeneracy is resolved, after turning on the interaction, into making the spin as large as possible, which agrees with the $Z^{1/3}$ size of spin if one believes in atomic shells. The study of spin is also of interest because it determines qualitative properties of the wave functions (see [HHS]).

Throughout the paper, C will be used to denote any irrelevant large constant, c any irrelevant small constant, and $C_1, C_2, \ldots, c_1, c_2, \ldots,$ will denote carefuly chosen large and small constants respectively.

1. Definitions, background and theorem.

Consider the atomic hamiltonian

$$H_{Z,N} = \sum_{i=1}^{N} \left(-\Delta_{r_i} - \frac{Z}{|r_i|} \right) + \sum_{i<j} \frac{1}{|r_i - r_j|} ,$$

and $E(Z,N)$ its lowest eigenvalue when acting on the Hilbert Space

$$\mathcal{H} = \bigwedge_{i=1}^{N} L^2 \left(\mathbb{R}^3 \times \mathbb{Z}_2 \right) .$$

The atomic ground-state energy is defined as

$$E(Z) = \inf_{N \geq 0} E(Z,N) .$$

It is a result of [Ru] and [Si] that $E(Z,N)$, which is decreasing in N, achieves the infimum at a finite N_c, which physically corresponds to the largest number of electrons an atom can bind; by the HVZ theorem (see [CFHS]), the ground state of the atom, which we denote by Ψ, is then defined as the eigenfunction of $H(Z,N_c)$ with eigenvalue $E(Z)$. It was proved in [Zh], [Li1] and [Li2] that

$$Z \leq N_c \leq 2\,Z .$$

Throughout this paper we will consider any N between Z and N_c, (the interesting cases corresponding, of course, to either $N = Z$ or $N = N_c$) and Ψ will denote any ground state of $H_{Z,N}$ with energy $E(Z,N)$.

As a consequence of Lieb's bound for N_c the trivial upper bound for the total spin is Z.

Here, we will use

$$x = (r, \sigma)$$

to denote the variable in $\mathbb{R}^3 \times \mathbb{Z}_2$, with $r \in \mathbb{R}^3$ the space variable and $\sigma = \sigma(x) = \pm 1/2$ the spin variable. The total spin operator is now given by

$$\mathrm{S} = \sum_{i=1}^{N} \sigma^{\uparrow}(x_i) - \sigma^{\downarrow}(x_i)\,,$$

where

$$\sigma^{\uparrow}(x) = \begin{cases} 1/2\,, & \text{if } \sigma(x) = +1/2\,, \\ 0\,, & \text{otherwise,} \end{cases}$$

$$\sigma^{\downarrow}(x) = \begin{cases} 1/2\,, & \text{if } \sigma(x) = -1/2\,, \\ 0\,, & \text{otherwise.} \end{cases}$$

Basic to our strategy is the theory of atomic (spectral) asymptotics, and some version of atomic electric neutrality, all well known, which we now briefly review.

Associated to the atomic hamiltonian there is the Thomas-Fermi energy ([Fe], [Th]), which equals $c_{\mathrm{TF}}\, Z^{7/3}$ for an explicit negative constant c_{TF} and satisfies

$$E(Z) = c_{\mathrm{TF}}\, Z^{7/3} + O\left(Z^{7/3-\varepsilon}\right)\,, \qquad \varepsilon > 0\,, \tag{1}$$

which was proved in [LS]. We also have the Thomas-Fermi density ρ^Z_{TF}, and the Thomas-Fermi potential V^Z_{TF}, which satisfy the perfect scaling conditions

$$V^Z_{\mathrm{TF}}(r) = Z^{4/3}\, V\left(Z^{1/3}\, r\right)\,, \qquad \rho^Z_{\mathrm{TF}}(r) = Z^2\, \rho\left(Z^{1/3}\, r\right)\,,$$

for universal functions $\rho(r)$ and $V(r)$, which satisfy the Thomas-Fermi equations

$$\rho(r) = \frac{1}{3\,\pi^2}\, V^{3/2}(r)\,, \qquad \Delta V(r) = 4\pi\, \rho(r)\,.$$

Note that our definition of the Thomas-Fermi potential is the negative of the usual one. We refer the reader to [Li] for a great exposition of Thomas-Fermi theory. We also have the bound [Hi]

$$(2)\qquad \begin{aligned} c\,\mathcal{S}(x) \le V_{\mathrm{TF}}^{Z}(x) \le C\,\mathcal{S}(x)\,, \qquad \mathcal{S}(x) = \min\left\{\frac{Z}{|x|}, \frac{1}{|x|^4}\right\},\\ \left|\nabla V_{\mathrm{TF}}^{Z}(x)\right| \le C\,\mathcal{S}(x)\,|x|^{-1}\,. \end{aligned}$$

The expansion (1) can be continued into what is called the Scott asymptotics, namely

$$(3)\qquad E(Z) = c_{\mathrm{TF}}\, Z^{7/3} + \frac{1}{4}\, Z^2 + O(Z^{2-\varepsilon})\;, \qquad \varepsilon > 0\,.$$

The Z^2 term is not semiclassical; its nature comes from the coulomb singularities and is therefore a genuine quantum effect. This was first realized in [Sc], and proved rigorously (in the atomic case only) in [Hu], [SW1], [SW2] and [SW3]. Its proof for molecules is in [IS].

A refinement of (3) is also known for atoms, and it has the form

$$(4)\quad E(Z) = c_{\mathrm{TF}}\, Z^{7/3} + \frac{1}{4}\, Z^2 + c_{\mathrm{DS}}\, Z^{5/3} + O\Big(Z^{5/3-\varepsilon}\Big)\;, \qquad \varepsilon > 0\,,$$

obtained in [Di] and [Sch], and proved rigorously in [FS1], [FS2], [FS3], [FS4], [FS5], [FS6], [FS7] and [FS8]. The corresponding molecular problem remains open, but the techniques in [IS] come very close to proving a similar expression. The expansion in powers of Z almost surely ends here (see [En], [CFS1] and [CFS2]).

Concerning the electronic neutrality problem, we only need the following two facts, which can be found in [FS9] and [SSS]; they depend on a number $b > 0$ which, after the accurate asymptotics in (4), or even (3) with $\varepsilon = 1/3$, can be taken to be $b = 2/3$. They are expressed in terms of the ground state density, which is defined as

$$\begin{aligned} \rho_\Psi(r) &= \rho_\Psi^{\uparrow}(r) + \rho_\Psi^{\downarrow}(r),\\ \rho_\Psi^{\uparrow}(r) &= \tfrac{1}{2}N \int_{[\mathbb{R}^3\times\mathbb{Z}_2]^{(N-1)}} \left|\Psi(r,\tfrac{1}{2};x_2;\cdots;x_N)\right|^2 dx_2\cdots dx_N,\\ \rho_\Psi^{\downarrow}(r) &= \tfrac{1}{2}N \int_{[\mathbb{R}^3\times\mathbb{Z}_2]^{(N-1)}} \left|\Psi(r,-\tfrac{1}{2};x_2;\cdots;x_N)\right|^2 dx_2\cdots dx_N. \end{aligned}$$

1. The main result in [FS9] and [SSS] is

$$N_c = \int_{\mathbb{R}^3} \rho_\Psi(r)\, dr = Z + O\left(Z^{1-3b/7}\right). \tag{5.a}$$

2. The following is the content of estimate (A) or Lemma 2.1 in [FS9], or Lemma 6 in [SSS]:

$$\left| \int_{\mathbb{R}^3} \rho_\Psi(r)\, \chi(r)\, dr - \int_{\mathbb{R}^3} \rho_{\mathrm{TF}}^Z(r)\, \chi(r)\, dr \right| \leq C\, Z^{(7/3-b)/2}\, \|\nabla \chi\|_2 , \tag{5.b}$$

where χ is a positive function equal to 1 in a ball of radius at least $C\,Z^{-2/3}$, 0 outside of its double, and bounded by 1.

A common feature in both the asymptotic analysis and the neutrality problem is Lieb's inequality which also plays a crucial role in our analysis, and is by now part of the mathematical physics folkore ([Li]; see also [SW2], and for improvements [FS7], [Ba] and [GS]). We will use it in the following precise form,

Theorem (Lieb). *Assume $\psi(x_1, \ldots, x_N)$, $(Z \leq N \leq 2Z)$ is such that*

$$\|\nabla \psi\|_2^2 \leq C\, Z^{7/3} .$$

Then, we have that

$$\langle H_{Z,N}\, \psi, \psi \rangle \geq \langle H_{Z,N}^{\mathrm{ind}} \psi, \psi \rangle - \frac{1}{2} \iint \frac{\rho_{\mathrm{TF}}^Z(x)\, \rho_{\mathrm{TF}}^Z(y)}{|x-y|}\, dx\, dy - C'\, Z^{5/3} ,$$

where

$$H_{Z,N}^{\mathrm{ind}} = \sum_{i=1}^{N} \left(-\Delta_{x_i} - V_{\mathrm{TF}}^Z(x_i) \right) .$$

The proof of this result can be found in Lemma 2 in [SW2], which is stated in a special case, but its proof shows exactly this. The role of this inequality is that it reduces the analysis of systems with interaction to a system without it. Technically, the problem reduces to an asymptotic estimate for the sum of the negative eigenvalues of a fixed Schrödinger operator in $\mathbb{R}^3$ (see below). For convenience, given an operator H, we denote the sum of its negatives eigenvalues by $\operatorname{sneg} H$. We denote by

H_N^Ω the corresponding operator with Neumann boundary conditions on Ω.

The asymptotic estimates we need began with the work of Lieb and Simon. Those estimates, more refined ones even, are now also part of the folkore. We reproduce here a variant which suffices for our theorem. This is essentially contained in [LS] and explicitly proven in [FS7]; we include a version of the proof here for the convenience of the reader, and to make this paper as self-contained as possible.

Lemma 1. *If Q is a cube of side L, and K is a number larger than $100\,L^{-2}$, we have that*

$$\text{sneg}\,(-\Delta - K)_N^Q \geq -\frac{1}{15\,\pi^2}\,K^{5/2}\,L^3 - C\,K^2\,L^2\,,$$

for a universal constant C. If $K \leq M\,L^{-2}$ we have trivially

$$\text{sneg}\,(-\Delta - K)_N^Q \geq -M'\,L^{-2}\,.$$

PROOF. If $K\,L^2 \geq 100$,

$$\begin{aligned}\text{sneg}\,(-\Delta - K)_N^Q &= \sum_{\substack{\pi^2(n_1^2+n_2^2+n_3^2)\leq K L^2\\ n_i\geq 0}} \left(\frac{\pi^2\,(n_1^2+n_2^2+n_3^2)}{L^2} - K\right)\\ &= \frac{1}{8}\int\limits_{|x|\leq\sqrt{K}L/\pi} \left(\left|\frac{\pi\,x}{L}\right|^2 - K\right)\,dx + O(K^2\,L^2)\,.\end{aligned}$$

Lemma 2. *Let W be any potential satisfying*

$$W(x) \sim \mathcal{S}(x)\,,\qquad |\nabla W(x)| \leq C\,\mathcal{S}(x)\,|x|^{-1}\,,$$
$$\mathcal{S}(x) = \min\left\{\frac{Z}{|x|}\,,|x|^{-4}\right\}. \tag{6}$$

Then,

$$\text{sneg}\,(-\Delta - W) \geq -\frac{1}{15\,\pi^2}\int W(x)^{5/2}\,dx - \overline{C}\,Z^{13/6}\,,$$

where $\overline{C}$ only depends on the constants in (6).

PROOF. We break up $\mathbb{R}^3$ into cubes Q_0, Q_ν and Q'_ν with the properties:

1. Q_0 is centered at the origin and has diameter $d_0 = C_1 Z^{-1}$.

2. The Q_ν are centered at x_ν, with $C_1 Z^{-1}/10 \le |x_\nu| \le c$, and have diameters d_ν which satisfy

$$d_\nu \sim S^{-1/4}(x_\nu)\,|x_\nu|^{1/2}\,. \tag{7}$$

3. The $Q_{\nu'}$ are centered at $x_{\nu'}$, with $|x_{\nu'}| \ge c'$, and have diameter $d_{\nu'}$ which satisfy

$$10^{-5}\,|x_{\nu'}| \le d_{\nu'} \le \frac{1}{100}\,|x_{\nu'}|\,. \tag{8}$$

Let us check that $\mathbb{R}^3$ can be broken into such cubes. We begin with a simple geometric observation: if $Q(r)$ denotes the cube of diameter r centered at 0, then $Q(3r) - Q(r)$ may be decomposed into cubes of diameter r. It follows that $Q(3r) - Q(r)$ may be decomposed into subcubes of diameter between $s/3$ and s, for any given $s \le r$.

Now let $r_k = C_1 Z^{-1} 3^k$ for $k \ge 0$, and break up $\mathbb{R}^3$ into $Q(r_0) = Q_0$, and $Q(r_{k+1}) - Q(r_k)$ for $k \ge 0$. For $k \ge 0$ such that $r_k \le c$ we break up $Q(r_{k+1}) - Q(r_k)$ into cubes Q_ν of diameter

$$d_k \sim s_k = \left(\min\left\{\frac{Z}{r_k}, r_k^{-4}\right\}\right)^{-1/4} r_k^{1/2}\,,$$

which is possible since $s_k \le r_k$.

For $k \ge 0$ such that $r_k > c$, we break up $Q(r_{k+1}) - Q(r_k)$ into cubes $Q_{\nu'}$ of diameter between $r_k/3 \cdot 10^4$ and $10^{-4}\,r_k$. One checks easily that the resulting decomposition into cubes satisfies 1, 2 and 3 above.

Note that, by (8), the number of $Q_{\nu'}$ with centers in a spherical shell of radii R and $2R$, is not more than a fixed large constant and therefore,

$$\text{number}\,\{Q_{\nu'} :\ R_1 \le |x_{\nu'}| \le R_2\} \le C\,\log(R_2/R_1)\,, \tag{9}$$

when $R_2 \ge R_1$. In preparation to use Lemma 1, we denote

$$w_\nu = \max_{x \in Q_\nu} W(x)\,,$$

and we note that, when $x \in Q_\nu$,

$$\begin{aligned}\left|W^{5/2}(x) - {w_\nu}^{5/2}\right| &\le C\,{w_\nu}^{3/2}\,|W(x) - W(x_\nu)| \\ &\le C\,{w_\nu}^{3/2}\,d_\nu \max_{Q_\nu} |\nabla W(x)| \\ &\underset{\text{using (7)}}{\le} C\,\mathcal{S}(x_\nu)^{3/2}\,\mathcal{S}^{-1/4}(x_\nu)\,|x_\nu|^{1/2}\,\mathcal{S}(x_\nu)\,|x_\nu|^{-1} \\ &\le C\,\mathcal{S}^{9/4}(x_\nu)\,|x_\nu|^{-1/2}\,.\end{aligned}$$

This implies that

$$\begin{aligned}\left|{w_\nu}^{5/2}\,d_\nu^3 - \int_{Q_\nu} W^{5/2}(x)\,dx\right| &\le C\,\mathcal{S}^{9/4}(x_\nu)\,|x_\nu|^{-1/2}\,d_\nu^3 \\ &\le C \int_{Q_\nu} \mathcal{S}^{9/4}(x)\,|x|^{-1/2}\,dx\,.\end{aligned} \tag{10}$$

For Q_0, we have that

$$\begin{aligned}\text{sneg}\,(-\Delta - W(x))_N^{Q_0} &\ge \text{sneg}\left(-\Delta - \frac{C\,Z}{|x|}\right)_N^{Q_0} \\ &= Z^2\,\text{sneg}\left(-\Delta - \frac{C}{|x|}\right)_N^{\tilde{Q}_0}\,,\end{aligned}$$

where $\tilde{Q}_0$ is the cube Q_0 dilated by Z, which is therefore of diameter C_1 and makes the sneg term above independent of Z.

After this, we turn to sneg's by writing

$$\begin{aligned}\text{sneg}\,(-\Delta - W(x)) &\ge \text{sneg}\,(-\Delta - W(x))_N^{Q_0} + \sum_\nu \text{sneg}\,(-\Delta - W(x))_N^{Q_\nu} \\ &\quad + \sum_{\nu'} \text{sneg}\,(-\Delta - W(x))_N^{Q_{\nu'}} \\ &\ge -C\,Z^2 + \sum_\nu \text{sneg}\,(-\Delta - w_\nu)_N^{Q_\nu} \\ &\quad + \sum_{\nu'} \text{sneg}\,(-\Delta - W(x))_N^{Q_{\nu'}} \\ &\ge -C\,Z^2 + \sum_\nu \left(w_\nu^{5/2}\,d_\nu^3 - C\,w_\nu^2\,d_\nu^2\right) \\ &\quad + \sum_{\nu'} \text{sneg}\,(-\Delta - W(x))_N^{Q_{\nu'}}\,.\end{aligned} \tag{11}$$

For the $Q_{\nu'}$, we use the trivial part of Lemma 1 to obtain

$$
\begin{aligned}
\sum_{\nu'} \operatorname{sneg}\,(-\Delta - W(x))_N^{Q_{\nu'}} &\geq -C \sum_{\nu'} d_{\nu'}{}^{-2} \\
&\geq -C \sum_{n=\sim 1}^{\infty} 2^{-2n} \sum_{2^n \leq |x_{\nu'}| \leq 2^{n+1}} 1 \\
&\geq -C\,.
\end{aligned}
\tag{12}
$$

For the Q_ν, we have

$$
w_\nu^2\, d_\nu^2 \leq C\, \mathcal{S}^{9/4}(x_\nu)\, |x_\nu|^{-1/2}\; |Q_\nu| \leq C \int_{Q_\nu} \mathcal{S}^{9/4}(x)\, |x|^{-1/2}\, dx\,.
$$

Putting this, with (10) and (12) into (11), we obtain

$$
\begin{aligned}
\operatorname{sneg}\,(-\Delta - W(x)) \geq &-\frac{1}{15\,\pi^2} \int_{\cup_\nu Q_\nu} W^{5/2}(x)\, dx \\
&- C \int_{\mathbb{R}^3} \mathcal{S}^{9/4}(x)\, |x|^{-1/2}\, dx - C \\
\geq &-\frac{1}{15\,\pi^2} \int_{\mathbb{R}^3} W^{5/2}(x)\, dx - C\, Z^{13/6}\,,
\end{aligned}
$$

as we claimed.

We are now ready to state and prove our main result:

Theorem 3. *If Ψ is the ground state for $H_{Z,N}$ for $Z \leq N \leq N_c$, then we have*

$$
|\langle \mathrm{S}\,\Psi, \Psi\rangle| \leq C\, Z^\gamma\,, \qquad \gamma < 1\,.
$$

PROOF. Let $\delta > 0$ be a small number to be chosen later, and consider a positive function χ, bounded by 1 and as smooth as possible, such that

$$
\chi(r) = \begin{cases} 1\,, & \text{if } |r| < Z^{-1/3+\delta}\,, \\ 0\,, & \text{if } |r| > 2\,Z^{-1/3+\delta}\,. \end{cases}
$$

For a real number μ in the range

$$
|\mu| \leq c_1\, Z^{-4\delta}\,,
\tag{13}
$$

consider the hamiltonian given by

$$H_\mu = H_{Z,N} + 2\,\mu\, Z^{4/3} \mathbf{S}_\chi \,,$$

where

$$\mathbf{S}_\chi = \sum_{i=1}^{N} \chi(r_i)\left(\sigma^\uparrow(x_i) - \sigma^\downarrow(x_i)\right) ,$$

and denote by $E_\mu(Z)$ the corresponding ground state energy. Note that

$$\mathbf{S} = \mathbf{S}_\chi + \mathbf{S}_{1-\chi} \,.$$

We will study $\mathbf{S}_\chi$ first using H_μ; later, $\mathbf{S}_{1-\chi}$ will be easily dominated using (5.).

We define the Thomas-Fermi approximation to E_ν,

$$\begin{aligned}\mathcal{E}_\mu(Z) = -\frac{Z^{7/3}}{15\,\pi^2} \int \Big(& \left(V(r) + \mu\,\chi(Z^{-\delta}\, r)\right)_+^{5/2} \\ & + \left(V(r) - \mu\,\chi(Z^{-\delta}\, r)\right)_+^{5/2} \Big)\, dr \\ & - \frac{1}{2} \iint \frac{\rho_{\mathrm{TF}}^Z(x)\, \rho_{\mathrm{TF}}^Z(y)}{|x-y|}\, dx\, dy \,,\end{aligned}$$

which plays the following role:

Proposition 4. *There is a constant C such that*

$$E_\mu(Z) \geq \mathcal{E}_\mu(Z) - C\, Z^{7/3-\varepsilon_1} \,, \qquad \varepsilon_1 = \frac{1}{6} \,,$$

uniformly for all $|\mu| \leq c_1\, Z^{-4\delta}$.

REMARK. Although the corresponding upper bound is most probably also true, we will have no need for it here, and we ignore the issue.

PROOF. Note first that our assumption (13) on μ implies that

$$(14) \qquad |\mu|\, Z^{4/3}\, \chi(x) \leq \frac{1}{2}\, V_{\mathrm{TF}}^Z(x) \,, \qquad \text{for all } x \,.$$

Indeed, this is clear for $|x| < Z^{-1/3}$, and is also obvious for $|x| \geq 2\, Z^{-1/3+\delta}$. For the other x, we have that $|\mu|\, Z^{4/3}\, \chi(x) \leq c_1\, Z^{4/3-4\delta}$, whereas $V^Z_{\mathrm{TF}}(x) \geq c\,|x|^{-4} \geq c\,Z^{4/3-4\delta}$, and (14) then follows by taking c_1 small enough. Estimate (14) in turn implies that

$$\frac{1}{2}\,V^Z_{\mathrm{TF}}(x) \leq V^Z_{\mathrm{TF}}(x) \pm \mu\,Z^{4/3}\,\chi(x) \leq \frac{3}{2}\,V^Z_{\mathrm{TF}}(x)\,. \tag{15}$$

In preparation to use Lieb's inequality, we compute the kinetic energy of a ground state Ψ_μ for E_μ, (or elements of a sequence with energy converging to E_μ) with a virial argument as follows: define

$$\mathrm{KE}(\psi) = \|\nabla\psi\|_2^2\,, \qquad \mathrm{PE}(\psi) = \langle V_{\mathrm{Coulomb}}\psi, \psi\rangle\,,$$

with

$$V_{\mathrm{Coulomb}}(x_1, \ldots, x_N) = -\sum_{i=1}^{N} \frac{Z}{|r_i|} + \sum_{i\neq j} \frac{1}{|r_i - r_j|}\,,$$

and denote the approximate ground-state sequence by $\Psi_{\mu,k}$. We denote their densities by $\rho_{\mu,k}$.

For $\lambda > 0$, denote

$$\Psi^\lambda_{\mu,k}(x_1, \ldots, x_N) = \lambda^{3N/2}\,\Psi_{\mu,k}(\lambda\,x_1, \ldots, \lambda\,x_N)\,,$$

and note that

$$\begin{aligned} f(\lambda) &= \langle H_\mu \Psi^\lambda_{\mu,k}, \Psi^\lambda_{\mu,k}\rangle \\ &= \lambda^2\,\mathrm{KE}(\Psi_{\mu,k}) + \lambda\,\mathrm{PE}(\Psi_{\mu,k}) \\ &\quad + \mu\,Z^{4/3} \int \chi(\lambda^{-1}x)\,\left(\rho^\uparrow_{\mu,k}(x) - \rho^\downarrow_{\mu,k}(x)\right)\,dx\,, \end{aligned}$$

is a smooth function which satisfies

$$\lim_{\lambda\to 0} f(\lambda) = 0\,, \qquad \lim_{\lambda\to\infty} f(\lambda) = \infty\,.$$

Also, using Ψ as trial function for H_μ and taking k large enough, we see that

$$f(1) \leq \frac{1}{2}\,E(Z) + c_1\,Z^{7/3-\delta}\,.$$

By (1), the right hand side is negative for all Z larger than a certain constant depending on c_1. Therefore f attains its minimum at some $0 < \lambda < \infty$ and, maybe by changing our sequence $\Psi_{\mu,k}$ to another whose energy converges faster to the ground state energy, we can rescale the

$\Psi_{\mu,k}$ so that the minimum of f is attained at $\lambda = 1$ and thus $f'(1) = 0$. This means that

$$2\,\mathrm{KE}(\Psi_{\mu,k}) + \mathrm{PE}(\Psi_{\mu,k}) \le 2\,|\mu|\,Z^{4/3} \int |\nabla\chi(x)|\,|x|\,\rho_{\Psi_{\mu,k}}(x)\,dx \le C\,Z^{7/3}.$$

Using $\Psi_{\mu,k}$ as trial function for $H_{Z,N}$, we see that

$$\mathrm{KE}(\Psi_{\mu,k}) + \mathrm{PE}(\Psi_{\mu,k}) \ge -C\,Z^{7/3}.$$

Altogether, we conclude that

$$\mathrm{KE}(\Psi_{\mu,k}) \le C\,Z^{7/3}.$$

In view of Lieb's inequality, it is then quite obvious that

$$E_\mu \ge \mathrm{sneg}\left(-\Delta - V_{\mathrm{TF}}^Z + \mu\,Z^{4/3}\,\chi\right) + \mathrm{sneg}\left(-\Delta - V_{\mathrm{TF}}^Z - \mu\,Z^{4/3}\,\chi\right) - \frac{1}{2}\iint \frac{\rho_{\mathrm{TF}}^Z(x)\,\rho_{\mathrm{TF}}^Z(y)}{|x-y|}\,dx\,dy - C\,Z^{5/3}.$$

Set

$$W(x) = V_{\mathrm{TF}}^Z(x) - \mu\,Z^{4/3}\,\chi(x),$$

and recall (2) and (15) which, with the equally trivial bound

$$\left|\mu\,Z^{4/3}\,\nabla\chi(x)\right| \le C\,|x|^{-5},$$

show that W satisfies (6). Lemma 2 then proves our result.

Now, we consider the following lemma:

Lemma 5. *$\mathcal{E}_\mu$, as a function of μ, is concave, and there is a constant C such that*

$$C^{-1}\,Z^{7/3} \le |\mathcal{E}_\mu(Z)| \le C\,Z^{7/3}, \qquad \left|\frac{\partial^2\mathcal{E}_\mu(Z)}{\partial\mu^2}\right| \le C\,Z^{7/3+\delta},$$

uniformly for all $|\mu| \le c_1\,Z^{-4\delta}$.

PROOF. After checking that (15) settles the first bounds in the statement of the lemma, a calculation gives

$$\begin{aligned}\left|\frac{\partial^2 \mathcal{E}_\mu(Z)}{\partial \mu^2}\right| &\le C\, Z^{7/3} \int_{\mathbb{R}^3} \Big(\big(V(x) + \mu\, \chi(Z^{-\delta}\, x)\big)^{1/2} \\ &\qquad + \big(V(x) - \mu\, \chi(Z^{-\delta}\, x)\big)^{1/2} \Big)\, \chi^2\, \left(Z^{-\delta}\, x\right)\, dx \\ &\le C\, Z^{7/3} \int\limits_{|x| \le 2\, Z^\delta} V^{1/2}(x)\, dx \\ &\le C\, Z^{7/3} \int\limits_{|x| \le 2\, Z^\delta} |x|^{-2}\, dx \\ &\le C\, Z^{7/3+\delta}\,.\end{aligned}$$

After this, we simply observe that $\mathcal{E}_0(Z) = c_{\mathrm{TF}}\, Z^{7/3}$ (again, see [Li]), and note that $\mathcal{E}_\mu(Z)$ is an even function of μ to conclude that, for μ in our range, we must have

$$\begin{aligned}\mathcal{E}_\mu(Z) &\ge c_{\mathrm{TF}}\, Z^{7/3} - \frac{\mu^2}{2} \sup_\mu \left|\frac{\partial^2 \mathcal{E}_\mu}{\partial \mu^2}\right| \\ &\ge c_{\mathrm{TF}}\, Z^{7/3} - C\, \mu^2\, Z^{7/3+\delta}\,,\end{aligned}$$

which implies

$$E_\mu(Z) \ge c_{\mathrm{TF}}\, Z^{7/3} - C\, \mu^2\, Z^{7/3+\delta} - C\, Z^{7/3-\varepsilon_1}\,. \tag{16}$$

On the other hand, if we denote by Ψ any ground state of the atom, we use it as a trial function to conclude that

$$E_\mu(Z) \le E(Z) + 2\, \mu\, Z^{4/3}\, \langle \mathrm{S}_\chi\, \Psi, \Psi \rangle\,.$$

If we now use as trial function the same Ψ, but with spins reversed, we obtain

$$E_\mu(Z) \le E(Z) - 2\, \mu\, Z^{4/3}\, \langle \mathrm{S}_\chi\, \Psi, \Psi \rangle\,.$$

Altogether, we obtain

$$E_\mu(Z) \le E(Z) - 2\, |\mu|\, Z^{4/3}\, \left|\langle \mathrm{S}_\chi\, \Psi, \Psi \rangle\right|\,.$$

Since

$$E(Z) = c_{\mathrm{TF}}\, Z^{7/3} + O(Z^2)\,,$$

we conclude that

$$E_\mu(Z) \le c_{\mathrm{TF}}\, Z^{7/3} - 2|\mu|\, Z^{4/3}\, |\langle \mathrm{S}_\chi\, \Psi, \Psi\rangle| + O(Z^2)\ .$$

Putting this together with (16), we obtain

$$|\mu|\, \frac{|\langle \mathrm{S}_\chi\, \Psi, \Psi\rangle|}{Z} \le C\, Z^{\delta}\, |\mu|^2 + C\, Z^{-\varepsilon_1}\,, \qquad |\mu| \le c_1\, Z^{-4\delta}\,.$$

If we choose now

$$|\mu| = c_1\, Z^{-4\delta}\,,$$

we obtain

$$\frac{|\langle \mathrm{S}_\chi\, \Psi, \Psi\rangle|}{Z} \le C\, Z^{-3\delta} + C\, Z^{-\varepsilon_1 + 4\delta}\,. \tag{17}$$

Finally, we have

$$|\langle \mathrm{S}_{1-\chi}\, \Psi, \Psi\rangle| \le N - \int \rho_\Psi(r)\, \chi(r)\, dr\,. \tag{18}$$

Since

$$\left| Z - \int \rho_{\mathrm{TF}}^{Z}(r)\, \chi(r)\, dr \right| \le C\, Z^{1-3\delta}\,,$$

we use (5.b) with $b = 2/3$ to conclude that

$$\begin{aligned} \left| N - \int \rho_\Psi(r)\, \chi(r)\, dr \right| &\le |Z - N_c| + C\, Z^{1-3\delta} + C\, Z^{(4/3+\delta)/2} \\ &\le C\, Z^{5/7} + C\, Z^{1-3\delta} + C\, Z^{(4/3+\delta)/2}\,. \end{aligned}$$

By (18), we conclude that

$$|\langle \mathrm{S}_{1-\chi} \Psi, \Psi\rangle| \le C\, Z^{5/7} + C\, Z^{1-3\delta} + C\, Z^{(4/3+\delta)/2}\,.$$

With $\varepsilon_1 = 1/6$, we choose $\delta = 1/42$ here and in (17) to conclude Theorem 3 with $\gamma = 13/14$.

Our proof of theorem 3 with $\gamma = 13/14$ was kept simple because we used a form of spectral asymptotics in Lemma 2 which is not very involved. If we used the sharper version given by Theorem 6 below, and the sharper atomic energy asymptotics in (4), then we would obtain, with the same arguments, a bound with $\gamma = 5/7$. But we would also drive the careful reader into the pain and suffering involved in reading the contents of [FS 2-8], which contains the proof of Theorem 6 below and (4). It is interesting to point out that the bound such analysis would yield, $\gamma = 5/7$, is the same as the bound we know for electric neutrality. And this is not because spin neutrality used electric neutrality: if we imposed electric neutrality to our atoms, by studying $H_{Z,Z}$ instead, we would obtain the same exponent.

We end by stating the theorem, proved in [FS5], which we mentioned above. Our potential W is easily checked to satisfy hypothesis (1), (2) and (3) below.

Theorem 6. *Suppose $W(r)$ is defined on $(0,\infty)$ and satisfies the following conditions:*

$$\left|\left(\frac{d}{dr}\right)^{\alpha} W(r)\right| \leq C_\alpha\, \mathcal{S}(r)\, r^{-\alpha}, \tag{1}$$

for all $r \in (0,\infty)$, $\alpha \geq 0$,

$$\left|\left(\frac{d}{dr}\right)^{\alpha} \left(W(r) - V_{\mathrm{TF}}^{Z}(r)\right)\right| \leq c_0\, \mathcal{S}(r)\, r^{-\alpha}, \tag{2}$$

for all $r \in (0,\infty)$, $\alpha = 0,1,2$, with $c_0 > 0$ determined by the C_α in (1),

$$\left|\left(\frac{d}{dr}\right)^{\alpha} \left(E_0 - \frac{Z}{r} + W(r)\right)\right| \leq C_\alpha\, Z^{3/2}\, r^{1/2-\alpha}, \tag{3}$$

for all $r \in (0, 2\,Z^{-3/5+2\varepsilon})$, $\alpha \geq 0$, with $c\,Z^{4/3} < E_0 < C\,Z^{4/3}$ and $0 < \varepsilon < 10^{-12}$.

Set Ω equal to the positive root of $\Omega(\Omega+1) = \max_{r>0} r^2\, W(r)$,

$$\left.\begin{aligned} \eta_l &= \frac{1}{\pi}\int_0^\infty \left(W(r) - \frac{l(l+1)}{r^2}\right)_+^{-1/2} dr \\ \phi_l &= \frac{1}{\pi}\int_0^\infty \left(W(r) - \frac{l(l+1)}{r^2}\right)_+^{1/2} dr \end{aligned}\right\} \qquad (1 \leq l \leq \Omega).$$

Then,

$$\begin{aligned} \operatorname{sneg}\left(-\Delta + W(|x|)\right) = &-\frac{1}{15\,\pi^2}\int_{\mathbb{R}^3} W^{5/3}(|x|)\,dx + \frac{Z^2}{8} \\ &- \frac{1}{48\,\pi^2}\int_{\mathbb{R}^3} W^{1/2}(|x|)\,\Delta W(|x|)\,dx \\ &+ \sum_{Z^{8/25+10\varepsilon} < l < \Omega} \frac{2l+1}{\eta_l}\,\mu(\phi_l) + \mathrm{Error}, \end{aligned}$$

with $|\text{Error}| \leq C' Z^{8/5+2\cdot 10^{-9}}$ *and* $\mu(t)$ *denotes the fractional part of* t. *The constant* C' *depends only on* C_α, c_0, C *and* ε *in* (1), (2) *and* (3). *Furthermore, the last sum is easily seen to be bounded by* $C\,Z^{5/3}$.

Acknowledgements. We wish to thank M. and T. Hoffmann-Ostenhof for getting us interested in the problem; one of us (L. S.) wishes to express his gratitude to them and to the Schrödinger Institute for their hospitality and financial support during a stay in Vienna, where part of this work was done. L. S. is also grateful to A. Córdoba and the Department of Mathematics of the Univeridad Autónoma de Madrid, were this work was finalized. This work is partially supported by an NSF grant (Ch. F.) and an NSERC and CICYT grant (L. S.).

References.

[AL] Aizenmann, M. and Lieb, E., Magnetic properties of some itinerant-electron systems at $T > 0$. *Phys. Rev. Lett.* **65** (1990), 1470-1477

[Ba] Bach, V., Accuracy of mean field approximations for atoms and molecules. *Comm. Math. Phys.* **155** (1993), 295-310.

[CFS1] Córdoba, A., Fefferman, Ch., Seco, L., A trigonometric sum relevant to the non-relativistic Theory of Atoms. *Proc. Nat. Acad. Sci. USA* **91** (1994), 5776-5778.

[CFS2] Córdoba, A., Fefferman, Ch., Seco, L., Weyl sums and atomic energy oscillations. *Revista Mat. Iberoamericana* **11** (1995), 167-228.

[CFHS] Cycon, H. L., Froese, R. G., Kirsch, W., Simon, B., Schrödinger operators. Springer Verlag, (1987).

[Di] Dirac, P., Note on exchange fenomena in the Thomas-Fermi atom. *Proc. Cambridge Phil. Soc.* **26** (1930), 376-385.

[En] Englert, B. G., *Semiclassical Theory of the Atom.* Springer Verlag Lecture Notes in Physics **301**, (1991).

[FS1] Fefferman, Ch., Seco, L., The ground-state energy of a large atom. *Bull. Amer. Math. Soc.* **23** (1990), 525-530.

[FS2] Fefferman, Ch., Seco, L., Eigenvalues and eigenfunctions of ordinary differential operators. *Advances in Math.* **95** (1992), 145-305.

[FS3] Fefferman, Ch., Seco, L., The eigenvalue sum for a one-dimensional potential. *Advances in Math.* **108** no. 2, (1994), 263-335.

[FS4] Fefferman, Ch., Seco, L., The density in a one-dimensional potential. *Advances in Math.* **107** (1994), 187-364.

[FS5] Fefferman, Ch., Seco, L., The eigenvalue sum for a three-dimensional radial potential. To appear in *Advances in Math.*

[FS6] Fefferman, Ch., Seco, L., The density in a three-dimensional radial potential. *Advances in Math.* **111** no. 1, (1995), 88-161.

[FS7] Fefferman, Ch., Seco, L., On the Dirac and Schwinger corrections to the ground-state energy of an atom. *Advances in Math.* **107** (1994), 1-185.

[FS8] Fefferman, Ch., Seco, L., Aperiodicity of the hamiltonian flow in the Thomas-Fermi potential. *Revista Mat. Iberoamericana* **9** (1993), 409 - 551.

[FS9] Fefferman, Ch., Seco, L., Asymptotic neutrality of large ions. *Comm. Math. Phys.* **128** (1990), 109-130.

[Fe] Fermi, E., Un metodo statistico per la determinazione di alcune priorieta dell'atome. *Rend. Accad. Naz. Lincei* **6** (1927), 602-607.

[GS] Graf, G. M., Solovej, J. P., A correlation estimate with applications to quantum systems with Coulomb interactions. *Reviews in Math. Phys.* **6**, no. 5a (1994), 977-997.

[Hi] Hille, E., On the Thomas-Fermi equation. *Proc. Nat. Acad. Sci. USA* **62** (1969) 7-10.

[HHS] Hoffmann-Ostenhof M., Hoffmann-Ostenhof, T., Stremnitzer, H., Local properties of Coulombic wave functions. *Commun. Math. Phys.* **163** (1994), 185-215.

[Hu] Hughes, W., An Atomic Energy Lower Bound that Agrees with Scott's Correction. *Advances in Mathematics* **79**, (1990) 213-270.

[IS] Ivrii, V., Sigal, I. M., Asymptotics of the ground state energies of large Coulomb systems. *Ann. of Math.* **138** no. 2, (1993) 243-335.

[Li] Lieb, E. H., A lower bound for Coulomb energies. *Phys. Lett.* **70A** (1979), 444-446.

[Li1] Lieb, E. H., Atomic and molecular ionization. *Phys. Rev. Lett.* **52** (1984) 315.

[Li2] Lieb, E. H., Bound on the maximum negative ionization of atoms and molecules. *Phys. Rev.* **A29** (1984), 3018-3028.

[LM] Lieb, E. H., Mattis, D., Theory of ferromagnetism and the ordering of electronic energy levels. *Phys. Rev.* **125** (1962), 164-170.

[LS] Lieb, E. H., Simon, B., Thomas-Fermi Theory of Atoms, Molecules and Solids *Advances in Math.* **23** (1977), 22-116.

[Ru] Ruskai, M. B., Absence of discrete spectrum in highly negative ions, I & II. *Comm. Math. Phys.* **82** (1982), 457-469 and **85**, 325-327.

[Si] Sigal, I. M. Geometric methods in the quantum many-body problem. Nonexistence of very negative ions. *Comm. Math. Phys.* **85** (1982), 309-324.

[Sch] Schwinger, J., Thomas-Fermi model: The second correction. *Phys. Rev.* *A*24 **5** (1981), 2353-2361.

[Sc] Scott, J. M. C., The binding energy of the Thomas-Fermi atom. *Phil. Mag.* **43** (1952), 859-867.

[SSS] Seco, L., Sigal, I. M., Solovej, J. P., Bound on the ionization energy of large atoms. *Comm. Math. Phys.* **131** (1990), 307-315.

[SW1] Siedentop, H., Weikard, R., On the leading energy correction for the statistical model of the atom: interacting case. *Comm. Math. Phys.* **112** (1987), 471-490.

[SW2] Siedentop, H., Weikard, R., On the leading correction of the Thomas-Fermi model: lower bound, and an appendix by A. M. K. Müller. *Invent. Math.* **97** (1989), 159-193.

[SW3] Siedentop, H., Weikard, R., A new phase space localization technique with applications to the sum of negative eigenvalues of Schrödinger operators. *Ann. Sci. Ecole Norm. Sup.* **24** (1991), 215-225.

[Th] Thomas, L. H., The calculation of atomic fields. *Proc. Cambridge Phil. Soc.* **23** (1927), 542-548.

[Zh] Zhislin, G., Discussion of the Spectrum of Schrödinger Operators for System of many Particles. *Tr. Mosk. Mat. Obs.* **9** (1960), 81-128.

Recibido: 14 de noviembre de 1.994

Charles L. Fefferman
Department of Mathematics
Princeton University
Princeton, NJ 08544, U.S.A.
cf@math.princeton.edu

and

Luis A. Seco
Department of Mathematics
University of Toronto
100 St. George St
Toronto, Ontario M5AS 1A1, CANADA
seco@math.toronto.edu

Revista Matemática Iberoamericana
Vol. 12, N.° 1, 1996

Correspondances géodésiques entre les surfaces euclidiennes à singularités coniques

Mohammed Mostefa Mesmoudi

Abstract. A. J. Montesinos has shown that a geodesic correspondence between two complete Riemannian manifolds with transitive topological geodesic flow is a homothety. In this paper we prove a similar result for a conformal geodesic correspondence between two singular flat surfaces with conical singularities and negative concentrated curvature.

Résumé. A. J. Montesinos a montré qu'une correspondance géodésique entre deux variétés riemanniennes complètes à flot géodésique topologiquement transitif est une homothétie. Dans le même esprit, nous montrons dans cet article qu'une correspondance géodésique conforme entre deux surfaces euclidiennes à singularités coniques avec des courbures concentrées négatives est une homothétie.

Introduction.

Une correspondance géodésique entre deux surfaces M et M', munies de deux métriques riemanniennes à singularités possible g et g' respectivement, est un difféomorphisme $f : M \longrightarrow M'$ qui envoie les géodésiques (non paramétrées) de (M,g) sur celles de (M',g'). Dans [Mon], Montesinos démontre le théorème suivant: Si (M,g) est une variété riemannienne de dimension $n > 1$, complète, à flot géodésique topologiquement transitif alors toute correspondance géodésique $f : (M,g) \longrightarrow (M',g')$, où (M',g') est une variété riemannienne de dimension $n > 1$, est une homothétie (c'est-à-dire il existe une constante c telle que $g'(dfv, dfw) = c\,g(v,w)$ pour tous v et w du fibré tangent TM).

Dans [Bon], F. Bonahon conjecture que tout homéomorphisme entre deux surfaces, munies de deux métriques riemanniennes à courbure strictement négative, qui envoie les géodésiques de la première surface sur les géodésiques de la deuxième surface est une homothétie. Dans le même contexte, on démontre dans cet article que toute correspondance géodésique conforme entre deux structures plates singulières est une homothétie.

Définition. *Soit M une surface sans bord. Une métrique plate singulière μ sur M est une métrique riemannienne sur $M \setminus \operatorname{Sing}\mu$, où $\operatorname{Sing}\mu$ est un sous-ensemble discret de points de M, vérifiant les deux conditions suivantes:*

1. *En dehors des points de $\operatorname{Sing}\mu$, la métrique μ est localement isométrique à la métrique euclidienne sur le plan.*

2. *Pour chaque point p de $\operatorname{Sing}\mu$, il existe un voisinage V_p de p et deux nombres réels $\theta_p \neq 2\pi$ et $\varepsilon > 0$ tels que V_p soit isométrique au cône d'angle conique θ_p défini par:*

$$\{(r,t) : \ 0 \leq r \leq \varepsilon \,, t \in \mathbb{R}/(\theta_p \mathbb{Z})\}/_{(0,t)\sim(0,t')}$$

muni de la métrique $ds^2 = dr^2 + r^2\,dt^2$. Le point p est appelé une singularité de μ et le nombre θ_p l'angle conique de la singularité p.

On étend cette définition d'une façon naturelle à une surface à bord. Un voisinage d'un point régulier sur le bord est isométrique à un voisinage d'un point sur le bord d'un demi-plan euclidien. Les singularités sur le bord sont des points anguleux dont l'angle correspondant est différent de π. Lorsque M est compacte, l'ensemble $\operatorname{Sing}\mu$ est fini.

Un segment géodésique dont l'intérieur ne contient pas de singularités est un segment euclidien. Un segment géodésique g dont l'une des extrémités est une singularité d'angle conique supérieur à 2π peut se prolonger au voisinage de cette singularité en une infinité de segments géodésiques tel que chacun des deux angles bordés par g et chacun de ces prolongés est supérieur à π, [fig.1].

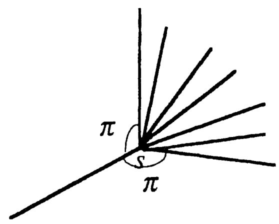


Figure 1

Proposition 1. *Sur une surface M simplement connexe munie d'une métrique plate singulière dont l'angle conique de chaque singularité est supérieur à 2π, chaque courbe qui est localement géodésique est une géodésique globale.*

DÉMONSTRATION. Rappelons d'abord que la caractéristique d'Euler d'une surface compacte munie d'une métrique plate singulière vérifie la formule de Gauss-Bonnet (voir par exemple [Tro]) suivante

$$2\pi\,\chi(M) = \sum_{s\in \mathrm{Sing}\,\mu\cap \mathrm{int}\, M} (2\pi - \theta_s) + \sum_{s\in \mathrm{Sing}\,\mu\cap\partial M} (\pi - \theta_s)\,,$$

où θ_s désigne l'angle conique d'une singularité s appartenant à l'intérieur de M dans la première somme et sur le bord de M dans la seconde. Soit g une courbe localement géodésique sur M. Supposons qu'il existe deux points a et b de g tels que le chemin le plus court ab qui les joint n'est pas inclus dans g. désignons par S le segment de g d'extrémités a et b. Le segment S borde avec ab au moins un disque polygonal. Soit D un tel disque. Ce disque a au plus deux sommets s_1 et s_2 d'angles (à l'intérieur de D) inférieurs strictement à π. La formule de Gauss-Bonnet appliquée à D donne

$$2\pi = \sum_{s\in \mathrm{Sing}\,\mu\cap \mathrm{int}\, D} (2\pi - \theta_s) + \sum_{s\in \mathrm{Sing}\,\mu - \{s_1,s_2\}\cap\partial D} (\pi - \theta_s) + (2\pi - \theta_{s_1} - \theta_{s_2})$$

ce qui implique

$$\theta_{s_1} + \theta_{s_2} = \sum_{s\in \mathrm{Sing}\,\mu\cap \mathrm{int}\, D} (2\pi - \theta_s) + \sum_{s\in \mathrm{Sing}\,\mu - \{s_1,s_2\}\cap\partial D} (\pi - \theta_s) < 0\,.$$

Ceci est évidemment impossible. Par conséquent, ab coïncide avec S.

Le corollaire suivant est immédiat.

Corollaire 1. *Sur une surface munie d'une métrique plate singulière dont l'angle conique de chaque singularité est supérieur à 2π, tout segment géodésique est unique dans sa classe d'homotopie relative.*

Par la suite, on suppose que M et M' sont deux surfaces compactes sans bord de même genre $g > 1$ et que ρ_1 et ρ_2 sont deux métriques plates définies respectivement sur M et M', singulières si $g > 1$ et telles que tous les angles coniques des singularités sont supérieurs à 2π . Pour chaque singularité s d'angle conique θ_s, le nombre $2\pi - \theta_s$ est appelé courbure concentrée de ρ_1 (ou ρ_2) en s. Ainsi on suppose que toutes les courbures concentrées sont négatives.

Proposition 2. *L'inverse d'une correspondance géodésique est une correspondance géodésique. De plus, si $f : (M, \rho_1) \longrightarrow (M', \rho_2)$ est une correspondance géodésique, l'image de tout point singulier de (M, ρ_1) est un point singulier dans (M', ρ_2).*

DÉMONSTRATION: Soit a un arc géodésique dans (M', ρ_2). Si $f^{-1}(a)$ n'est pas géodésique, l'image par f d'un segment géodésique a' homotope à $f^{-1}(a)$ à extrémités fixes est un segment géodésique dans (M', ρ_2) homotope au segment a à extrémités fixes. Ceci est en contradiction avec le Corollaire 1.

Un segment géodésique a dont l'une des extrémités est une singularité s peut se prolonger en une infinité des segements géodésiques [fig. 1]. Les images de ces segments par f sont des segments géodésiques qui prolongent $f(a)$ en $f(s)$. Le point $f(s)$ est alors une singularité. La proposition est démontrée.

Proposition 3. *Si s est une singularité de ρ_1 ayant un angle conique θ_s, alors $f(s)$ est une singularité de ρ_2 ayant le même angle conique θ_s .*

Pour démontrer cette proposition on a besoin de deux lemmes.

Lemme 1. *L'image par f d'un secteur basé en s d'angle π est un secteur basé en $f(s)$ d'angle π .*

DÉMONSTRATION. Soit a un segment géodésique sur (M, ρ_1) dont s est l'une des extrémités. Le segment a se prolonge au point s en une infinité de géodésiques (un faisceau de géodésiques) bornées par deux géodésiques a_1 et a_2 qui font avec a un angle π [fig.1 ou 2]. Ces géodésiques s'envoient par f en des géodésiques de (M', ρ_2) ayant toutes le segment $f(a)$ en commun. Par conséquent, les angles que fait $f(a)$ avec $f(a_1)$ et $f(a_2)$ sont tous supérieurs à π. S'il existe une géodésique g dans (M', ρ_2) prolongeant $f(a)$ au point $f(s)$ et n'appartenant pas à l'image du faisceau de géodésiques borné par a_1 et a_2 alors, d'après la proposition précédente, $f^{-1}(g)$ est une géodésique sur (M, ρ_1) prolongeant a et n'appartenant pas au faisceau borné par a_1 et a_2. Ceci est évidemment impossible. Par conséquent, les angles au point $f(s)$ entre $f(a)$ et $f(a_1)$ et entre $f(a)$ et $f(a_2)$ sont égaux à π. L'image d'un secteur d'angle π est alors un secteur d'angle π.

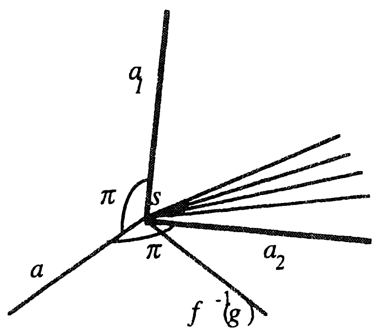


Figure 2

Lemme 2. *Si l'angle conique θ_s de s est compris entre $k\pi$ et $(k+1)\pi$, pour un certain entier naturel k, alors l'angle conique $\theta_{f(s)}$ de $f(s)$ est aussi compris entre $k\pi$ et $(k+1)\pi$.*

DÉMONSTRATION: Supposons que $\theta_s = k\pi + \alpha$ avec $0 \leq \alpha < \pi$. Traçons un secteur basé en s d'angle α délimité par deux segments géodésiques notés a_0 et a_k. Subdivisons le secteur restant en k secteurs chacun d'angle π. Chaque secteur est délimité par deux segments géodésiques notés a_i et a_{i+1}, l'indice i varie entre 0 et k [fig. 3]. Supposons que $\theta_{f(s)}$ est strictement inférieur à $k\pi$. Supposons enfin que $(m-1)\pi <$

$\theta_{f(s)} < m\pi$ avec $m <$ k. Le lemme précédent entraîne que l'image des $(m-1)$ premiers secteurs basés en s et délimités par les segments a_0 et a_{m-1} est un secteur basés en f(s) d'angle $(m-1)\pi$ et délimité par $f(a_0)$ et $f(a_{m-1})$. Les secteurs restants basés en s s'envoient tous sur un secteur basé en $f(s)$ d'angle inférieur à π qui ne peut pas contenir de géodésiques passant par $f(s)$, voir les parties hachurées de la Figure 3. Ceci est évidemment impossible et par conséquent $\theta_{f(s)} \geq k\pi$. Puisque l'image inverse de l'application f est une correspondance géodésique, le même raisonnement entraîne que si $\theta_{f(s)}$ est supérieur à $(k+1)\pi$ alors $\theta_{f^{-1}(f(s))} = \theta_s$ est aussi supérieur à $(k+1)\pi$. Ce qui n'est pas possible. D'où l'on déduit le lemme.

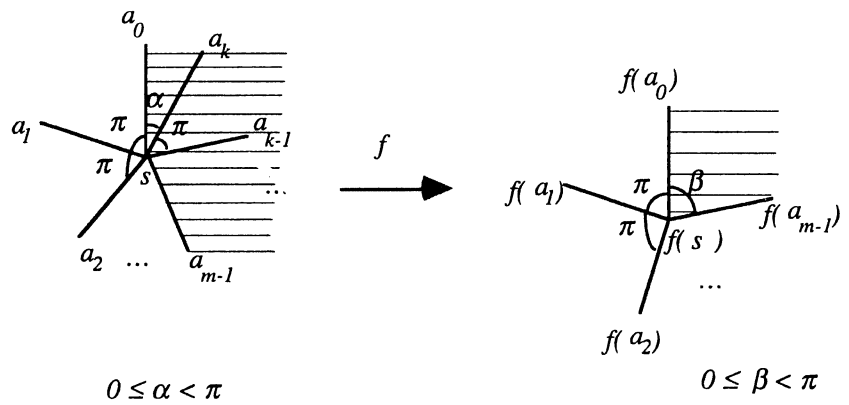


Figure 3

DÉMONSTRATION DE LA PROPOSITION. Si l'angle conique θ_s est un multiple de π, le Lemme 2 implique que $\theta_{f(s)} = \theta_s$. Supposons maintenant que θ_s n'est pas un multiple de π et que l'on a $\theta_{f(s)} > \theta_s$ et $(k-1)\pi < \theta_s < k\pi$ pour un certain entier naturel k. Cet encadrement entraîne, d'après ce qui précède, que l'on a $(k-1)\pi < \theta_{f(s)} < k\pi$. Notons pour simplifier $\alpha = k\pi - \theta_s$ et $\beta = k\pi - \theta_{f(s)}$. Puisque l'image de θ_s par f est $\theta_{f(s)}$ et f préserve les angles multiples de π, en chaque singularité, l'image de tout secteur d'angle α est un secteur d'angle β. Notons ceci par $f(\alpha) = \beta$. Par conséquent, l'image de deux secteurs voisins d'angle α chacun sont deux secteurs voisins d'angle β chacun, d'où $f(2\alpha) = 2\beta$. Par récurrence, on montre que $f(n\alpha) = n\beta$ pour tout entier naturel n. L'inégalité $\theta_{f(s)} > \theta_s$ implique que $\beta < \alpha$. Puisque

$\alpha - \beta$ est différent de zéro, la différence $n\alpha - n\beta$ tend vers l'infini quand n croît. Soit n' le premier entier naturel tel que :

$$n'\,\alpha = m\pi + \alpha' \qquad \text{et} \qquad n'\,\beta = m\pi + \beta'\,,$$

où $m \in \mathbb{N}$, $\pi \leq \alpha'$ et $0 \leq \beta' < \pi$.

Puisque $f(n\alpha) = n\beta$ et $f(n\pi) = n\pi$ quel que soit l'entier naturel n, on a $f(\alpha') = \beta'$. Ceci est impossible car le secteur d'angle α' contient des géodésiques passant par son sommet s, tandis que le secteur image d'angle β' ne contient pas de géodésiques passant par son sommet $f(s)$ [fig. 4]. Par conséquent, on obtient $\theta_{f(s)} \leq \theta_s$.

Puisque f^{-1} est une correspondance géodésique, la même méthode appliquée à f^{-1} donne l'inégalité $\theta_{f(s)} \geq \theta_{f^{-1}(f(s))=s}$. D'où l'on déduit la proposition.

Corollaire 2. *Soient s un point singulier de la métrique ρ_1 et θ_s l'angle conique en s. Supposons que θ_s est rationnellement indépendant de π. L'image par f de tout secteur d'ouverture α' basé en s est un secteur d'ouverture α' basé en $f(s)$. Autrement dit, f respecte les angles en s.*

DÉMONSTRATION. Supposons que $\theta_s = k\pi + \alpha$ pour un certain entier naturel k et $0 < \alpha < \pi$. Un simple calcul montre que l'angle α est rationnellement indépendant avec θ_s. Un résultat de topologie implique que les multiples relatifs de α modulo θ_s forment un ensemble dense dans l'intervalle $[0, \theta_s]$. D'après la proposition précédente, on a $f(n\alpha) = n\alpha$ pour tout entier n. Chaque angle α' peut être approché par une suite de multiples de α modulo θ_s . L'image de cette suite par f est elle même. Par conséquent, elle tends vers α'. Ainsi $f(\alpha') = \alpha'$.

Les deux corollaires suivants sont immédiats.

Corollaire 3. *L'image par f d'un cylindre euclidien sur (M, ρ_1) est un cylindre euclidien sur (M', ρ_2).*

Corollaire 4. *Si $\{s_1, \ldots, s_k\}$ est l'ensemble des singularités de ρ_1, alors l'ensemble des singularités de ρ_2 est exactement $\{f(s_1), \ldots, f(s_k)\}$*

et, pour chaque $i \in \{1, \ldots, k\}$, le point singulier $f(s_i)$ a le même angle conique que s_i.

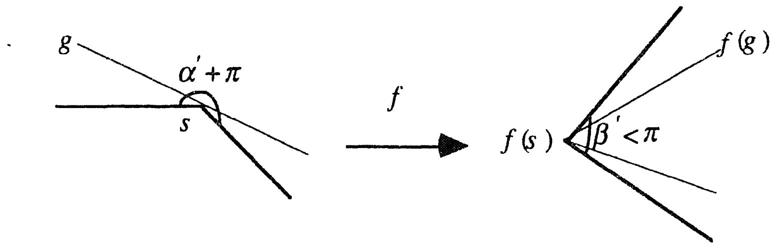


Figure 4.

Théorème. *Soient M et M' deux surfaces compactes sans bord de même genre $g \geq 1$. Soient ρ_1 et ρ_2 deux métriques plates singulières définies respectivement sur M et M' telles que tous les angles coniques des singularités des deux métriques sont strictement supérieure à 2π. Toute correspondance geodésique conforme $f : (M, \rho_1) \longrightarrow (M', \rho_2)$ est une homothétie.*

DÉMONSTRATION. Remarquons d'abord que l'hypothèse du théorème entraîne que dans le cas $g = 1$ les métriques plates ρ_1 et ρ_2 ne sont pas singulières. Notons $\overline{\rho}_2$ la métrique réciproque de ρ_2 par f. Les métriques $\overline{\rho}_2$ et ρ_2 sont isométriques. Ce qui implique que $\overline{\rho}_2$ est une métrique plate singulière sur M. Puisque f est conforme, la métrique $\overline{\rho}_2$ est conforme à ρ_1 (dans le sens où il existe une fonction positive h telle que $\overline{\rho}_2 = h\rho_1$) et a les mêmes géodésiques (non paramétrées) que ρ_1. D'après le corollaire précédent, si M est de genre $g > 1$ alors ρ_1 et $\overline{\rho}_2$ ont les mêmes points singuliers avec les mêmes angles coniques respectifs. Comme les métriques ρ_1, $\overline{\rho}_2$ sont conformes et ont mêmes singularités avec les mêmes angles coniques respectifs (quand $g > 1$), alors, d'après le Théorème de Troyanov [Tro] de classification des surfaces euclidiennes à singularités coniques, la fonction h est une constante c. On a alors, $\overline{\rho}_2 = c\rho_1$. Il en résulte que l'application f est une homothétie.

REMARQUE. La condition du théorème demandant que ρ_1 et $\overline{\rho}_2$ soient conformes est essentielle au moins en genre 1. Dans ce cas si la correspondance géodésique n'est pas conforme, le théorème est faux. En effet, un tore T muni des deux métriques plates $\rho_1 = dx^2 + dy^2$ et $\rho_2 = dx^2 + 2\,dy^2$ admet l'identité comme correspondance géodésique alors que les métriques ρ_1 et ρ_2 ne sont pas proportionnelles.

Remerciements. Je remercie Marc Troyanov pour ses remarques et suggestions sur ce travail.

Références.

[Bon] Bonahon, F., Surfaces with the same marked lengh spectrum. *Topology and its applications* **50** (1993), 55-62.

[Mes] Mesmoudi, M., M., Métriques plates singulières et feuilletages mesurés sur les surfaces, Thèse de doctorat, prépublication de l'IRMA, 1994.

[Mon] Montesinos, A. J., Sur les variétés riemanniennes à flot géodésique topologiquement transitif. *Rev. Mat. Iberoamericana* **6** (1990), 109-124.

[Tro] Troyanov, M., Les surfaces euclidiennes à singularités coniques. *Enseig. Math.* **32** (1986), 79-94.

Recibido: 14 de noviembre de 1.994
Revisado: 22 de febrero de 1.995

Mohammed Mostefa Mesmoudi
Institut de Recherche Mathématique Avancée
Université Louis Pasteur et C.N.R.S.
7, rue René Descates 67084
Strasbourg Cedex, FRANCE
mesmoudi@math.u-strasbg.fr

Revista Matemática Iberoamericana
Vol. 12, N.º 1, 1996

The range of Toeplitz operators on the ball

Boris Korenblum and John E. McCarthy

0. Introduction.

Let B_d be the unit ball in $\mathbb{C}^d$, S_d be the boundary of B_d, and σ_d be normalized Lebesgue measure on S_d. The Hardy space $H^2(B_d)$ is the closure in $L^2(S_d, \sigma_d)$ of the analytic polynomials. The space $H^\infty(B_d)$ of bounded functions in $H^2(B_d)$ is precisely the space of functions that are radial limits (σ_d-almost everywhere) of bounded analytic functions on B_d. Let P denote the orthogonal projection from $L^2(S_d, \sigma_d)$ onto $H^2(B_d)$. If m is in $H^\infty(B_d)$, the co-analytic Toeplitz operator $T_{\bar{m}}^{H^2(B_d)}$ is defined by

$$T_{\bar{m}}^{H^2(B_d)} f = P\bar{m}f\,.$$

The purpose of this paper is to study the common range of all the co-analytic Toeplitz operators $T_{\bar{m}}^{H^2(B_d)}$.

For the case $d = 1$, it was shown in [2] that a function f is in the range of every non-zero co-analytic Toeplitz operator $T_{\bar{m}}^{H^2(B_1)}$ if and only if the Taylor coefficients of f at zero satisfy

$$\hat{f}(n) = O(e^{-c\sqrt{n}})$$

for some $c > 0$. It was also shown that, for the case $d > 1$, if the Taylor coefficients of some f in $H^2(B_d)$ satisfy

$$\hat{f}(\alpha) = O(e^{-c|\alpha|^{d/(d+1)}})$$

for some $c > 0$, then the function f will be in the range of every non-zero co-analytic Toeplitz operator on $H^2(B_d)$. It was asked if this sufficient condition were also necessary. Our main theorem answers this question in the negative:

Theorem 1. *Let $f(z_1, \ldots, z_d) = f_1(z_1) = \sum_{n=0}^{\infty} a_n z_1^n$, let $\varepsilon > 0$, and suppose that $a_n = O(e^{-cn^{1/2+\varepsilon}})$ for some $c > 0$. Then f is in the range of the Toeplitz operator $T_{\bar{m}}^{H^2(B_d)}$ for every non-zero m in $H^\infty(B_d)$.*

The exponent $n^{1/2+\varepsilon}$ is not optimal -using results of [3] it can be improved to $\sqrt{n}\log n$. We do not know what necessary and sufficient conditions are for a function to be in the range of every non-zero co-analytic Toeplitz operator.

In dimension $d = 1$, Szegö's theorem [9] states that a necessary and sufficient condition for a positive bounded function g on the circle to be the modulus of a non-zero function in $H^\infty(B_1)$ is

$$\int_{S_d} log(g)\, d\sigma_d > -\infty\,. \tag{0.1}$$

For $d > 1$, condition (0.1) is necessary and sufficient for g to be the modulus of a function in the larger Nevanlinna class $N(B_d)$, consisting of those holomorphic functions f on the ball for which

$$T(f,1) := \sup_{0<r<1} \int_{S_d} \log^+ |f(r\zeta)|\, d\sigma_d(\zeta) < \infty$$

[7, Theorem 10.11]. It is no longer sufficient, however, for g to be the modulus of a bounded analytic function, because the function

$$\zeta \mapsto \operatorname*{ess\,sup}_{-\pi\le\theta\le\pi} |m(e^{i\theta}\zeta)|$$

must be lower semi-continuous on S_d if m is in $H^\infty(B_d)$ [7]. In [7, Theorem 12.5], Rudin proves that if g is log-integrable, and there exists some non-zero f in $H^\infty(B_d)$ with $g \ge |f|$ almost everywhere and $g/|f|$ lower semi-continuous, then there does exist m in $H^\infty(B_d)$ with $g = |m|$ almost everywhere. We show

Theorem 2. *Let $d \ge 2$. There is a non-negative continuous function g on S_d, with $\int_{S_d} log(g)\, d\sigma_d > -\infty$, and which vanishes at only one* *point, but such that for no non-zero function m in $H^\infty(B_d)$ is $|m| \leq g$ almost everywhere with respect to σ_d.*

This answers Question 15 of [7] in the negative.

When the original version of this paper was circulated in preprint form (see the announcement in [1]), H. Alexander (private communication) produced a very simple constructive example of a function g satisfying the conclusion of Theorem 2, obviating the complicated construction in our proof. However, as we think our construction may be of some use in solving the problem of characterising exactly which functions are moduli of $H^\infty(B_d)$ functions, we elected to retain the proof of Theorem 2 in this paper.

1. Preliminary Lemmata.

We need to know explicitly the projection from $L^2(B_d)$ onto $H^2(B_d)$. Let $\alpha = (\alpha_1, \ldots, \alpha_d)$ be a multi-index and $\zeta = (z_1, \ldots, z_d)$ a point in $\mathbb{C}^d$. The function ζ^α then maps ζ to $z_1^{\alpha_1} \cdots z_d^{\alpha_d}$. The notation $|\alpha|$ stands for $\alpha_1 + \cdots + \alpha_d$, and $\alpha! = \alpha_1! \cdots \alpha_d!$.

Lemma 1.1.

$$\int_{S_d} \zeta^\alpha \overline{\zeta^\beta}\, d\sigma_d = \delta_{\alpha,\beta} \frac{(d-1)!\,\alpha!}{(d-1+|\alpha|)!}\,. \tag{1.2}$$

Moreover, if $P_{H^2(B_d)}$ denotes the projection from $L^2(\sigma_d)$ onto $H^2(B_d)$, then

$$\begin{aligned} &P_{H^2(B_d)} |z_2^{\alpha_2}|^2 \cdots |z_d^{\alpha_d}|^2 \overline{z_1^j} z_1^i \\ &= \begin{cases} 0\,, & \text{if } i < j\,, \\ \dfrac{(d-1+i-j)!\, i!\, \alpha_2! \cdots \alpha_d!}{(i-j)!\,(d-1+i+\alpha_2+\cdots+\alpha_d)!} z_1^{i-j}\,, & \text{if } i \geq j\,. \end{cases} \end{aligned} \tag{1.3}$$

PROOF. Formula (1.2) is proved in [6]. The expression on the left-hand side of (1.3) is orthogonal to every monomial except z_1^{i-j}; taking inner products gives the constant.

We need to consider co-analytic Toeplitz operators on different spaces. If μ is a compactly supported measure on $\mathbb{C}^d$, let $P^2(\mu)$ denote the closure of the polynomials in $L^2(\mu)$, and let $P_{P^2(\mu)}$ denote the orthogonal projection from $L^2(\mu)$ onto $P^2(\mu)$. If m is a bounded analytic function on the support of μ, the co-analytic Toeplitz operator $T_{\bar{m}}^{P^2(\mu)}$ is defined by

$$T_{\bar{m}}^{P^2(\mu)} f \;=\; P_{P^2(\mu)}\, \bar{m} f\,.$$

When μ is σ_d, the space $P^2(\mu)$ is the Hardy space $H^2(B_d)$, and we recover our original definition.

In order to transfer information about co-analytic Toeplitz operators with the same symbol on different spaces, we use the following lemma, whose proof is immediate:

Lemma 1.4. *Let $\mathcal{H}$ be a Hilbert space of holomorphic functions on B_d in which the monomials are mutually orthogonal. Let $m(z_1, \ldots, z_d) = \sum_{\beta \in \mathbb{N}^d} b_\beta \zeta^\beta$. Then*

$$T_{\bar{m}}^{\mathcal{H}} \frac{\zeta^\alpha}{\|\zeta^\alpha\|_{\mathcal{H}}^2} = \sum_{\beta \leq \alpha} \bar{b}_{\alpha-\beta} \frac{\zeta^\beta}{\|\zeta^\beta\|_{\mathcal{H}}^2}\,. \tag{1.5}$$

This lemma also allows us to define Toeplitz operators with an unbounded conjugate analytic symbol. The formal definition (1.5) defines an upper triangular operator, with respect to the orthonormal basis of normalized monomials. It therefore has a domain which contains all the polynomials; we extend its domain to include all functions on which $T_{\bar{m}}$, thought of as a formal operator on the power series, gives a power series whose coefficients are the Taylor coefficients of some function in $\mathcal{H}$.

Lemma 1.6. *Let g be in the Nevanlinna class $N(B_1)$, with $g(0) \neq 0$, and $1 \leq \alpha < 2$. Then*

$$\int_{B_1} (\log^- |g|)^\alpha \, dA \leq K\,,$$

where K is some constant depending only on $T(g,1)$, $|g(0)|$ and α.

PROOF. The proof is in two parts. First we prove it for g zero-free, then we prove it for g a Blaschke product. As $\log g$ is the sum of the logarithms of two such terms, this suffices.

a) Suppose g has no zeroes in B_1, and without loss of generality assume $\|g\|_\infty < 1$. Then there is a singular measure μ_s such that, for any $0 < r < 1$,

$$\log^- |g(re^{i\theta})| = \frac{1}{2\pi} \int_0^{2\pi} P_{re^{i\theta}}(e^{i\phi}) (\log^- |g(e^{i\phi})| \, d\phi + d\mu_s(\phi)),$$

where $P_{re^{i\theta}}(e^{i\phi})$ is the Poisson kernel. Therefore

$$\begin{aligned}
&\int_{B_1} (\log^- |g(re^{i\theta})|)^\alpha \, r \, dr \, d\theta \\
&= \int_0^1 r \, dr \int_0^{2\pi} d\theta \Big(\int_0^{2\pi} P_{re^{i\theta}}(e^{i\phi}) (\log^- |g(e^{i\phi})| \, d\phi + d\mu_s(\phi)) \Big)^\alpha \\
&\le \Big(\int_0^{2\pi} \Big(\int_0^1 r \, dr \int_0^{2\pi} d\theta \, (P_{re^{i\theta}}(e^{i\phi}))^\alpha \Big)^{1/\alpha} (\log^- |g(e^{i\phi})| \, d\phi + d\mu_s(\phi)) \Big)^\alpha
\end{aligned}$$

by Minkowski's inequality. As the $L^\alpha(A)$ norm of the Poisson kernel for a fixed boundary point is at most $(8/(2-\alpha))^{1/\alpha}$, we get

$$\int_{B_1} (\log^- |g(re^{i\theta})|)^\alpha \, r \, dr \, d\theta \le \frac{8}{2-\alpha} \Big(\frac{1}{\log |g(0)|} \Big)^\alpha .$$

b) Suppose now g is a Blaschke product with zero-set $\{w_n\}$. Then

$$\begin{aligned}
&\Big(\int_{B_1} |\log |g(z)||^\alpha \, dA(z) \Big)^{1/\alpha} \\
&= \Big(\int_{B_1} \Big(\sum_{n=0}^{\infty} \log \Big| \frac{1 - \bar{w}_n z}{z - w_n} \Big| \Big)^\alpha dA(z) \Big)^{1/\alpha} \qquad (1.7)\\
&\le \sum_{n=0}^{\infty} \Big(\int_{B_1} \Big(\log \Big| \frac{1 - \bar{w}_n z}{z - w_n} \Big| \Big)^\alpha dA(z) \Big)^{1/\alpha} .
\end{aligned}$$

Now let us estimate

$$\int_{B_1} \Big(\log \Big| \frac{1 - \bar{w}_n z}{z - w_n} \Big| \Big)^\alpha dA(z).$$

The terms for $|w_n| \le 1/2$ are dominated by $T(g,1) + \log^- |g(0)|$, by Jensen's formula. For convenience, assume w is positive, and make the change of variables $\zeta = re^{i\theta} = (z-w)/(1-wz)$. Then

$$
\begin{aligned}
&\int_{B_1} \left(\log\left|\frac{1-wz}{z-w}\right|\right)^\alpha dA(z) \\
(1.8) \qquad &= \frac{1}{\pi}\int_{B_1} \left(\log\frac{1}{r}\right)^\alpha \frac{(1-w^2)^2}{|1-wre^{i\theta}|^4}\, r\,dr\,d\theta \\
&= 2\,(1-w^2)^2 \int_0^1 \left(\log\frac{1}{r}\right)^\alpha \frac{1+(rw)^2}{(1-(rw)^2)^3}\, r\,dr\,.
\end{aligned}
$$

Break the integral (1.8) into two pieces: from 0 to $1/e$, where the integrand is bounded by some constant C_1 independent of w, and from $1/e$ to 1. For the latter integral, use the inequality $\log(1/r) \leq 2\,(1-wr)$. One gets that (1.8) is bounded by $C_2\,(1-w)^\alpha/(2-\alpha)$, where C_2 depends on neither w nor α. So (1.7) is dominated by $(C_2/(2-\alpha))^{1/\alpha}\sum_{n=0}^{\infty}(1-|w_n|)$, and Jensen's formula again means we can dominate everything by a constant depending on α, $\log|g(0)|$ and $T(g,1)$.

Let A^{-n} consist of all holomorphic functions m in the unit disk that satisfy $|m(z)| = O((1-|z|)^{-n})$. The space A^0 is $H^\infty(B_1)$.

Lemma 1.9. *Let f be in A^{-n} for some n, and $0<\alpha<2$. Then*

$$\int_{B_1} (\log^-|f|)^\alpha\, dA < \infty\,.$$

PROOF. We can assume that $f(0) \neq 0$. As f need not be in $N(B_1)$ we cannot apply Lemma 1.6 directly; but f is in the Nevanlinna class of certain smaller domains that touch the boundary of B_1 at only one point, and we shall average over these.

Fix p strictly between 1 and $2/\alpha$, let $a = \alpha p < 2$, let $q = p/(p-1)$ and let $N > q$. Let D_1 be a smoothly bounded convex domain inside the disk, containing $\{z : |z| < 1/2\}$, whose closure touches the unit circle only at 1, and which has a high degree of tangency at 1: let the boundary of D_1 be $\{\rho(\theta)e^{i\theta} : -\pi \leq \theta \leq \pi\}$, and assume $1-\rho(\theta) \sim |\theta|^N$. For any other point $\zeta = e^{i\theta_0}$ on the boundary of the unit disk, let $D_\zeta = e^{i\theta_0}D_1$.

Let ψ_ζ be the Riemann map of D_ζ onto B_1 that takes 0 to 0 and ζ to ζ. As the boundary of D_ζ is smooth, it follows from the Kellog-Warschawski theorem (see *e.g.* [5]) that ψ_ζ and its derivatives extend continuously to the closure of D_ζ, so distances before and after the conformal mapping are comparable.

If $r < 1/N$, then f is in $H^r(D_\zeta)$, and $\sup_{\zeta\in S_1} \|f \circ \psi_\zeta^{-1}\|_{H^r} < \infty$. Thus

$$\int_{D_{e^{i\theta}}} (\log^- |f|)^a \, dA \leq C\,, \qquad \text{for all } e^{i\theta}\,.$$

Integrating with respect to θ and changing the order of integration yields

$$\int_{B_1} (\log^- |f(re^{i\phi})|)^a \,(1-r)^{1/N}\, r\, dr\, d\phi < \infty\,.$$

Now

$$\int_{B_1} (\log^- |f|)^\alpha \, dA$$
$$\leq \Big(\int_{B_1} (\log^- |f|)^{\alpha p}\,(1-r)^{p/N}\, dA\Big)^{1/p} \Big(\int_{B_1} (1-r)^{-q/N}\, dA\Big)^{1/q} < \infty\,.$$

Let μ_n be the measure on the unit disk given by $d\mu_n(z) = \pi^{-1}(1-|z|^2)^n\, dA(z)$, and let $\mathcal{H}_n$ be $P^2(\mu_n)$. It is routine to verify that in $\mathcal{H}_n$ the monomials are mutually orthogonal, and

$$\|z^k\|^2_{\mathcal{H}_n} = \frac{n!}{(k+1)\cdots(k+n+1)}\,.$$

The space $\mathcal{H}_0$ is the usual Bergman space for the disk. The following lemma is proved in [3] (in fact a slightly sharper form is proved). We include the following proof, which is sufficient for our purposes, for completeness:

Lemma 1.10. *Let $n \geq 0$, and m be a function in A^{-n}, not identically zero. Suppose $f(z) = \sum_{k=0}^\infty a_k z^k$ where $a_k = O(e^{-ck^{1/2+\varepsilon}})$ for some ε and c greater than 0. Then for any $s \geq 2n$ there exists g in $\mathcal{H}_s$ such that $T_{\bar m}^{\mathcal{H}_s} g = f$.*

PROOF. First, observe that $f = T_{\bar m}^{\mathcal{H}_s} g$ for some g if and only if there is a constant C such that for all polynomials p

$$|\langle p, f\rangle_{\mathcal{H}_s}| \leq C\sqrt{\int |p|^2|m|^2\, d\mu_s}\,.$$

So it is sufficient to prove that

$$\Big|\sum_{k=0}^{\infty} \bar{a}_k\, \hat{p}(k) \frac{1}{(k+1)\cdots(k+s+1)}\Big| \leq C \sqrt{\int |p|^2 |m|^2 \, d\mu_s}\,.$$

This in turn will follow from the Banach-Steinhaus theorem if we can show that for any function h in $P^2(|m|^2\mu_s)$,

$$\hat{h}(k) = O(e^{ck^{1/2+\epsilon}})\,. \tag{1.11}$$

Now Stoll showed in [8] that if h satisfies

$$\int_{B_1} (\log^+ |h|)^\alpha \, dA < \infty$$

for some $\alpha > 0$ then $\hat{h}(k) = O(e^{o(2/(2+\alpha))})$. We can assume ε is small, and take $\alpha = (2-4\varepsilon)/(1+2\varepsilon)$. As h is in $P^2(|m|^2\mu_s)$, $h(z)m(z)(1-|z|^2)^{s/2} := k(z)$ is in $L^2(dA)$, and

$$\log^+|h| \leq \log^+|k| + \log^-|(1-|z|^2)^{s/2}| + \log^-|m|\,.$$

The first two terms on the right are clearly integrable to the α^{th} power, and so is the third by Lemma 1.9; therefore h satisfies (1.11) as desired.

We want to be able to restrict functions in the ball to planes and factor out zeros without losing control of the size of the function; the next lemma allows us to do this.

Lemma 1.12. *Let m be holomorphic on B_d and satisfy*

$$|m(z_1,\ldots,z_d)| \leq C\,(1-\sqrt{|z_1|^2+\ldots|z_d|^2})^{-s}\,.$$

Suppose also that

$$m(z_1,\ldots,z_d) = z_d^t\, m_2(z_1,\ldots,z_d) + z_d^{t+1}\, m_3(z_1,\ldots,z_d)\,,$$

where m_2 and m_3 are analytic. Let

$$m_1(z_1,\ldots,z_{d-1}) = m_2(z_1,\ldots,z_{d-1},0)\,.$$

Then

$$|m_1(z_1,\ldots,z_{d-1})| \leq (3\,d)^{s+t}\, C\,(1-\sqrt{|z_1|^2+\ldots|z_{d-1}|^2})^{-(s+t)}\,.$$

PROOF. Let $(z_1, \ldots, z_{d-1})$ be in B_{d-1}, and let

$$\varepsilon = \frac{1}{3d}\left(1 - \sqrt{|z_1|^2 + \cdots + |z_{d-1}|^2}\right).$$

Then the polydisk centered at $(z_1, \ldots, z_{d-1}, 0)$ with multi-radius $(\varepsilon, \ldots, \varepsilon)$ is contained in $(1-\varepsilon)\, B_d$. Integrating on the distinguished boundary of the polydisk we get

$$\begin{aligned} |m_1(z_1, \ldots, z_{d-1})| &= |m_2(z_1, \ldots, z_{d-1}, 0)| \\ &= \Big| \int_{(z_1,\ldots,z_{d-1},0)+\varepsilon T^d} \frac{m(\zeta_1, \ldots, \zeta_d)}{\zeta_3^t} \Big| \leq \frac{C}{\varepsilon^{s+t}} . \end{aligned}$$

2. Common Range of $T_{\bar{m}}$.

We can now prove that a function that depends on only one variable is in the range of every $T_{\bar{m}}^{H^2(B_d)}$ if its Taylor coefficients decay like $e^{-ck^{1/2+\epsilon}}$.

Theorem 1. *Let $f(z_1, \ldots, z_d) = f_1(z_1) = \sum_{n=0}^{\infty} a_n z_1^n$, let $\varepsilon > 0$, and suppose that $a_n = O(e^{-cn^{1/2+\epsilon}})$ for some $c > 0$. Then f is in the range of the Toeplitz operator $T_{\bar{m}}^{H^2(B_d)}$ for every non-zero m in $H^\infty(B_d)$.*

PROOF. For $d = 1$, this is proved (without the ε) in [2], so assume $d \geq 2$. Fix m in $H^\infty(B_d)$;

$$m(z_1, \ldots, z_d) = \sum_{i_1,\ldots,i_d=0}^{\infty} b_{i_1,\ldots,i_d}\, z_1^{i_1} \ldots z_d^{i_d} .$$

Let

$$S = \{(i_2, \ldots, i_d) : \text{ for some } i_1 ,\ b_{i_1,\ldots,i_d} \neq 0\} .$$

Define

$$t_d = \inf\{i_d : \text{ for some } i_2, \ldots, i_{d-1} ,\ (i_2, \ldots, i_{d-1}, i_d) \in S\} ,$$

and define t_k inductively by

$$t_k = \inf\{i_k : \text{ for some } i_2, \ldots, i_{k-1},\ (i_2, \ldots, i_{k-1}, i_k, t_{k+1}, \ldots, t_d) \in S\} .$$

Let $n = t_2 + \cdots + t_d$.

Case a) $n = 0$. Then the function

$$m_1(z_1) = m(z_1, 0, \ldots, 0)$$

is not identically zero, and is in $H^\infty(B_1)$. By Lemma 1.1,

$$T_m^{H^2(B_d)} z_1^i = \sum_j \bar{b}_{j,0,\ldots,0} \frac{(i-j+1)\cdots(i-j+d-1)}{(i+1)\cdots(i+d-1)} z_1^{i-l}.$$

So by Lemma 1.4, if one can solve the equation

$$T_{\bar{m}_1}^{\mathcal{H}_{d-2}} g_1 = f_1 \tag{2.1}$$

for some g_1 in $\mathcal{H}_{d-2}$, then $g(z_1, \ldots, z_d) = g_1(z_1)$ solves

$$T_{\bar{m}}^{H^2(B_d)} g = f,$$

and, by equation (1.2), $\|g\|_{H^2(B_d)} = \sqrt{(d-1)!}\ \|g_1\|_{\mathcal{H}_{d-1}} < \infty$. By Lemma (1.10), equation (2.1) has a solution.

Case b) $n > 0$. One can decompose m as

$$m(z_1, \ldots, z_d) = z_2^{t_2} \cdots z_d^{t_d} m_2(z_1, \ldots, z_d) + m_3(z_1, \ldots, z_d);$$

where each term in the expansion of m_3 is divisible by some $z_k^{t_k+1}$. Applying Lemma 1.12 inductively, $m_1(z) = m_2(z, 0, \ldots, 0)$ is in A^{-n}, and by the choice of $t_2, \ldots, t_d$, it is not identically zero. Consider the function

$$f_2(z) = \sum_{k=0}^\infty a_k (k+d)(k+d+1)\cdots(k+dn+1) z^k.$$

As $d \geq 2$, we can apply Lemma 1.10 with $s = dn$, so there is

$$g_2(z) = \sum_{k=0}^\infty \gamma_k (k+1)(k+2)\cdots(k+dn+1) z^k$$

in $\mathcal{H}_{dn}$ with

$$T_{\bar{m}_1}^{\mathcal{H}_{dn}} g_2 = f_2. \tag{2.2}$$

Define g by

$$g(z_1, \dots, z_d) = \frac{1}{t_2! \cdots t_d!} z_2^{t_2} \cdots z_d^{t_d} \cdot \sum_{k=0}^{\infty} \gamma_k (k+1)(k+2) \cdots (k+n+d-1) z_1^k .$$

The function g is in $H^2(B_d)$ because

$$\begin{aligned} \|g\|^2_{H^2(B_d)} &= \frac{(d-1)!}{t_2! \cdots t_d!} \sum_{k=0}^{\infty} |\gamma_k|^2 (k+1) \cdots (k+n+d-1) \\ &\le \frac{(d-1)!}{t_2! \cdots t_d!} \sum_{k=0}^{\infty} |\gamma_k|^2 (k+1) \cdots (k+dn+1) \\ &= \frac{(d-1)!}{t_2! \cdots t_d!} \|g_2\|^2_{\mathcal{H}_{(d-1)n}} < \infty . \end{aligned}$$

Moreover

$$T_{\bar{m}}^{H^2(B_d)} g = T^{H^2(B_d)}_{\bar{z}_2^{t_d} \cdots \bar{z}_d^{t_d} \overline{m_1(z_1)}} g$$

is a function of z_1 only; it is, in fact, f. For if $T_{\bar{m}}^{H^2(B_d)} g = \sum_{k=0}^{\infty} e_k z_1^k$, and $m_1(z) = \sum_{k=0}^{\infty} c_k z^k$, then taking the inner product with z_1^j we get

$$\begin{aligned} \frac{(d-1)!}{(j+1)\cdots(j+d-1)} e_j &= \langle T_{\bar{m}}^{H^2(B_d)} g, z_1^j \rangle_{H^2(B_d)} \\ &= \langle g, z_2^{t_2} \cdots z_d^{t_d} m_1 z_1^j \rangle_{H^2(B_d)} \\ &= (d-1)! \sum_{k=j}^{\infty} \gamma_k \bar{c}_{k-j} . \end{aligned} \tag{2.3}$$

Taking the inner product with z^j in equation (2.2), we get

$$\begin{aligned} \frac{1}{(j+1)\cdots(j+d-1)} a_j &= \langle T_{\bar{m}_1}^{\mathcal{H}_{dn}} g_2, z^j \rangle_{\mathcal{H}_{dn}} \\ &= \langle g_2, m_1 z^j \rangle_{\mathcal{H}_{dn}} \\ &= \sum_{k=j}^{\infty} \gamma_k \bar{c}_{k-j} . \end{aligned} \tag{2.4}$$

Comparing equations (2.3) and (2.4), we see that $T_{\bar{m}}^{H^2(B_d)} g = f$, as desired.

3. Boundary moduli.

Define $F_{c,w}$ by

$$F_{c,w}(z) = \exp\left(c\frac{1-|w|^2}{(1-\langle z,w\rangle)^{d+1}}\right). \tag{3.1}$$

We need the following two results. The first was proved by Drewnowski; a proof is given in [4, Lemma 3.2].

Lemma 3.2 (Drewnowski).

$$\lim_{c\to 0}\sup_{w\in B_d}\int_{S_d}\log(1+|c\,F_{c,w}|)\,d\sigma_d = 0\,.$$

The second result, due to Nawrocki, estimates the growth of the Taylor coefficients of $F_{c,w}$. We are interested in $w = re_1 = (r,0,\ldots,0)$; in this case all the Taylor coefficients of F_{c,re_1} are positive, and the following follows easily from the proof of [4, Lemma 3.3]:

Lemma 3.3 (Nawrocki). *For each $c>0$ there exists $\varepsilon>0$ such that*

$$\inf_{i\in\mathbb{N}}\sup_{0<r<1}\sqrt{\frac{(d-1+i)!}{(d-1)!\,i!}}\;\hat{F}_{c,re_1}(i,0,\ldots,0)\,e^{-\varepsilon i^{d/(d+1)}} > 0\,.$$

We can now use our knowledge of the common range of co-analytic Toeplitz operators to prove:

Theorem 2. *Let $d\geq 2$. There is a continuous non-negative function g on S_d, vanishing only at the point e_1, and satisfying $\int_{S_d}log(g)d\sigma_d > -\infty$, with the property that the only function m in $H^\infty(B_d)$ with $|m|\leq g$ almost everywhere with respect to σ_d is the zero function.*

PROOF. Let

$$V_n = \left\{\zeta\in S_d:\ |\zeta-e_1|\geq\frac{1}{n}\right\}.$$

By Lemma 3.3, for any sequence c_n tending to zero, one can choose i_n and r_n such that

$$\hat{F}_{c_n,r_ne_1}(i_n,0,\ldots,0) > \frac{n}{c_n}\,e^{(i_n)^{4/7}} \tag{3.4}$$

(because $4/7 < d/(d+1)$). Moreover, by passing to a subsequence, one can assume that

$$\sup_{\zeta \in V_n} c_n \, |F_{c_n, r_n e_1}(\zeta)| \leq \frac{1}{2^n}, \tag{3.5}$$

because $\zeta \in V_n$ implies that

$$|1 - \langle \zeta, r_n e_1 \rangle| \geq \frac{1}{2n^2},$$

and that

$$\int_{S_d} \log(1 + |c_n \, F_{c_n, r_n e_1}|) \, d\sigma_d \leq \frac{1}{2^n},$$

by Lemma 3.2. Define g by

$$g(\zeta) = \sqrt{\frac{1}{1 + \sum_{n=1}^{\infty} |c_n \, F_{c_n, r_n e_1}(\zeta)|^2}}.$$

It follows from (3.5) that g is continuous and vanishes only at e_1. Moreover

$$\begin{aligned} \int_{S_d} \log g \, d\sigma_d &= -\frac{1}{2} \int_{S_d} \log \left(1 + \sum_{n=1}^{\infty} |c_n \, F_{c_n, r_n e_1}|^2\right) d\sigma_d \\ &> - \int_{S_d} \log \prod_{n=1}^{\infty} (1 + |c_n \, F_{c_n, r_n e_1}|)^2 \, d\sigma_d \\ &= -2 \sum_{n=1}^{\infty} \int_{S_d} \log(1 + |c_n \, F_{c_n, r_n e_1}|) \, d\sigma_d \geq -2. \end{aligned}$$

Now suppose there is a non-zero m in $H^\infty(B_d)$ with $|m| \leq g$ almost everywhere. Then each of the functions $c_n F_{c_n, r_n e_1}$, being analytic in the ball of radius $1/r_n$, is in $P^2(|m|^2 \sigma)$; moreover they are all of norm less than one in this space, because

$$\int_{S_d} |c_n \, F_{c_n, r_n e_1}|^2 \, |m|^2 \, d\sigma \leq \int_{S_d} |c_n \, F_{c_n, r_n e_1}|^2 \, g^2 \, d\sigma < 1.$$

Let

$$f(z_1, \ldots, z_d) = \sum_{k=0}^{\infty} e^{-k^{4/7}} \frac{(k + d - 1)!}{(d-1)! k!} z_1^k.$$

By Theorem 1, there is a function h in $H^2(B_d)$ with

$$T_{\bar{m}}^{H^2(B_d)} h = f \, .$$

It follows that the linear map

$$\Gamma : p \mapsto \langle p, f \rangle_{H^2(B_d)} \, ,$$

defined a priori on the polynomials, extends by continuity to a bounded linear map on $P^2(|m|^2\sigma)$, as

$$|\Gamma(p)| = |\langle p, P(\bar{m}h)\rangle| = \left| \int p \, m \, \bar{h} \, d\sigma_d \right| \leq \|h\|_{H^2(B_d)} \, \|p\|_{P^2(|m|^2\sigma)} \, .$$

Moreover, each function $c_n F_{c_n, r_n e_1}$ is uniformly approximated on S_d by the partial sums of its Taylor series; hence

$$\Gamma(c_n \, F_{c_n, r_n e_1}) = \sum_{k=0}^{\infty} c_n \, \hat{F}_{c_n, r_n e_1}(k) \, e^{-k^{4/7}} \, . \tag{3.6}$$

But all the terms on the right-hand side of (3.6) are positive, and the i_n^{th} term is at least n by equation (3.4). This contradicts the boundedness of Γ.

References.

[1] Korenblum, B. I., M^c^Carthy, J. E., Non-attainable boundary values for H^∞ functions. *Extracta Math.* **8** (1993), 138-141.

[2] M^c^Carthy, J. E., Common range of co-analytic Toeplitz operators. *J. Amer. Math. Soc.* **3** (1990), 793-799.

[3] M^c^Carthy, J. E., Coefficient estimates on weighted Bergman spaces. *Duke Math. J.* **76** (1994), 751-760.

[4] Nawrocki, M., Linear functionals on the Smirnov class of the unit ball in $\mathbb{C}^n$. *Ann. Acad. Sci. Fenn.* **14** (1989), 369-379.

[5] Pommerenke, Ch., *Boundary behaviour of conformal maps.* Springer-Verlag, 1992.

[6] Rudin, W., *Function Theory in the unit ball of* $\mathbb{C}^n$. Springer-Verlag, 1980.

[7] Rudin, W., *New constructions of functions holomorphic in the unit ball of* $\mathbb{C}^n$. C.B.M.S. **63**. Amer. Math. Soc., 1986.

[8] Stoll, M., Mean growth and Taylor coefficients of some topological algebras of analytic functions. *Ann. Polon. Math.* **35** (1977), 139-158.

[9] Szegö, G., Über die Randwerten einer analytischen Funktionen. *Math. Ann.* 84 (1921), 232-244.

Recibido: 6 de noviembre de 1.994
Revisado: 2 de marzo de 1.995

Boris Korenblum
State University of New York
Albany, New York 12222, U.S.A.

John E. McCarthy*
Washington University
St. Louis, Missouri 63130, U.S.A.
mccarthy@math.wustl.edu

* Partially supported by the National Science Foundation grant DMS 9296099

Revista Matemática Iberoamericana
Vol. 12, N.º 1, 1996

Generalized Fock spaces, interpolation, multipliers, circle geometry

Jaak Peetre, Sundaram Thangavelu and Nils-Olof Wallin

donec eris felix, multos numerabis amicos
OVID

Abstract. By a (generalized) Fock space we understand a Hilbert space of entire analytic functions in the complex plane $\mathbb{C}$ which are square integrable with respect to a weight of the type $e^{-Q(z)}$, where $Q(z)$ is a quadratic form such that $\operatorname{tr} Q > 0$. Each such space is in a natural way associated with an (oriented) circle $\mathcal{C}$ in $\mathbb{C}$. We consider the problem of interpolation between two Fock spaces. If $\mathcal{C}_0$ and $\mathcal{C}_1$ are the corresponding circles, one is led to consider the pencil of circles generated by $\mathcal{C}_0$ and $\mathcal{C}_1$. If H is the one parameter Lie group of Moebius transformations leaving invariant the circles in the pencil, we consider its complexification H^c, which permutes these circles and with the aid of which we can construct the "Calderón curve" giving the complex interpolation. Similarly, real interpolation leads to a multiplier problem for the transformation that diagonalizes all the operators in H^c. It turns out that the result is rather sensitive to the nature of the pencil, and we obtain nearly complete results for elliptic and parabolic pencils only.

Introduction.

In this paper we shall understand by a *generalized Fock space* a Hilbert space of entire analytic functions in the complex plane $\mathbb{C}$ which are square integrable with respect to a weight of the type $e^{-Q(z)}$, where Q is a real quadratic form such that $\operatorname{tr} Q > 0$.

Such a quadratic form can be written as

$$Q(z) = k\,|z|^2 - \operatorname{Re}(l\,z^2)\,,$$

where k is a positive number ($k > 0$) and l is a complex number. Indeed, putting $z = x + iy$, we have

$$Q(z) = (k - \operatorname{Re} l)\,x^2 + (k + \operatorname{Re} l)\,y^2 + 2\,(\operatorname{Im} l)\,x\,y\,,$$

so that there are enough parameters to describe the most general real quadratic form. Moreover, we have

$$\operatorname{tr} Q = (k - \operatorname{Re} l) + (k + \operatorname{Re} l) = 2k > 0\,,$$

while

$$\det Q = (k - \operatorname{Re} l)(k + \operatorname{Re} l) - (\operatorname{Im} l)^2 = k^2 - |l|^2\,.$$

Thus, our spaces are labelled by pairs (k, l) and shall henceforth be denoted by $F_{(k,l)}$. If $f \in F_{(k,l)}$, its norm $\|f\|_{(k,l)}$ will be defined by

$$\|f\|^2_{(k,l)} = \frac{k^{1/2}}{\pi} \int_{\mathbb{C}} |f(z)|^2\, e^{-k|z|^2+\operatorname{Re}(lz^2)}\, dm(z)\,,$$

where we have written $dm(z) = dx\,dy$ (Euclidean measure). The corresponding inner product will be written $\langle f, f_1\rangle_{(k,l)}$ if $f, f_1 \in F_{(k,l)}$.

More generally, for $0 < p \le \infty$ we let $F^p_{(k,l)}$ be the space of entire analytic functions f such that (with the usual interpretation as a supremum if $p = \infty$)

$$\|f\|^p_{(k,l);p} = \frac{k^{1/2}}{\pi} \int_{\mathbb{C}} \Big(|f(z)|e^{-(k|z|^2+\operatorname{Re}(lz^2))/2}\Big)^p\, dm(z) < \infty\,;$$

for $p \ge 1$ the expression $\|f\|_{(k,l);p}$ is a norm and we have a Banach space; if $p < 1$ it is a quasi-norm and we have a quasi-Banach space. (In Section 4 we shall also briefly say a few words about (generalized)

Orlicz-Fock spaces $F^{\Phi}_{(k,l)}$.) Most of the time we shall however take $p = 2$.

EXAMPLE 1. In the paper [6] two special cases were considered:

$$F_{(k,0)} =: F_k \quad - \quad \text{weight } e^{-k|z|^2},$$
$$F_{(k,k)} =: G_k \quad - \quad \text{weight } e^{-2ky^2};$$

there the normalization was a slightly different one.

It will also be convenient to consider a certain limiting case of the spaces $F_{(k,l)}$, namely the case $l = ke^{i\theta}$, $k \to \infty$. To fix the ideas take first $\theta = 0$. Then formally

$$\frac{k^{1/2}}{\pi} \int_{\mathbb{C}} |f(z)|^2 \, e^{-2ky^2} \, dx \, dy \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} |f(x)|^2 \, dx\,.$$

This is because

$$\frac{k^{1/2}}{\pi} e^{-2ky^2} \, dy \to \frac{1}{\sqrt{2\pi}} \, \delta(y) \qquad \text{(Dirac measure)}.$$

Thus we are led to the space $L^2(\mathbb{R})$ of square integrable (non-analytic!) functions on the real line $\mathbb{R}$ equipped with the measure $dx/\sqrt{\pi}$. In the same way, for general θ we obtain the space $L^2(e^{-i\theta}\mathbb{R})$ of square integrable functions on the line $e^{-i\theta/2}\mathbb{R}$. We shall, alternatively, denote this space by S_θ (*Schrödinger space*). Its exact significance will be made more clear later on (see Example 1 in Section 1). We remark however right away that it should be viewed not primarily as a Lebesgue space, but as the completion in the metric in question of a space of certain analytic functions. The spaces S_θ have also a nice interpretation in terms of the heat equation (*cf.* [9]), but this point of view will not be pursued here.

The space $F := F_{(1,0)}$ will be called the *standard* Fock space and its norm will be written $\|\cdot\| = \|\cdot\|_{(1,0)}$.

In [6], among other things, the question of interpolation of the two scales of spaces F^p_k and G^p_k was raised.

1) Regarding complex interpolation the following result was established:

$$[F^{p_0}_{k_0}, F^{p_1}_{k_1}]_\theta = F^{p_\theta}_{k_\theta}\,, \qquad [G^{p_0}_{k_0}, G^{p_1}_{k_1}]_\theta = G^{p_\theta}_{k_\theta}\,,$$

where in both cases p_θ is given by

$$\frac{1}{p_\theta} = \frac{1-\theta}{p_0} + \frac{\theta}{p_1}, \qquad 0 < \theta < 1,$$

while k_θ is in the former case a weighted geometric mean of k_0 and k_1:

$$k_\theta = k_0^{1-\theta} k_1^\theta$$

and in the latter case the corresponding weighted harmonic mean:

$$\frac{1}{k_\theta} = \frac{1-\theta}{k_0} + \frac{\theta}{k_1}.$$

2) What real interpolation concerns only a reduction to a multiplier problem was indicated in the case of the scale F_k^p (with p fixed, k variable).

This curious simultaneous occurence of both the geometric and the harmonic mean in essentialy the same context, already recorded in [6], has rised our curiosity. It is one of the objects of this paper to clarify this point and it is for this reason that it was decided that it is necessary to put oneself on the level of the generalized Fock spaces. At the same time we shall also settle the issue of real interpolation, at least in the two cases just indicated.[1]

It turns out that the subject is intimately connected with classical end 19th century higher geometry (German: "höhere Geometrie"), especially circle geometry. Namely, each space $F_{(k,l)}$ is in a natural way associated with a certain circle $C_{(k,l)}$ (or, perhaps rather, a disk $D_{(k,l)}$). And the problem of interpolation between two spaces $F_{(k_0,l_0)}$ and $F_{(k_1,l_1)}$ leads one to consider the pencil of circles generated by $C_{(k_0,l_0)}$ and $C_{(k_1,l_1)}$. There are basically three different types of pencils which we have decided to term elliptic, parabolic and hyperbolic. We have been able to settle most of our question in the elliptic and, to some extent, in the parabolic case but in the hyperbolic case some unexpected difficulties turn up so in this case our results are so far less complete.

Eventually we would like to extend the theory developed in the present paper to the case of several variables. We expect that the rôle

[1] As we shall see, the occurence of these special means is, however, a delusion to some extent!

played by the unit disk here will be taken over by a bounded symmetric domain of tube type, in the first instance one of type III. But this has to wait for the future ...

The plan of the paper is as follows. In Section 1 we set the foundations for the theory of generalized Fock spaces. In particular, we begin to uncover the geometric and the group theoretic aspects of the matter. In an appendix to Section 1 we discuss of the possiblity of assigning spaces not only to proper disks (not containing the point at infintity) but also to arbitrary disks on the Riemann sphere S^2. In the next two sections the interpolation theory of generalized Fock spaces will be developed, complex interpolation in Section 2 and real interpolation in Section 3. The short Section 4 contains some auxiliary results not directly related to the main theme of the paper. The theorems, lemmas etc. are numbered independently in each section.

1. Gauss-Weierstrass functions and Shale-Weil operators. Segal bundle.

We shall study our generalized Fock spaces $F_{(k,l)}$ with the aid of the family of functions e_{ac},

$$e_{ac}(z) = e^{(az^2+cz)/2},$$

where a and c are arbitrary complex numbers. In [9][2] these functions were referred to as *Gauss-Weierstrass functions*; other names current in the literature are: coherent states, Gabor wavelets etc. In our theory they serve as "atoms".

From [9] we take over the following formula:

$$\|e_{ac}\|^2 = \exp\left(\frac{\operatorname{Re} a\bar{c}^2 + |c|^2}{1-|a|^2}\right)(1-|a|^2)^{-1/2} \tag{1}$$

or in polarized form

$$\langle e_{ac}, e_{bd}\rangle = \exp\left(\frac{\dfrac{a\bar{d}^2 + \bar{b}c^2}{2} + c\bar{d}}{1-a\bar{b}}\right)(1-a\bar{b})^{-1/2}, \tag{2}$$

[2] We would like to turn the reader's attention to the circumstance that there is, regretfully, an abundance of misprints in [9]; this is most unfortunate as this reference was meant to be "a small compendium of useful formulae connected with . . . Fock space".

the proper interpretation of these formulae being that $e_{ac} \in F$ if and only if $|a| < 1$. In particular, the system of functions $\{e_{ac}\}$, with $|a| < 1$, $c \in \mathbb{C}$, is total in F.

Next we perform a reduction to standard form. We rewrite the norm in our space $F_{(k,l)}$ as follows:

$$(3) \qquad \|f\|_{(k,l)} = \|f e^{lz^2/2}\|_{(k,0)} = \left\| \frac{1}{k^{1/4}} f\left(\frac{z}{k^{1/2}}\right) e^{lz^2/2k} \right\| .$$

In other words, we have a unitary map

$$(4) \qquad \begin{aligned} V : F_{(k,l)} &\longrightarrow F \\ f(z) &\mapsto \frac{1}{k^{1/4}} f\left(\frac{z}{k^{1/2}}\right) e^{lz^2/2k} , \end{aligned}$$

that is,

$$\|f\|_{(k,l)} = \|Vf\| , \quad \text{if } f \in F_{(k,l)} .$$

The inverse map reads

$$(5) \qquad \begin{aligned} V^{-1} : F &\longrightarrow F_{(k,l)} \\ f(z) &\mapsto k^{1/4} f(k^{1/2} z) e^{-lz^2/2} . \end{aligned}$$

Using (3) in conjunction with (1) we can formally give an expression for the norm of a Gauss-Weierstrass function in the space $F_{(k,l)}$:

$$(6) \quad \|e_{ac}\|^2_{(k,l)} = k^{1/2} \exp\left(\frac{\operatorname{Re}\left(\dfrac{(a+l)\,\bar{c}^2}{k^2}\right) + \dfrac{|c|^2}{k}}{1 - \dfrac{|a+l|^2}{k^2}} \right) \left(1 - \frac{|a+l|^2}{k^2}\right)^{-1/2} .$$

Indeed, with the above notation we have

$$V e_{ac} = \frac{1}{k^{1/4}}\, e_{(a+l)/k, ck^{-1/2}} , \qquad \|e_{ac}\|_{(k,l)} = \|V e_{ac}\| ,$$

so, using (1) and (4), (6) readily follows. (The reader will have no difficulty in writing down the corresponding polarized identity.) The interpretation of (6) is the following:

$$e_{ac} \in F_{(k,l)} \quad \text{if and only if} \quad |a + l| < k .$$

Thus, *to the space* $F_{(k,l)}$ *there corresponds the disk* $D_{(k,l)}$ *with radius* k *and center at the point* $-l$, $D_{(k,l)} = \{a : |a+l| < k\}$. We denote by $\mathcal{C}_{(k,l)}$ the circle which constitutes the boundary of $D_{(k,l)}$, that is, $\mathcal{C}_{(k,l)} = \{a : |a+l| = k\}$. We put $D = D_{(1,0)}$ and $\mathcal{C} = \mathcal{C}_{(1,0)}$, unit disk and unit circle respectively. A total system of functions in this case is $\{e_{ac}\}$, with $a \in D_{(k,l)}$, $c \in \mathbb{C}$.

EXAMPLE 1. In [6] the following instances of this are found:

- to the space F_k there corresponds the disk $D_{k,0} = \{a : |a| < k\}$,
- to the space G_k there corresponds the disk $D_{k,k} = \{a : |a+k| < k\}$.

To this we may now add:

- to the space S_θ there corresponds the halfplane

$$P_\theta = \{a : \operatorname{Re} ae^{-i\theta} < 0\} \qquad \text{(a generalized disk).}$$

We see that S_θ has the interpretation as the closure of the functions $\{e_{ac}\}$ with $a \in P_\theta$, $c \in \mathbb{C}$ in a suitable metric.

EXAMPLE 2. As another application of formula (6) let us record the following formula for the reproducing kernel in the space $F_{(k,l)}$:

$$K(z,w) = k^{1/2}\, e^{-l(z^2+\bar{w}^2)/2+kz\bar{w}}\,.$$

If $k = 1$, $l = 0$ it reduces, of course, to the well-known expression for the reproducing kernel in the standard Fock space F:

$$K(z,w) = e^{z\bar{w}}\,;$$

see *e.g.* [6, formula (7.2)] with $\alpha = 0$ and $n = 1$. The reproducing kernel will not play any rôle in our discussion.

Returning to the general discusion, let us note that the intersection of two Fock spaces $F_{(k_0,l_0)}$ and $F_{(k_1,l_1)}$ is non-nil,

$$F_{(k_0,l_0)} \cap F_{(k_1,l_1)} \neq \{0\}\,,$$

provided the corresponding disks have non-empty intersection,

$$D_{(k_0,l_0)} \cap D_{(k_1,l_1)} \neq \varnothing\,.$$

This is follows from the fact that $e_{ac} \in F_{(k_0,l_0)} \cap F_{(k_1,l_1)}$ if $a \in D_0 \cap D_1$, $c \in \mathbb{C}$.[3]

Next, we put into play the *Shale-Weil operators*. Let $G^c = \mathrm{Sp}(2,\mathbb{C})$ the group of complex 2×2 matrices $g = \left(\begin{smallmatrix}\alpha & \beta \\ \gamma & \delta\end{smallmatrix}\right)$ with $\alpha\delta - \beta\gamma = 1$. To such a matrix g we associate an integral operator

$$(7)\quad T_g f(z) = \frac{B^{1/2}}{\pi} \int_{\mathbb{C}} \exp\Big(\frac{Az^2 + 2Bz\bar{w} + C\bar{w}^2}{2}\Big) f(w)\, e^{-|w|^2}\, dm(w),$$

where

$$A = \frac{\beta}{\delta}, \quad B = \frac{1}{\delta}, \quad C = -\frac{\gamma}{\delta}.$$

We proceed somewhat informally. We think of T_g as being defined on a suitable (preferably dense) subspace of our standard Fock space F and, for the time being (*cf.* Remark 1 below), we let T_g go undefined if $\delta = 0$. In addition, due to the ambiguity in the definition of the square root $B^{1/2}$, T_g is actually determined only up to sign $\pm$.

In [9] the following statement was proved:

T_g is unitary if and only if g is a pseudo-unitary matrix, i.e.

$$g \in G = \mathrm{SU}(1,1);$$

we consider the previous group G^c as the complexification of the group G. It was also shown in [9] that the composition of two such operators T_{g_1} and T_{g_2}, if it makes sense, is again an operator of the same type; indeed, one has $T_{g_1}T_{g_2} = \pm T_{g_1 g_2}$. In other words, we have a unitary representation of a suitable double cover of $G = \mathrm{SU}(1,1) \approx \mathrm{Sp}(2,\mathbb{R})$ (the symplectic group), *viz.* the *metaplectic group* $\widetilde{G} = \mathrm{Mp}(2,\mathbb{R})$. It is the ambiguity in the definition of the square root that forces us to pass to a cover. A typical element of $\tilde{G}$ is given by a pair $\tilde{g}$, an element $g = \left(\begin{smallmatrix}\alpha & \beta \\ \gamma & \delta\end{smallmatrix}\right)$ of G *plus* a determination of the square root of δ, the composition being defined as follows: If we have a second element $\tilde{g}'$, then the composition $\tilde{g}'' = \tilde{g}'\tilde{g}$ is found exploiting the identity

$$\sqrt{\delta''} = \sqrt{1 + \frac{\gamma'}{\delta'}\, g0}\, \sqrt{\delta'}\, \sqrt{\delta}\,.$$

---

[3] As the referee has pointed out to us, it is likely that, conversely, $F_{(k_0,l_0)} \cap F_{(k_1,l_1)} \neq \{0\}$ implies $D_{(k_0,l_0)} \cap D_{(k_1,l_1)} \neq \varnothing$, but we do not know how to prove such a result. We are indepted to him for several other precious remarks as well.

Added in proof (Nov. 95). This question has now been affirmatively settled by the authors.

We use the fact that, as $|\gamma'/\delta'| < 1$, one can define $a \mapsto (1+(\gamma'/\delta')a)^{1/2}$ as an analytic function in the unit disk $D(1,0)$ taking the value 1 at the origin.

This representation is known in the literature under various names: oscillator, harmonic, Bargmann-Segal, Shale-Weil, etc. representation. One speaks also, referring to the operators T_g, of the *oscillator group.* If one restricts attention to matrices g with the property that the corresponding Moebius transformation $a \mapsto ga = (\alpha a + \beta)/(\gamma a + \delta)$ maps the unit disk $D(0,1)$ into itself, not onto, then one obtains instead the *oscillator semi-group* (*cf.* [5], [10]).

Another formula, in [9] established for the group G, is

$$T_g e_{ac}(z) = \frac{1}{(\gamma a + \delta)^{1/2}} \exp\Big(- \frac{\gamma c^2}{2(\gamma a + \delta)}\Big) e_{ga, c/(\gamma a + \delta)}(z), \tag{8}$$

which is easy to verify at least on the formal level.

REMARK 1. Note that, in contradistinction to (7) above, this formula (8) makes sense even if $\delta = 0$. Thus (8) may serve as a definition of T_g in this case; we view then T_g as a linear operator on the linear hull of the family of functions $\{e_{ac}\}$. We must only make sure that $ga \neq \infty$ or that $\gamma a + \delta \neq 0$.

Using (1) we find from this

$$\begin{aligned} \|T_g e_{ac}\|^2 &= \frac{1}{\pi} \frac{1}{|\gamma a + \delta|} \exp\Big(-\operatorname{Re} \frac{\gamma c^2}{\gamma a + \delta}\Big) \\ &\quad \cdot \exp\left(\frac{\operatorname{Re} ga \Big(\overline{\frac{c}{\gamma a + \delta}}\Big)^2 + \frac{|c|^2}{|\gamma a + \delta|^2}}{1 - |ga|^2} \right) (1 - |ga|^2)^{-1/2}. \end{aligned} \tag{9}$$

EXAMPLE 3. Let $g = \left(\begin{smallmatrix} \alpha & \beta \\ 0 & \delta \end{smallmatrix}\right)$, so that $\delta = \alpha^{-1}$. Then $T_g f(z) = \alpha^{1/2} f(\alpha z) e^{i\beta\alpha z}$ for a general function f, while

$$T_g e_{ac}(z) = \alpha^{1/2} e_{(\alpha a + \beta)/\delta,\, c/\delta}(z).$$

Let us write $\alpha = k^{-1/2}$, $\beta = k^{-1/2} l$, so that $\delta = k^{1/2}$. (The corresponding Moebius transformation is thus $a \mapsto (a + l)/k$.) Then we see that $V = T_g$ with $g = \left(\begin{smallmatrix} k^{-1/2} & k^{-1/2} l \\ 0 & k^{1/2} \end{smallmatrix}\right)$. That is, we have $\|f\|_{(k,l)} = \|T_g f\|$ for $f \in F_{(k,l)}$.

The above suggests to consider in general Hilbert spaces with a norm of the type $\|T_g f\|$ for some $g \in G^c$, where $\|\cdot\|$ stands for the standard Fock norm in our standard Fock space F.

In this direction we can establish the following basic result.

Theorem 1. *Let g be an element of G^c such that $g^{-1}(D) = D_{(k,l)}$ (in particular, one has $\infty \notin g^{-1}(D)$). Here $D = D_{(1,0)}$ is the unit disk. Then $T_g^{-1}(F) = F_{(k,l)}$. Moreover, we have $\|f\|_{(k,l)} = \|T_g f\|$ for $f \in F_{(k,l)}$.*

REMARK 2. The transformations g occurring in the statement give an element of $G^c \backslash G$, that is, a residue class modulo G in G^c.

The proof of this theorem which will be based on the following two lemmata may be of independent interest.

Lemma 1 (generalized Lagrange identity). *Let $g = \left(\begin{smallmatrix} \alpha & \beta \\ \gamma & \delta \end{smallmatrix}\right)$ be a matrix in G^c such that the inverse image of the unit disk D is the disk $D_{(k,l)}$. Then*

$$1 - |ga|^2 = \frac{1}{|\gamma a + \delta|^2} \, \frac{k^2 - |a + l|^2}{k} \,. \tag{10}$$

Moreover, one has

$$k = \frac{1}{|\alpha|^2 - |\gamma|^2} \,, \qquad l = -\frac{\bar{\alpha}\beta - \bar{\gamma}\delta}{|\alpha|^2 - |\gamma|^2} \,. \tag{11}$$

Lemma 2. *Let the matrix g and the parameters k and l be as in Lemma 1. Then the following identity holds*

$$(\bar{\alpha}\bar{a} + \bar{\beta})\, k - (\gamma a + \delta)(\bar{a} + \bar{l}) = \gamma (k^2 - |a + l|^2). \tag{12}$$

REMARK 3. If $g(D) = D$, *i.e.* if $g \in G$, then we get from Lemma 1 the well-known identity

$$1 - |ga|^2 = \frac{1}{|\gamma a + \delta|^2} \, (1 - |a|^2) \,, \tag{13}$$

often used in function theory; in [9] it was called *Lagrange identity*. (The reason for this choice of name is the following: let us put ourselves in the case of the fundamental symmetry of D interchanging an arbitrary point $b \in D$ and the origin 0, that is, the mapping $a \mapsto (b-a)/(1-a\bar{b})$. Then (10) becomes

$$|1-a\bar{b}|^2 - |b-a|^2 = (1-|a|^2)(1-|b|^2)\,.$$

Introducing homogeneous coordinates (writing $a = a_1/a_0$, etc.), this gives

$$|\bar{a}_0\bar{b}_0 - \bar{a}_1\bar{b}_1|^2 = (|a_0|^2 - |a_1|^2)(|b_0|^2 - |b_1|^2)\,.$$

This is the expression of the norm of a bivector in the pseudo-Hermitean metric $|a_0|^2 - |a_1|^2$.) In the same way from Lemma 2 we obtain the usual condition for a complex unimodular matrix to be in G, *viz.* $\bar{\alpha} = \delta$, $\bar{\beta} = \gamma$.

PROOF OF LEMMA 1. After having chased a denominator we perform the following chain of transformations:

$$\begin{aligned}|\gamma a + \delta|^2 - |\alpha a + \beta|^2 &= |\gamma|^2\,|a|^2 + 2\,\mathrm{Re}\,\gamma\bar{\delta}a + |\delta|^2 \\ &\quad - (|\alpha|^2\,|a|^2 + 2\,\mathrm{Re}\,\alpha\bar{\beta}a + |\beta|^2) \\ &= -(|\alpha|^2 - |\gamma|^2)\left|a + \frac{\bar{\alpha}\beta - \bar{\gamma}\delta}{|\alpha|^2 - |\gamma|^2}\right|^2 \\ &\quad + \left(|\delta|^2 - |\beta|^2 + \frac{|\bar{\alpha}\beta - \bar{\gamma}\delta|^2}{|\alpha|^2 - |\gamma|^2}\right).\end{aligned}$$

The last term in the last expression can be rewritten as

$$\begin{aligned}|\delta|^2 - |\beta|^2 + \frac{|\bar{\alpha}\beta - \bar{\gamma}\delta|^2}{|\alpha|^2 - |\gamma|^2} &= \frac{|\alpha|^2\,|\delta|^2 + |\beta|^2\,|\gamma|^2 - 2\,\mathrm{Re}\,\alpha\delta\bar{\beta}\bar{\gamma}}{|\alpha|^2 - |\gamma|^2} \\ &= \frac{|\alpha\delta - \beta\gamma|^2}{|\alpha|^2 - |\gamma|^2} = \frac{1}{|\alpha|^2 - |\gamma|^2}\,,\end{aligned}$$

where we in the last step used $\alpha\delta - \beta\gamma = 1$. Taking now (10) as definition of the numbers k and l, we formally arrive at formula (11). It is however readily seen from this equality that these parameters must have the desired significance.

PROOF OF LEMMA 2. *Step* 1. First we observe that if (12) holds for a matrix g, unimodular or not, then it holds for any multiple $t \cdot g$, where $t \in \mathbb{R}$.

Step 2. Reduction to the case $k = 1$, $l = 0$. Put $a_1 = (a + k)/k$. Then $|ga| < 1$ if and only if $|a_1| < 1$, while (12) can be written

$$\left(\bar{\alpha}\bar{a}_1 + \frac{\bar{\beta} - \bar{l}\bar{\alpha}}{k}\right) - \left(\gamma a_1 + \frac{\delta - l\gamma}{k}\right)\bar{a}_1 = \gamma(1 - |a_1|^2).$$

But this is nothing but (12) for $g_1 = \left(\begin{smallmatrix} \alpha & \beta_1 \\ \gamma & \delta_1 \end{smallmatrix}\right)$, where $\beta_1 = (\beta - l\alpha)/k$, $\delta_1 = (\delta - l\gamma)/k$, and this matrix represents the transformation $a_1 \mapsto b$, where $b = ga$. Clearly $\det g_1 = \alpha\delta_1 - \beta_1\gamma = 1/k \in \mathbb{R}$. By Step 1 the same equality holds then also for the corresponding unimodular matrix.

Step 3. The case $k = 1$, $l = 0$. In this case, as is we have already noted (see Remark 1) that in this case $\bar{\alpha} = \delta$, $\bar{\beta} = \gamma$. So then (12) is equivalent to the absolutely trivial relation

$$(\delta\bar{a} + \gamma) - (\gamma a + \delta)\,\bar{a} = \gamma(1 - |a|^2).$$

Next we proceed to the proof of the Theorem 1.

PROOF OF THEOREM 1. Let g be a matrix such that $g^{-1}(D) = D_{(k,l)}$. It suffices to show that

$$\|T_g e_{ac}\| = \|e_{ac}\|_{(k,l)}, \qquad \text{for } a \in D_{(k,l)},\ c \in \mathbb{C}. \tag{14}$$

For then we have by polarization

$$\langle T_g e_{ac}, T_g e_{bd}\rangle = \langle e_{ac}, e_{bd}\rangle_{(k,l)}, \qquad \text{for } a, b \in D_{(k,l)},\ c, d \in \mathbb{C},$$

whence, by considering linear combinations of Gauss-Weierstrass functions and applying a density argument (the Gauss-Weierstrass functions form a total set), it follows that $\|T_g f\| = \|f\|_{kl}$ for all $f \in F_{(k,l)}$.

Now we verify (14). To this end we must do some transformations in formula (9) showing that it reduces to (6).

First, we observe that the two exponent free factors combine in view of (10) in Lemma 1 to a factor

$$\frac{k^{1/2}}{\pi}\left(1 - \frac{|a + l|^2}{k^2}\right)^{-1/2}.$$

Next, we look at the exponential factors. Using Lemma 1 once more we see that the coefficient of $|c|^2$ in the exponent becomes

$$\frac{1}{1-\dfrac{|a+l|^2}{k^2}}\,,$$

as it should. Similarly, formula (2) in Lemma 2 helps us to bring the c^2-term into the right shape. Indeed, we find (this is what stands after the sign Re after we have combined the two exponential factors)

$$\begin{aligned}
&-\frac{\gamma\,c^2}{\gamma\,c+\delta}+\frac{\overline{g\bar{a}}}{1-|ga|^2}\,\frac{c^2}{(\gamma\,a+\delta)^2}\\
&\qquad=\Big(-\frac{\gamma\,c^2}{\gamma\,c+\delta}+\frac{\overline{\alpha a+\beta}}{(1-|ga|^2)\,|\gamma\,a+\delta|^2\,(\gamma\,a+\delta)}\Big)\,c^2\\
&\qquad=\frac{-(k^2-|a+l|^2)+\overline{\alpha a+\beta}k}{(\gamma\,a+\delta)\,(k^2-|a+l|^2)}\,c^2\\
&\qquad=\frac{a+l}{k^2-|a+l|^2}\,c^2\,,
\end{aligned}$$

which is precisely what is desired (see (6)). (In [9] the corresponding computations were done when $g\in G$.)

Let us also indicate an alternative less direct approach. Although it is apparently shorter than our previous proof, we prefer the form because of its constructive flavor involving also the beautiful identities in Lemma 1 and Lemma 2, which, as we have hinted at, may well be of independent interest.

ALTERNATIVE PROOF OF THEOREM 1. We begin by noting that the unitary map V in (4) obviously corresponds to the standard affine map $g_0 : D_{(k,l)} \to D$ given by $g_0 a = (a+l)/k$, *i.e.* $V = T_{g_0}$. It follows that if g is any element of G^c such that $g^{-1}(D) = D_{(k,l)}$ then we have $g = hg_0$ for some $h \in G$ (a pseudo-unitary matrix). But then $T_g = \pm T_h T_{g_0} = \pm T_h V$. As T_h is a unitary map on the Hilbert space F, this again implies that

$$\|T_g f\| = \|T_h V f\| = \|V f\| = \|f\|_{(k,l)}\,,$$

where we in the last step used (4).

Theorem 1 has an obvious generalization to the spaces F^p, $p \geq 1$.

Corollary. *Let g be as in Theorem 1. Let $p \geq 1$. Then again $T_g^{-1}(F^p) = F^p_{(k,l)}$. Moreover, we have the norm equivalence $\|f\|_{(k,l);p} \approx \|T_g f\|_p$ for $f \in F^p_{(k,l)}$.*

PROOF. In [9] it was shown that the group G^c acts on the spaces F^p (for the case $p < 1$ see Section 4). Therefore the previous (alternative) proof of Theorem 1 extends to the present situation without any changes.

Appendix to Section 1. Non-existence of a certain bundle.[4]

Now we have settled our main question (see the above Theorem 1) but only in the auxiliary assumption that the inverse image of the unit disk under g does not contain the point at infinity, ∞. It is a legitimate question whether it might be possible to free oneself of this assumption. In this appendix we give a brief discussion of this issue. However, it is mainly a negative experience.

First we recall that there are on the Riemann sphere S^2 three kinds of (generalized) disks:

1) proper disks;
2) halfplanes (limiting case of a disk);
3) exteriors of proper disk.

Alternatively, we could speak of oriented (generalized) circles: if an oriented circle is given, we pick up the disk that is to its "left". Thus there is a 1:1 correspondence

$$\text{disks} \longleftrightarrow \text{oriented circles}.$$

The question is thus whether it is possible to associate in a natural way to a generalized disk D on S^2 a "Fock space" $\mathfrak{F}_D$, extending the previous correspondence $D_{(k,l)} \mapsto F_{(k,l)}$. Introducing the notation $\mathfrak{M}$ for the manifold of disks (oriented circles) this would yield a bundle of Fock spaces $\mathfrak{F}$ over $\mathfrak{M}$, say. (Let us remark that in previous work

[4] A reader who is only interested in analysis questions (interpolation, multipliers) can safely omit this appendix. The senior author would like to thank Johan Råde for an illuminating discussion helping to clarifying some questions connected with the topology of the Lie groups G, $\tilde{G}$ and G^c.

one of us has already encountered several occurences of vector bundles (of infinite rank) over complex manifolds: the Fock bundle [11] and the Fischer bundle [12]; here we have a manifold $\mathfrak{M}$ which is not complex.) However, this seems to be a chimera: *the bundle $\mathfrak{F}$ does not exist.* Let us indicate why this is so.

Let us fix the disk D. Then D can be mapped conformally onto the unit disk $D(0,1)$ but not in a unique way. Any such map comes from a certain element g of the group G^c, for reference, let us call it a *frame*. It is natural to try to define the fiber $\mathfrak{F}_D$ as a kind of pullback of the standard Fock space $F = F_{(1,0)}$. More exactly, given any two frames coming from group elements g and g_1 we can write $g_1 = ug$ with $u \in G$ and one is then led to consider two functions f_1 and f in F as representatives of one and the same element of $\mathfrak{F}_D$ if $f_1 = T_u f$. However, by the above T_u is defined only up to sign $\pm$, which seems to be an unsurmountable difficulty and so our approach breaks down. It is only when we restrict ourselves to suitable open subsets of $\mathfrak{M}$ that we can make it work, for instance, when we consider the subset of all disks avoiding one point, say, the point at infinity, but then we are back in the situation considered already in Section 1.

A possible way out would be to count elements of F modulo sign but this would then essentially lead to a projective bundle, not a vector bundle, but this is not exactly what we desire.

One can give the above somewhat heuristic considerations also a somewhat more rigorous formulation using the language of principal bundles and their associated bundles, which we now indicate very quickly.[5] Let us denote by $\mathfrak{R}$ the manifold of all frames. Of course, we have the trivial identification $\mathfrak{R} \approx G^c$. Moreover, we can identify $\mathfrak{M}$ with a certain space of cosets of G^c, $\mathfrak{M} \approx G\backslash G^c$. It follows that $\mathfrak{R}$ can be viewed as a principal bundle over $\mathfrak{M}$ with G as structure group. If V is a any vector space on which G acts (a representation space), there is an associated vector bundle $\mathfrak{V}$ on which G acts. In our case we would like to take $V = F$ but the trouble is that its double cover the metaplectic $\tilde{G}$ acts on F, not G itself. There seems to be no way out of this dilemma. This is connected with the fact that while G admits a double cover, its complexification G^c does not. This again depends on the following facts: On the one hand, as G^c as a topological space

[5] We are now adressing ourselves to those readers who are familiar we the rudiments of this theory (see *e.g.* the book [8]). Notice that in the conventional treatment the structure group usually acts from the right, while in our formulation we have a group action (of the group G) from the left.

is simply connected its fundamental group is trivial, $\pi_1(G^c) = 1$, while, on the other hand, G is contractible to a circle S^1 and thus has the fundamental group $\pi_1(G) = \mathbb{Z}$.

2. Complex interpolation. Circle geometry.

Now we begin to interpolate. In this Section we shall deal with complex interpolation exclusively, thus relegating real interpolation to Section 3.

Our objective is to determine the complex interpolation spaces between two given Fock spaces $F_{(k_0,l_0)}$ and $F_{(k_1,l_1)}$. Assuming that their intersection is not nil, $F_{(k_0,l_0)} \cap F_{(k_1,l_1)} \neq \{0\}$, we shall show that the interpolation space $[F_{(k_0,l_0)}, F_{(k_1,l_1)}]_\theta$ again is a certain Fock space $F_{(k_\theta,l_\theta)}$. Indeed, if $C_{(k_0,l_0)}$ and $C_{(k_1,l_1)}$ are the circles corresponding to the spaces $F_{(k_0,l_0)}$ and $F_{(k_1,l_1)}$, then the circle $C_{(k_\theta,l_\theta)}$ corresponding to $F_{(k_\theta,l_\theta)}$ belongs to the pencil of circles generated by the two given circles $C_{(k_0,l_0)}$ and $C_{(k_1,l_1)}$.

First we recall some general facts about complex interpolation (for details, consult the excellent book Bergh-Löfström [2]).

Consider quite generally any Banach couple (A_0, A_1), *i.e.* A_0 and A_1 are two Banach spaces (over $\mathbb{C}$) both continuously imbedded in a Hausdorff topological vector space $\mathcal{A}$. An element a in the linear hull $A_0 + A_1$ of A_0 and A_1 in $\mathcal{A}$ is said to be in the complex interpolation space $[A_0, A_1]_\theta$, where $0 < \theta < 1$, if, informally speaking, there is a complex curve through a connecting A_0 and A_1. More exactly, we require that there exists a holomorphic function $f(\zeta)$, where $\zeta = \xi + i\eta$ is a complex variable, defined in the strip $0 < \operatorname{Re}\zeta < 1$ with values in $A_0 + A_1$ such that $a = f(\theta)$ and such that its boundary values satisfy $f(i\eta) \in A_0$, $f(1 + i\eta) \in A_1$. In addition, some growth conditions must be satisfied, and we have not told in what sense the boundary values are taken, but we shall not enter into such technicalities here.

Next, let us specialize to the case when

$$\begin{aligned} A_0 &= E = \text{a given Banach space,} \\ A_1 &= D(\Lambda) = \text{the domain of a closed} \\ &\qquad\qquad \text{unbounded operator } \Lambda \text{ acting in } E. \end{aligned}$$

Then one expects that, in suitable assumptions, one has $[E, D(\Lambda)]_\theta = D(\Lambda^\theta)$, where Λ^θ stands for the suitably defined θ-th power of Λ. For instance, it suffices that imaginary powers $\Lambda^{i\eta}$ make sense and satisfy a suitable growth estimate, *e.g.* $\|\Lambda^{i\eta}\| \leq C(1+|\eta|)^m$ or even $\|\Lambda^{i\eta}\| \leq Ce^{|\eta|^l}$, $l < 1$, will do and certainly $\|\Lambda^{i\eta}\| = 1$ (isometry). Then the canonical quasi-optimal choice of the function in the above construction is $f(\zeta) = \Lambda^{\theta-\zeta}e$, where e is an element of the space E. In particular, the following situation is allowed: E = a Hilbert space, Λ = a positive self-adjoint operator in E.

In the Fock case there is a natural choice for the operators Λ^{ζ}, namely $\Lambda^{\zeta} = T_{g_\zeta}$, where the transformations g_ζ form a certain complex one parameter subgroup of G^c leaving invariant the pencil generated by the given circles $C_{(k_0,l_0)}$ and $C_{(k_1,l_1)}$. Before making this more precise let us review some basic facts about circle geometry (classical references for "higher geometry" are Klein [7] and Blaschke [1][6]).

The equation of a (generalized) circle $\mathcal{C}$ on the Riemann sphere S^2 can be written

$$(1) \qquad A a\bar{a} + 2\,\mathrm{Re}\, B\bar{a} + C = 0$$

(or equivalently $Aa\bar{a} + B\bar{a} + \bar{B}a + C = 0$), where A and C are real numbers, while B is a complex quantity. Thus (1) means one of the following: a genuine (real) circle; in a limiting case, a line (a circle through the point at infinity); a point circle; an imaginary circle. We see that each circle $\mathcal{C}$ gives a triple $\phi = (A, B, C)$ determined up to a non-zero real multiple. Note that such a triple consists of two real and one complex numbers; alternatively, splitting B into its real and imaginary parts, we could likewise have spoken of a quadruple of real numbers, thus a point in $\mathbb{R}^4$. (Sometimes it is also convenient to put $\phi = (A, B, \bar{B}, C)$.) A *pencil of circles* is a one parameter of family of circles of the form

$$(A_0 + tA_1)\,a\bar{a} + 2\,\mathrm{Re}(B_0 + tB_1)\,\bar{a} + (C_0 + tC_1) = 0\,, \qquad t \in \mathbb{R}\,.$$

We say that the pencil is generated by the circles $\mathcal{C}_0$ and $\mathcal{C}_1$ corresponding to the triples $\phi_0 = (A_0, B_0, C_0)$ and $\phi_1 = (A_1, B_1, C_1)$. It is the sign of the *discriminant* $D = AC - |B|^2$ that determines the geometric meaning of the equation (1): assuming that $A \neq 0$

if $D < 0$ it is a real circle;

if $D = 0$ it is a point circle;

[6] The former book was actually edited by Blaschke.

if $D > 0$ it is an imaginary circle.

The assignment $\mathcal{C} \mapsto (A, B, C)$ thus defines a mapping from the space of all (generalized) circles to real projective space $\mathbb{PR}^3$ equipped with a distinguished quadric

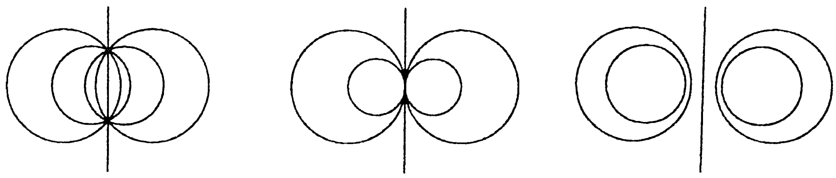
$$\mathcal{Q} : (\phi, \phi) = D = AC - |B|^2 = 0 ,$$

corresponding to a quadratic form in $\mathbb{R}^4$ of signature $+ - - -$ (or index of inertia $(1, 3)$).

We can thus set up a small dictionary.

space of circles	projective space $\mathbb{PR}^3$
point circle	point on $\mathcal{Q}$
circle	point not on $\mathcal{Q}$
pencil of circles	line
the group G^c	the Lorentz group $SO(1, 3)$

We can apply the insights gained above to describe the structure of pencils of circles. There are essentially three cases depending on the mutual position of the corresponding line in $\mathbb{PR}^3$ and the quadric $\mathcal{Q}$. This is depicted in the figure on this page.



hyperbolic (time-like) parabolic (light-like) elliptic (space-like)

Thus in the former case the circles go through two real points, in the middle case they are tangent at a real point and in the last case they go through two imaginary points (and do not meet in the real).

It is clear that the spaces F_k correspond to an elliptic situation (the concentric circles $|a| < k$), while the spaces G_k correspond to a parabolic situation (the circles $|a + k| < k$ tangent to the imaginary axes at the origin). *This explains, in particular, their different interpolational behavior* (see Introduction).

REMARK 1. Note also that a pencil of circles contains in general a unique line called its *power line*. The exception is when we have a pencil of circles through the point at infinity. Then all elements of the pencil are lines, of course.

REMARK 2. Notice also that there is a duality for pencils of circles. The dual pencil consists of all circles orthogonal to the circles of the given pencil. This duality interchanges elliptic and hyperbolic, respects parabolic.

Now we discuss the subgroup of G^c which preserves a given pencil of circles. Let C_0 and C_1 a pair of generating circles and denote by H the group of transformations leaving each of them invariant. Then we have the following lemma, which is the key to our discussion of complex interpolation of Fock spaces in general.

Lemma 1. *Let H^c be the complexification of the group H. Then H^c preserves the pencil (that is, the transformations in H^c map each circle in the pencil onto another circle of the same pencil -we say that they permute the circles in the pencil).*

PROOF. It can be shown (inspection!) that the group H is a one dimensional Lie group, hence commutative. So, using the exponential mapping, its element can be written in the form g_ξ, where ξ is a real parameter ($\xi \in \mathbb{R}$). Similarly the transformations in the complexification H^c will be written g_ζ, where ζ is a complex parameter ($\zeta = \xi + i\eta \in \mathbb{C}$). To fix the ideas, let us assume that we are in the hyperbolic case, denoting the points through which the circles go by p and q. (The other two cases are dealt with in a similar fashion.) Thus we have two equations of the type $g_\xi p = p$ and $g_\xi q = q$ ($\xi \in \mathbb{R}$). Then it is manifest that they remain true also after passing to the complexification (with ξ replaced by ζ). In other words, we have $g_\zeta p = p$ and $g_\zeta q = q$ ($\zeta = \xi + i\eta \in \mathbb{C}$). So if C is any circle passing through p and q, then its image $g_\zeta(C)$ under g_ζ is a circle which still passes through p and q and so belongs to the given pencil; but in general it is not the same circle ($g_\zeta(C) \neq C$).

Let us look at the three cases (hyperbolic, parabolic and elliptic) separately.

1. *Elliptic case.* Making a preliminary conformal transformation

we may pass to the normal form when it is question of concentric circles about the origin. The group H fixing any two of these circles, and thus all of them, consists of the maps $g_{e^{i\theta}} = e^{i\theta}a$ -rotation about the origin. Complexifying we get the transformations $g_\zeta a = \zeta a$, where we have put $\zeta = re^{i\theta}$ -rotations followed by dilation. (Note that here we made a passage from additive language to multiplicative language.)

2. *Parabolic case.* Now we may assume that we are dealing with straight lines parallel to the real axis -this is a pencil of degenerate circles. (In the case corresponding to the spaces G_k this can be achieved by applying the Bargmann transformation whereby Fock space G_k gets replaced by the Schrödinger space S_θ; see Introduction.) The maps in H consist of translations $a \mapsto a + \beta$ with β real. Complexifying yields the corresponding transformations with β complex. Note that in this limiting case the full group preserving the pencil is the 3-dimensional "$(\alpha a + \beta)$-group" with $\alpha \neq 0$ real, β complex.[8]

3. *Hyperbolic case.* As normal form we may use the straight lines through the origin. Then the transformations preserving the pencil are formally the same as in Case 1, $a \mapsto \zeta a$, the difference being that it is when we take the variable ζ real that we get the maps that leave invariant each element of the pencil (a degenerate circle).

REMARK 3. We note that H is compact precisely in the elliptic case.

Next we turn to the problem of the analytic description of the group H or H^c. Recall that if $\phi = (A, B, C)$ is the triple corresponding to a circle $\mathcal{C}$, we have already introduced the metric form

$$\langle \phi, \phi \rangle = AC - |B|^2 . \tag{2}$$

If we have one more circle $\mathcal{C}'$ corresponding to the triple $\phi' = (A', B', C')$, we obtain by polarization the inner product

$$\langle \phi, \phi' \rangle = \frac{1}{2}(AC' + CA') - \operatorname{Re} B\bar{B}' . \tag{2'}$$

EXAMPLE 1. A point circle can be identfied to the triple $\phi_a = (1, -a, |a|^2)$. Then (1) can, in view of (2′), be written as

$$\langle \phi, \phi_a \rangle = 0 .$$

8 Since the letter a is occupied by the variable, we cannot speak of the $(ax+b)$-group!

We can also rewrite our previous formula for norm of the function e_{ac} (see Section 1) as

$$\|e_{ac}\|^2_{(k,l)} = \exp\left(\frac{\mathrm{Re}\left(\frac{\partial\langle\phi,\phi_a\rangle}{\partial\bar{a}}\bar{c}^2\right) - (-\langle\phi,\phi\rangle)^{1/2}|c|^2}{-\langle\phi,\phi_a\rangle}\right)\left(\frac{\langle\phi,\phi_a\rangle}{(-\langle\phi,\phi\rangle)}\right)^{-1/2},$$

which perhaps looks more convincing; here $\phi = (1, l, |l|^2, -k^2)$.

Let us now fix a pencil of circles. To determine the corresponding group H, the latter being one dimensional and hence commutative, it suffices to determine its infinitesimal generator X (forming a basis for the Lie algebra $\mathfrak{h}$ of H). This is essentially an exercise in linear algebra.

The image of the pencil under the circle-to-point map $\mathcal{C} \mapsto \phi = (A, B, C)$ is, by what we have said, a line L in $\mathbb{PR}^3$. Let $\phi_0 = (A_0, B_0, C_0)$ and $\phi_1 = (A_1, B_1, C_1)$ be on L. We seek a linear map $\hat{X}$ on $\mathbb{R}^4$ which vanishes on the span of the vectors ϕ_0 and ϕ_1, and is skew-Hermitean with respect to the metric $\langle\phi,\phi\rangle$. Clearly, X is the inverse image of $\hat{X}$.

It is easily seen that $\hat{X}$ is given by the the condition

$$i\begin{vmatrix} A' & B' & \bar{B}' & C' \\ A & B & \bar{B} & C \\ A_0 & B_0 & \bar{B}_0 & C_0 \\ A_1 & B_1 & \bar{B}_1 & C_1 \end{vmatrix} = \langle\phi', \hat{X}\phi\rangle\,, \quad \text{where } \phi' = (A', B', C') \text{ (and } i^2 = -1).$$

Expanding the determinant and comparing with (2′) shows that

$$\hat{X} : (A, B, C) \mapsto i\left(-\begin{vmatrix} A & B & \bar{B} \\ A_0 & B_0 & \bar{B}_0 \\ A_1 & B_1 & \bar{B}_1 \end{vmatrix}, \begin{vmatrix} A & B & C \\ A_0 & B_0 & C_0 \\ A_1 & B_1 & C_1 \end{vmatrix}, \begin{vmatrix} B & \bar{B} & C \\ B_0 & \bar{B}_0 & C_0 \\ B_1 & \bar{B}_1 & C_1 \end{vmatrix}\right).$$

Putting $\phi^* = \hat{X}\phi = (A^*, B^*, C^*)$, we can write this, expanding the 3×3 determinants also, as

$$(3)\quad \begin{cases} A^* = i\left(-\begin{vmatrix} B_0 & \bar{B}_0 \\ B_1 & \bar{B}_1 \end{vmatrix} A + \begin{vmatrix} A_0 & \bar{B}_0 \\ A_1 & \bar{B}_1 \end{vmatrix} B - \begin{vmatrix} A_0 & B_0 \\ A_1 & B_1 \end{vmatrix} \bar{B} + 0\right), \\ B^* = i\left(-\begin{vmatrix} B_0 & C_0 \\ B_1 & C_1 \end{vmatrix} A - \begin{vmatrix} A_0 & C_0 \\ A_1 & C_1 \end{vmatrix} B + 0 + \begin{vmatrix} A_0 & B_0 \\ A_1 & B_1 \end{vmatrix} \bar{B}\right), \\ C^* = i\left(0 + \begin{vmatrix} \bar{B}_0 & C_0 \\ \bar{B}_1 & C_1 \end{vmatrix} B - \begin{vmatrix} B_0 & C_0 \\ B_1 & C_1 \end{vmatrix} \bar{B} + \begin{vmatrix} B_0 & \bar{B}_0 \\ B_1 & \bar{B}_1 \end{vmatrix} C\right). \end{cases}$$

That is, we have, in matrix form,

$$(4)\quad \hat{X} = i\begin{pmatrix} -\begin{vmatrix} B_0 & \bar{B}_0 \\ B_1 & \bar{B}_1 \end{vmatrix} & \begin{vmatrix} A_0 & \bar{B}_0 \\ A_1 & \bar{B}_1 \end{vmatrix} & -\begin{vmatrix} A_0 & B_0 \\ A_1 & B_1 \end{vmatrix} & 0 \\ -\begin{vmatrix} B_0 & C_0 \\ B_1 & C_1 \end{vmatrix} & -\begin{vmatrix} A_0 & C_0 \\ A_1 & C_1 \end{vmatrix} & 0 & \begin{vmatrix} A_0 & B_0 \\ A_1 & B_1 \end{vmatrix} \\ 0 & \begin{vmatrix} \bar{B}_0 & C_0 \\ \bar{B}_1 & C_1 \end{vmatrix} & -\begin{vmatrix} B_0 & C_0 \\ B_1 & C_1 \end{vmatrix} & \begin{vmatrix} B_0 & \bar{B}_0 \\ B_1 & \bar{B}_1 \end{vmatrix} \end{pmatrix}.$$

EXAMPLE 2 (The case of concentric circles). We can take $\phi_0 = (1, 0, 1)$, $\phi_1 = (1, 0, 2)$. Then $A^* = C^* = 0$, $B^* = B$. This corresponds to the circle transformations $B \mapsto e^{i\theta}B$, again induced by the point transformations $z \mapsto e^{i\theta}z$ (rotations about the origin). This we know, of course.

The map $\hat{X}$ is an element of the Lie algebra $\mathfrak{so}(1,3)$. Now we seek the corresponding element X in $\mathfrak{g}^c = \mathfrak{sl}(2,\mathbb{C})$.

First we work on the group level. Let $g = \left(\begin{smallmatrix} \alpha & \beta \\ \gamma & \delta \end{smallmatrix}\right)$ be in G^c. Then a circle $\mathcal{C}$ corresponding to the quadruple $\phi = (A, B, \bar{B}, C)$ is mapped into a circle $\mathcal{C}^*$ corresponding to the quadruple $\phi = (A^*, B^*, \bar{B}^*, C^*)$, where

$$\begin{aligned} A^* &= A\alpha\bar{\alpha} + B\gamma\bar{\alpha} + \bar{B}\alpha\bar{\gamma} + C\gamma\bar{\gamma}\,, \\ B^* &= A\beta\bar{\alpha} + B\delta\bar{\alpha} + \bar{B}\beta\bar{\gamma} + C\delta\bar{\gamma}\,, \\ C^* &= A\beta\bar{\beta} + B\delta\bar{\beta} + \bar{B}\beta\bar{\delta} + C\delta\bar{\delta}\,. \end{aligned}$$

Thus the point transformation g induces the circle transformation

$$\hat{g} = \begin{pmatrix} \alpha\bar{\alpha} & \gamma\bar{\alpha} & \alpha\bar{\gamma} & \gamma\bar{\gamma} \\ \beta\bar{\alpha} & \delta\bar{\alpha} & \beta\bar{\gamma} & \delta\bar{\gamma} \\ \alpha\bar{\beta} & \gamma\bar{\beta} & \alpha\bar{\delta} & \gamma\bar{\delta} \\ \beta\bar{\beta} & \delta\bar{\beta} & \beta\bar{\delta} & \delta\bar{\delta} \end{pmatrix}.$$

Passing to the infinitesimal (algebra) level we see that to the matrix

$$X = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \mathfrak{sl}(2,\mathbb{C})$$

there corresponds the matrix

$$(5)\qquad \hat{X} = \begin{pmatrix} 2\,\mathrm{Re}\,\alpha & \gamma & \bar{\gamma} & 0 \\ \beta & -i\,\mathrm{Im}\,\alpha & 0 & \bar{\gamma} \\ \bar{\beta} & 0 & i\,\mathrm{Im}\,\alpha & \gamma \\ 0 & \bar{\beta} & \beta & -2\,\mathrm{Re}\,\alpha \end{pmatrix} \in \mathfrak{so}(1,3).$$

(Here we use $\alpha + \delta = 0$, corresponding to $\alpha\delta - \beta\gamma = 1$.)

Now we compare the general formula (4) to (3). This gives in our case

$$X = i \begin{pmatrix} \begin{vmatrix} A_0 & C_0 \\ A_1 & C_1 \end{vmatrix} - \begin{vmatrix} B_0 & \bar{B}_0 \\ B_1 & \bar{B}_1 \end{vmatrix} & -\begin{vmatrix} B_0 & C_0 \\ B_1 & C_1 \end{vmatrix} \\ \begin{vmatrix} A_0 & \bar{B}_0 \\ A_1 & \bar{B}_1 \end{vmatrix} & -\begin{vmatrix} A_0 & C_0 \\ A_1 & C_1 \end{vmatrix} + \begin{vmatrix} B_0 & \bar{B}_0 \\ B_1 & \bar{B}_1 \end{vmatrix} \end{pmatrix} \tag{6}$$

This is the sought infinitesimal generator of the Lie algebra $\mathfrak{h}$.

We may summarize the preceding discussion as follows.

Lemma 2. *The Lie group H fixing the two circles C_0 and C_1 corresponding to the triple $\phi_0 = (A_0, B_0, C_0)$ and $\phi_1 = (A_1, B_1, C_1)$ is generated by the matrix given by formula* (6).

Let us give another example.

EXAMPLE 3. Consider the hyperbolic pencil of circles through the points 1 and -1. These circles correspond to the parameters $k = \sqrt{1+m^2}$, $l = im$ with m real. We may take $\phi_0 = (1, 0, -1)$ (unit circle), $\phi_1 = (0, i, 0)$ (real axis). Then (6) readily gives

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

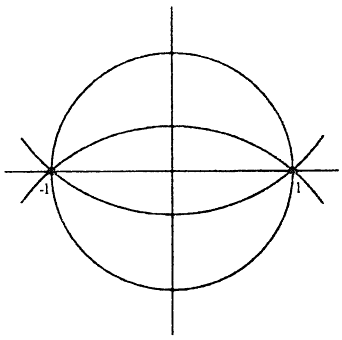
Thus integrating we get the transformations

$$g_\zeta = \begin{pmatrix} \cosh\zeta & \sinh\zeta \\ \sinh\zeta & \cosh\zeta \end{pmatrix}.$$

Each of the maps g_ζ preserves, if ζ is real, any of the circles of the pencil. If we let ζ assume complex values, we obtain transformations that permute the circles. If ζ is purely imaginary, $\zeta = i\eta$, then the image of the unit circle corresponds to

$$k = \frac{1}{\cos 2\eta} = \sec 2\eta\,, \quad l = i\,\frac{\sin 2\eta}{\cos 2\eta} = i\,\tan 2\eta\,, \qquad (m = \tan 2\eta)\,. \tag{7}$$

In particular, for $\eta = \pi/4$ the unit circle is mapped onto the real axis (Cayley transformation).



Now that we have a rather complete picture of the transformations permuting the circles of a given pencil, that is, of the complexification H^c of the group H of transformations fixing any two of them, it is possible also to answer the initial question of complex interpolation of two given spaces $F_{(k_0,l_0)}$ and $F_{(k_1,l_1)}$ corresponding to any two circles $\mathcal{C}_0 = C_{(k_0,l_0)}$ and $\mathcal{C}_1 = C_{(k_1,l_1)}$ in the pencil.

We let $D_0 = D_{(k_0,l_0)}$ and $D_0 = D_{(k_1,l_1)}$ be the corresponding disks (the disks bounding the circles and not containing the point at infinity). We assume that we have

$$D_0 \cap D_1 \neq \varnothing . \tag{8}$$

As was already recorded in Section 1, this implies that $F_{(k_0,l_0)} \cap F_{(k_1,l_1)} \neq \{0\}$. Changing somewhat the notation we may assume that the maps $g_{i\eta}$ ($\eta \in \mathbb{R}$) in H^c leave $\mathcal{C}_0$ and $\mathcal{C}_1$ invariant. We may also assume that $g_1(\mathcal{C}_1) = \mathcal{C}_0$. (This amounts to normalizing the group parameter.)

Lemma 3. *It is possible to choose g_1 such that $g_1(D_1) = D_0$.*

PROOF. By inspection. Expect in the hyperbolic case this is automatic. In the latter case we first choose g_1 to be minimal, that is, $g_1(\mathcal{C}_1) = \mathcal{C}_0$ but $g_\xi(\mathcal{C}_1) \neq \mathcal{C}_0$ for $0 < \xi < 1$. Then either $g_1(D_1) = D_0$ or else $g_1(D_1) = \tilde{D}_0$, where $\tilde{D}_0$ is the complementary disk $\tilde{D}_0 = S^2 \backslash \bar{D}_0$. In this case the group generated by g_1 must be compact. (On the other hand, the one generated by g_i equals H and is not compact *cf.* Remark 3.)

Therefore it must be periodic. If τ is the period, we can now achieve $g_1(D_1) = D_0$ replacing if necessary g_1 by $g_{1/2-\tau}$.

Corollary. *It follows that* $T_{g_1}(F_{(k_1,l_1)}) = F_{(k_0,l_0)}$ *and in particular that* $\|f\|_{(k_1,l_1)} = \|T_{g_1}f\|_{k_0}^{l_0}$ *for* $f \in F_{(k_1,l_1)}$.

For $\theta \in (0,1)$ let now the circle $\mathcal{C}_\theta$ be chosen in such a way that $g_\theta(\mathcal{C}_\theta) = \mathcal{C}_0$. We let further D_θ be the disk corresponding to the circle $\mathcal{C}_\theta$.

Lemma 4. *We have* $\infty \notin g_\theta(D_1)$, $\theta \in (0,1)$. *In particular, we have* $g_\theta(D_\theta) = D_0$.

PROOF. By a continuity argument. The elliptic and parabolic cases are quite obvious, because then the circles $\mathcal{C}_\theta$ lie all between $\mathcal{C}_0$ and $\mathcal{C}_1$. So let us again look at the hyperbolic case. In this case it is clear that the relation $g_\theta(D_\theta) = D_0$ holds true at least for θ close to 0. If the assertion were not true, then it is easy to see that for some particular value $\theta_0 \in (0,1)$ the corresponding circle $\mathcal{C}_{\theta_0}$ degenerates and becomes a line, the power line of our pencil (see Remark 3). But at that moment the corresponding disk degenerates into a halfplane. Continuing the parameter θ beyond the value θ_0 it is now easy to arrive at a contradiction, namely that $g_1(D_1) = \tilde{D}_0$, where again $\tilde{D}_0$ stands for the complementary disc.

We can now announce the following result.

Theorem 1. *Let* $F_{(k_0,l_0)}$ *and* $F_{(k_1,l_1)}$ *be the generalized Fock spaces corresponding to the circles* $C_{(k_0,l_0)}$ *and* $C_{(k_1,l_1)}$. *If* D_0 *and* D_1 *be the corresponding disks, we assume that* (8) *holds true. Let* g_ξ *be the one parameter group of conformal maps as defined above in the course of the discussion of Lemma 2 and 3. (In particular thus* $g_1(D_1) = D_0$.) *For* $0 < \theta < 1$ *define the circle* $C_{(k_\theta,l_\theta)}$ *by* $g_\theta(C_{(k_\theta,l_\theta)}) = C_{(k_0,l_0)}$. *Then we have the isometry*

$$[F_{(k_0,l_0)}, F_{(k_1,l_1)}]_\theta = F_{(k_\theta,l_\theta)}\,, \qquad 0 < \theta < 1\,. \tag{9}$$

PROOF. This follows from the general facts about complex interpolation which we recalled in the beginning of this section. In particular, the rôle of the operators Λ^ζ is now played by the maps T_{g_ζ}, as follows readily from Section 1, Theorem 1. The crucial thing is that for purely imaginary values of ζ these are unitary maps in $F_{(k_0,l_0)}$, $\|T_{g_{i\eta}}\| = 1$ for $\eta \in \mathbb{R}$. So there is really nothing to prove.

EXAMPLE 4. It is clear that Theorem 1 contains as special cases the results from [6] for $p = 2$ with the spaces F_k and G_k which were recalled in the Introduction. These are elliptic and parabolic cases respectively. A concrete example in a hyperbolic situation can easily be constructed at the hand of Example 3 *ultra*. Let us fix attention to the circles in the hyperbolic pencil there which lie in the upper halfplane, that is, if $k \geq 1$ is, as usual, the radius then the second parameter l is determined by $l = i\sqrt{1-k^2}$ (with the positive sign of the square root). We are thus lead to consider the family of spaces E_k of entire analytic functions f with the metric

$$\|f\|^2 = \frac{k^{1/2}}{\pi} \int_{\mathbb{C}} e^{-k|z|^2 - 2\sqrt{1-k^2}y^2} \, |f(z)|^2 \, dm(z) .$$

In agreement with our previous notation (*cf.* Introduction) we have in particular $E_1 = F_1 = F$ (our standard Foch space) and $E_0 = G_1$. Thus this connects the spaces F_1 and G_1. We conclude that we have the interpolation formula

$$[E_{k_0}, E_{k_1}]_\theta = E_{k_\theta} , \qquad 0 < \theta < 1 ,$$

where k_θ is obtained from k_0 and k_1 according to the following rule: if we write $k_0 = \sec 2\eta_0$ and $k_1 = \sec 2\eta_1$ then $k_\theta = \sec 2\eta$ with $\eta = (1-\theta)\eta_0 + \theta\eta_1$.

REMARK 5. The recepee for computing the "mean" of the parameters k_0 and k_1 is thus rather complicated in this case. That the rule has such a simple form in the case of the families F_k (geometric mean) and G_k (harmonic mean) is rather exceptional. In particular, the homogeneity is accounted for by the fact that the corresponding pencils are dilation invariant then. The phenomenon we initially set out to clarify in this paper has turned to be an exception!

So far we have only dealt with Hilbert spaces, that is, the problem of complex interpolation of the scale of generalized Fock spaces $F_{(k,l)}$. Now we pass to the corresponding problem for the Banach spaces $F^p_{(k,l)}$

($1 < p < \infty$). On a formal basis we expect that the obvious analogue of (9), *viz.* the interpolation formula,

$$[F^p_{(k_0,l_0)}, F^p_{(k_1,l_1)}]_\theta = F^p_{(k_\theta,l_\theta)}\,, \qquad 0 < \theta < 1\,, \tag{10}$$

to be true, perhaps not isometrically but at least up to an equivalence of norm. (For simplicity we keep the parameter p fixed taking $p_0 = p_1 = p$ and interpolate only k and l; how to treat the case $p_0 \neq p_1$ is indicated in [6].) The main difficulty is again to estimate the operator norm of $T_{g_{i\eta}}$, this time in the space $F^p_{(k_0,l_0)}$. It turns out that the different cases (elliptic, etc.) behave differently.

Let us first look at the elliptic case. Putting into play the map $V = V_0 : F^p_{(k_0,l_0)} \to F^p$, which by Section 1 is an isometry (only the case $p = 2$ was worked out there), we can reduce to the case when $\mathcal{C}_{(k_0,l_0)}$ is the unit circle $\mathcal{C}_{(1,0)}$. We recall from [9], Section 8, that the metaplectic group $\tilde{G}$ acts continuously on the spaces F^p, but this action is not isometric if $p \neq 2$. Indeed the operators T_g, $g \in \tilde{G}$, admit in F^p the following norm estimate:

$$\|T_g\|_p \approx |\delta|^{|1/p-1/2|}\,. \tag{11}$$

In the present elliptic case the one parameter group $g_{i\eta}$ is compact (cf. Remark 3). Therefore it follows from (11) that we have $\|T_{i\eta}\|_p \approx C$. So we are in business. We get thus back the result for the spaces F^p_k ([6, Theorem 9.3, the case $p_0 = p_1$]).

REMARK 6. In this situation we could have used instead of V_0 another more cleverly chosen Shale-Weil transformation reducing ourselves to the case when $\mathcal{C}_{(k_1,l_1)}$ is a concentric circle $\mathcal{C}_{(k,0)}$. Then we are back in the set up of [6].

Next, let us look look at the parabolic situation. It is then readily seen that the matrices $g_{i\eta}$ are conjugated to the matrices

$$\begin{pmatrix} 1 + i\frac{\eta}{2} & i\frac{\eta}{2} \\ -i\frac{\eta}{2} & 1 - i\frac{\eta}{2} \end{pmatrix}$$

by a fixed matrix. It follows then from (11) that now $\|T_{i\eta}\|_p \approx (1+|\eta|)^m$ for some number m, that is, we have power-like growth. We are again in business and have, in particular, essentially recovered the corresponding result for the spaces G^p_k ([6, formula 11.6, the case $p_0 = p_1$]).

Finally, we turn to the hyperbolic situation. From Example 3 it is seen that now the $g_{i\eta}$ are conjugated to the matrices

$$g_{i\eta} = \begin{pmatrix} \cosh\eta & \sinh\eta \\ \sinh\eta & \cosh\eta \end{pmatrix} .$$

But according to (11) this gives exponential growth of the norm and we must conclude that the general theory is not applicable. The different cases thus behave essentially differently. We may summarize the preceding discussion as follows.

Theorem 2. *We return to the set up of Theorem* 1, *replacing everywhere the space* $F_{(k,l)}$ *by* $F^p_{(k,l)}$, $p \geq 1$. *Then we have the interpolation formula* (10)

$$(10) \qquad [F^p_{(k_0,l_0)}, F^p_{(k_1,l_1)}]_\theta = F^p_{(k_\theta,l_\theta)}, \qquad 0 < \theta < 1,$$

which is an isomorphism (equality up to equivalence of norm), provided the pencil generated by the circles $C_{(k_0,l_0)}$ *and* $C_{(k_1,l_1)}$ *is either elliptic or parabolic. If however this pencil is hyperbolic the natural approach fails and we do not know whether* (10) *is true or not.*

3. Real interpolation. Multipliers.

Now we turn to real interpolation. The problem is thus to describe the real interpolation spaces between two given space $F^p_{(k_0,l_0)}$ and $F^p_{(k_1,l_1)}$. First we recall some salient facts about real interpolation in general (and we refer again to [2] for details).

If (A_0, A_1) is any Banach couple, one begins by defining the K-functional: for any element a in the sum $A_0 + A_1$ and any t with $0 < t < \infty$ one puts[8]

$$K(t,a) = \inf\{\|a_0\|_{A_0} + t\|a_1\|_{A_1}\},$$

[8] If X is any Banach space, we use a subsript X to designate the corresponding norm, thus writing $\|\cdot\|_X$.

where the infimum extends over all decompositions of a of the form $a = a_0 + a_1$ with $a_0 \in A_0$, $a_1 \in A_1$. One says that a belongs to the K-space $(A_0, A_1)_{\theta,q}$, where $0 < \theta < 1$ and $0 < q \leq \infty$, if and only if

$$\|a\|_{(A_0,A_1)_{\theta,q}} \stackrel{\text{def}}{=} \Big(\int_0^\infty (t^{-\theta} K(t,a))^q \, \frac{dt}{t} \Big)^{1/q} < \infty \, ;$$

if $q = \infty$ the left hand side of the inequality is interpreted as a supremum.[9] In order to obtain a concrete representation of the K-spaces one has to compute the K-functional, at least approximately.

If we are in the situation of an "operator pair" $(E, D(\Lambda))$ (see Section 2), it is natural to try to exploit the functional or spectral calculus associated with the operator Λ in question. More specifically, in some situations one can prove that one has an estimate of the type

$$K(t,a) \approx \|\varphi(t\Lambda)a\|_E \tag{1}$$

with a suitable scalar function φ. (In particular, the couple $(E, D(\Lambda))$ is thus "quasi-linearizable" in a certain technical sense.) Then one has

$$a \in (E, D(\Lambda))_{\theta,q} \Longleftrightarrow \Big(\int_0^\infty (t^{-\theta} \|\varphi(t\Lambda)a\|_E)^q \, \frac{dt}{t} \Big)^{1/q} < \infty \, . \tag{2}$$

Let us look at some special cases.

1) In the Hilbert case (*viz.* E = a Hilbert space, Λ a positive self-adjoint operator acting in E; *cf.* Section 2), there are plenty of such functions: any function φ defined on $(0,\infty)$ such that $\varphi(\lambda) \approx \min(1,\lambda)$ will do. With the aid of this one can prove that in case $q = 1/2$ indeed holds $(E, D(\Lambda))_{\theta,1/2} = D(\Lambda^\theta)$, up to equivalence of norm. (Indeed, there is now a canonical choice for the function φ: $\varphi(\lambda) = (1+\lambda^{-2})^{-2}$; with this choice one has even an isometry, provide the K-functional is replaced by what is known as the K_2-functional.) Thus, in this case, and in general only in this case, the two approaches -real and complex interpolation- produce the same result.

2) In the general case, it is natural to try to exploit the resolvent $R(t) = (1 + t\Lambda)^{-1}$ $(t > 0)$. If one has the estimate $\|R(t)\| \leq C$, with C independent of t, it is not very hard to show that (1) is fulfilled with

---

9 In the literature there is a J-functional and J-spaces, but these need not bother us here.

$\varphi(\lambda) = \lambda/(1+\lambda)$; operators Λ with this property are called *positive*.[10] Thus (2) in this case means that

$$(2') \qquad a \in (E, D(\Lambda))_{\theta,q} \Longleftrightarrow \Big(\int_0^\infty (t^{-\theta}\|R(t)\,t\Lambda a\|_E)^q \,\frac{dt}{t}\Big)^{1/q} < \infty\,.$$

3) If Λ is the generator of a semi-group of operators, it is again natural to use instead the semi-group of operators $G(t) = e^{t\Lambda}$. If one has $\|G(t)\| \le C$, with C independent of t, one speaks of a bounded semi-group and in this case one has the following result:

$$K(t,a) \approx \inf_{s\le t} \|G(s)a - a\|_E\,.$$

Although this is formally weaker than (1), it is nevertheless sufficient for establishing the desired analogue of (2), *viz.*

$$(2'') \quad a \in (E, D(\Lambda))_{\theta,q} \Longleftrightarrow \Big(\int_0^\infty (t^{-\theta}\|G(t)a - a\|_E)^q \,\frac{dt}{t}\Big)^{1/q} < \infty\,.$$

Let us mention that if Λ is the generator of a bounded semi-group of operators, then Λ is positive.

REMARK 1. Above we have summarized some classical results due to Lions, Grisvard, and others. Besides [2] we can also refer to the books Butzer-Behrens [3], Triebel [16]. Case 3) will not be used here, but we have now made mention of this case anyhow. (Perhaps somebody in the future might want to use semigroups in the Fock context ...)

After this long digression let us return to our generalized Fock spaces for good. More specifically, we are addressing ourselves to the problem of quasi-linearizability. The case $p = 2$ is essentially trivial, because as we have seen in 1) *ultra* in the Hilbert space case in general we can, in principle, even obtain an exact result. Therefore we proceed directly to the case of general p. (The problem of interpolation between two Fock spaces with different p's seems to be very hard; at least, it is not likely that one has a quasi-linearizable couple in that case.) So we are given two spaces $F^p_{(k_0,l_0)}$ and $F^p_{(k_1,l_1)}$ with $F^p_{(k_0,l_0)} \cap F^p_{(k_1,l_1)} \ne \{0\}$, assuming also that the corresponding disks have non-empty intersection, $D_{(k_0,l_0)} \cap D_{(k_1,l_1)} \ne \varnothing$. There is a natural candidate for the operator Λ,

10 For Hilbert spaces the two notions of positivity coincide.

namely the operator $T_1 = T_{g_1}$ constructed in Section 2. So the problem becomes to decide when operator T_1 is positive in the above sense, that is, when the operators $(1 + tT_1)^{-1}$ are uniformly bounded. Again this is basically a *multiplier problem* in the natural basis provided by the spectral theorem (if $p = 2$) for which T_1 comes in diagonal form. The nature of its solution depends on the type of the corresponding pencil of circle. Therefore we shall proceed by case by case study.

1. *Elliptic case.* Making a preliminary conformal mapping (see Section 1), one can put oneself in the situation of the two spaces $F^p = F_1^p = F_{(1,0)}^p$ and $F_k^p = F_{(k,0)}^p$, where we can assume, with no loss of generality, that $k > 1$. In this case we have $T_1 f(z) = f(k^{-1/2}z)$ (*cf.* Section 1, Corollary to Theorem 1). It will be expedient to write $\delta = k^{-1/2}$, so that $0 < \delta < 1$. In particular, we have then $T_1 : z^n \mapsto \delta^n z^n$, so a basis in which T_1 is in diagonal form is provided by the monomials $\{z^n\}$ $(n \neq 0)$. Moreover, we have as a consequence $R(t) : z^n \mapsto (1 + t\,\delta^n)^{-1} z^n$. This suggests to look quite generally at multiplier transforms on the space F^p.

Given a bounded function $\omega(n)$ defined on the set $\mathbb{N}$ of non-negative integers, $\mathbb{N} = \{0, 1, 2, \dots\}$, we define an operator R_ω by setting

$$R_\omega f(z) = \sum_{n=0}^{\infty} \omega(n)\, a_n z^n$$

whenever $f \in F^p$ has the expansion

$$f(z) = \sum_{n=0}^{\infty} a_n z^n .$$

It follows that R_ω maps each basis vector z^n into a multiple of itself, $R_\omega : z^n \mapsto \omega(n)\, z^n$. We are interested in the boundedness of R_ω on F^p. First we establish an easy transference result which reduces the study of R_ω to the study of Fourier multipliers.

Let us define the operator $\tilde{R}_\omega$ on $L^p(\mathbb{T})$, where $\mathbb{T}$ is the unit circle parametrized by arc length θ, by

$$\tilde{R}_\omega f(\theta) = \sum_{n=0}^{\infty} \omega(n) a_n e^{in\theta}$$

whenever $f \in L^p(\mathbb{T})$ has the Fourier expansion

$$f(z) = \sum_{n=0}^{\infty} a_n e^{in\theta} .$$

Proposition 1. *The operator R_ω is bounded on F^p whenever $\tilde{R}_\omega$ is bounded on $L^p = L^p(\mathbb{T})$.*

PROOF. We have

$$\int_{\mathbb{C}} |R_\omega f(z)|^p \, e^{-p|z|^2/2} \, dm(z) = \int_0^\infty \int_0^{2\pi} |R_\omega f(re^{i\theta})|^p \, d\theta \, e^{-pr^2/2} \, r \, dr$$
$$= \int_0^\infty \|\tilde{R}_\omega f_r(\,\cdot\,)\|_p^p \, e^{-pr^2/2} \, r \, dr \,,$$

where we have written $f_r(\theta) = f(re^{i\theta})$. If we assume that $\tilde{R}_\omega : L^p \to L^p$ is bounded, then

$$\|\tilde{R}_\omega f_r\|_p^p \le C \int_0^{2\pi} \Big| \sum_{n=0}^\infty a_n r^n e^{in\theta} \Big|^p d\theta \,.$$

It therefore follows that

$$\int_{\mathbb{C}} |R_\omega f(z)|^p \, e^{-p|z|^2/2} \, dm(z) \le C \int_{\mathbb{C}} |f(z)|^p \, e^{-p|z|^2/2} \, dm(z) \,.$$

We can now prove the following theorem; we assume now that ω is defined for all $\xi \ge 0$.

Proposition 2. *Let ω a bounded function on $(0, \infty)$ such that*

$$\int_0^\infty |\omega'(\xi)| \, d\xi < \infty \,.$$

Then for $1 < p < \infty$ the multiplier R_ω is bounded on F^p.

PROOF. In view of the previous Proposition 1 it is enough to show that $\widetilde{R}_\omega$ is bounded on L^p, $1 < p < \infty$. Again, using another transference result between multipliers for the Fourier series and multipliers for the Fourier transform (see [14, Chapter VII, Theorem 3.8]) it is enough to show that the operator

$$S_\omega f(x) = \int_{-\infty}^\infty e^{ix\xi} \, \omega(\xi) \, \hat{f}(\xi) \, d\xi$$

is bounded on $L^p(\mathbb{R})$. But this follows from the Marcinkiewicz multiplier theorem (*cf.* [4, Proposition 4.1]) under the above hypothesis on ω.

REMARK 2. The proof of the Marcinkiewicz multiplier theorem actually gives a bound for the norm of R_ω on F^p. In fact,

$$\|R_\omega f\|_p \leq \left(\|\omega\|_\infty + \int_0^\infty |\omega'(\xi)|\, d\xi\right) \|f\|_p\,, \qquad 1 < p < \infty\,.$$

In applications one encounters multipliers of the form $\omega_t(\xi) = \varphi(t\,\psi(\xi))$, $t > 0$, where ψ is a positive monotone function. If we assume that φ is bounded and that $\int_0^\infty |\varphi'(\xi)|\, d\xi < \infty$, then the operator R_{ω_t} will be bounded on F^p with a bound independent of t, $t > 0$. Indeed, as $\omega_t'(\xi) = \varphi'(t\,\psi(\xi))\, t\,\psi'(\xi)$ we have $\|\omega\|_\infty + \int_0^\infty |\omega'(\xi)|\, d\xi \leq C$ with C independent of t, so that Proposition 2 is applicable (see Remark 1). In particular, taking $\psi(\xi) = \delta^\xi$, where $0 < \delta < 1$, we see that the operators

$$R_t f(z) = \sum_{n=0}^{\infty} a_n\,(1 + t\,\delta^n)^{-1}\, z^n \tag{3}$$

are uniformly bounded on F^p, $1 < p < \infty$. In view of the general remarks in the beginning of this Section we have thus established the following theorem, which thus in particular settles in part a question left over in [6].

Theorem 1. *Let $1 < p < \infty$, $0 < q \leq \infty$. Then the operators R_t as defined in* (3) *(with $\delta = k^{-1/2}$) are uniformly bounded in $F^p = F_1^p$ and for $f \in F^p = F^p + F_k^p$ we have*

$$f \in (F^p, F_k^p)_{\theta,q} \quad \Longleftrightarrow \quad \left(\int_0^\infty (t^{-\theta}\,\|R_t f\|_{F^p})^q\, \frac{dt}{t}\right)^{1/q} < \infty\,.$$

REMARK 3. Thus the pair (F^p, F_k^p) is quasi-linearizable in the technical sense.

There remain the cases $p = 1$, $p < 1$ and $p = \infty$. Here we shall only consider the former case. It is not hard to see that for each $t > 0$ the operator R_t is bounded on F^1. But unfortunately they are not uniformly bounded. This indicates that it is very unlikely that the Banach couple (F^1, F_k^1) is quasi-linearizable, again indicating that no result of the type of (2) can be true in this case. For reference, we state the result as a theorem.

Theorem 2. *The operators R_t are not uniformly bounded on F^1.*

The proof is by contradiction, but as it is rather long we prefer to split it up into several steps in the form Propositions 3-5 below. Let us begin by explaining the basic underlying idea.

Suppose that the operators R_t are uniformly bounded on F^1. Taking $f(z) = e_{0,c}(z) = e^{cz}$ ($c \in \mathbb{C}$) and noting that

$$\int_{\mathbb{C}} |e_{0,c}(z)|\, e^{-|z|^2/2}\, dm(z) = e^{|c|^2/2},$$

we see that one must have the estimate

$$\int_{\mathbb{C}} |R_t e_{0,c}(z)|\, e^{-|z|^2/2}\, dm(z) \leq C\, e^{|c|^2/2}. \tag{4}$$

(As the functions serve as "atoms" in the space F^1, one sees that, conversely, (4) implies uniform boundedness; *cf.* [6, Theorem 8.1.]) In what follows we shall show that (4) cannot hold true, proving the theorem. In doing this we may as well assume that c is a positive number.

Let us set

$$f(z,t,\delta) = \sum_{n=0}^{\infty} \frac{z^n}{n!} \frac{1}{1+t\,\delta^n}, \qquad t > 0, \quad 0 < \delta < 1. \tag{5}$$

We wish thus to test the hypothesis

$$\int_{\mathbb{C}} |f(cz,t,\delta)|\, e^{-|z|^2/2}\, dm(z) \leq C\, e^{c^2/2} \tag{6}$$

for $c > 0$ and C independent of t and c.

First we replace $f(z,t,\delta)$ by chain of simpler functions ending up with the function $j(z,t,\delta)$ in formula (13) below and then proceed to the study of that function. Our first intermediary result is thus the following.

Proposition 3. *If inequality* (6) *holds with $f(cz,t,\delta)$, then it holds with $f(cz,t,\delta)$ replaced by $j(cz,t,\delta)$.*

PROOF. Performing a Mellin transformation with respect to t gives the representation

$$f(z,t,\delta) = \frac{1}{2i}\int_{\gamma-i\infty}^{\gamma+i\infty} \frac{t^{-\lambda}e^{z\delta^{-\lambda}}}{\sin\pi\lambda}\,d\lambda \tag{7}$$

for $\lambda = \gamma + i\mu$, $0 < \gamma < 1$. It follows from (7) that

$$f(z,t,\delta) - f(z,t^{-1},\delta) = -\int_{-\infty}^{\infty} \frac{\sin(\mu\ln t)}{\sinh\pi\mu}\, e^{z\delta^{-i\mu}}\,d\mu\,, \tag{8}$$

where we also have moved the path of integration to the left so that it passes through the origin. (Note that the identity (8) is a special case of a more general formula stated in [13].) The essential feature of the right hand member of (8) is the factor $e^{z\delta^{-i\mu}}$ for which

$$\int_{\mathbb{C}} |e^{cz\delta^{-i\mu}}|\, e^{-|z|^2/2}\,dm(z) = 2\pi\, e^{c^2/2}\,. \tag{9}$$

Introducing the "c-norm" of an entire function f by

$$\|f\|_c = \int_{\mathbb{C}} |f(cz)|\, e^{|z|^2/2}\,dm(z)$$

(in the notation of the Introduction it is up to a factor just the norm in the space $F^1_{1/\sqrt{c}}$), we see that the right hand side of (8) is a superposition of functions all having the c-norm equal to $e^{c^2/2}$. Now, t is going to be large so the term $f(z,t^{-1},\delta)$ can easily be seen to satisfy (6). Thus the right hand member of (8) is essentially a representation of $f(z,t,\delta)$. Due to rapid convergence at infinity of the integral in (8) we can pass to the function

$$g(z,t,\delta) = \int_{-A}^{A} \frac{\sin(\mu\ln t)}{\sinh\pi\mu}\, e^{z\delta^{-i\mu}}\,d\mu \tag{10}$$

with A fixed > 0. This again may be replaced by

$$\begin{aligned} h(z,t,\delta) &= \int_{-A}^{A} \frac{\sin(\mu\ln t)}{\mu}\, e^{z\delta^{-i\mu}}\,d\mu \\ &= \int_{-A\ln(1/\delta)}^{A\ln(1/\delta)} \frac{\sin(\mu\ln t/\ln(1/\delta))}{\mu}\, e^{ze^{i\mu}}\,d\mu\,, \end{aligned} \tag{11}$$

where in the last equality we have made the change of variable $\mu \mapsto \mu/\ln(1/\delta)$. Choosing A so that $A\ln(1/\delta) = \pi$ we get

$$h(z,t,\delta) = \int_{-\pi}^{\pi} \frac{\sin(\mu \ln t/\ln(1/\delta))}{\mu} e^{ze^{i\mu}} \, d\mu. \tag{12}$$

Finally, we pass to the function

$$j(z,t,\delta) = \int_{-\pi}^{\pi} \frac{\sin(\mu \ln t/\ln(1/\delta))}{\sin(\mu/2)} e^{ze^{i\mu}} \, d\mu. \tag{13}$$

It is easy to estimate the error thereby committed but we shall not enter into details. Thus, *testing* (6) *for* $f(z,t,\delta)$ *is completely equivalent to testing it for* $j(z,t,\delta)$.

Next we establish the following result.

Proposition 4. *Let* φ *be a bounded radial function in* $\mathbb{C}$: $\varphi(z) = \sum_k \hat{\varphi}(k)\, e^{-ik\theta}$. *Then one has the identity*

$$\begin{aligned} &\int_{\mathbb{C}} j(cz, t_N, \delta)\, \varphi(z)\, e^{-|z|^2/2} \, dm(z) \\ &\qquad = (2\pi)^2 \sum_{n=0}^{N} \frac{c^n}{n!} 2^{n/2} \Gamma(\frac{n}{2}+1)\, \hat{\varphi}(-n), \end{aligned} \tag{14}$$

where $\log t_N = (N+1/2)\log(1/\delta)$.

PROOF. For every bounded function φ we clearly have

$$\begin{aligned} &\int_{\mathbb{C}} j(z,t,\delta)\, \varphi(z)\, e^{-|z|^2/2} \, dm(z) \\ &= \Big(\int_{-\pi}^{\pi} \frac{\sin(\mu \ln t/\ln(1/\delta))}{\sin(\mu/2)} \, d\mu\Big) \Big(\int_{\mathbb{C}} e^{cze^{i\mu}} \varphi(z)\, e^{-|z|^2/2} \, dm(z)\Big). \end{aligned} \tag{15}$$

Setting $z = re^{i\theta}$, writing

$$\varphi(z) = \varphi(re^{i\theta}) = \sum_{n=-\infty}^{\infty} \hat{\varphi}(n,r)\, e^{in\theta},$$

(for our purposes we may without loss of generality assume that this is a finite sum) and using the expansion

$$e^{cze^{i\mu}} = \sum_{n=0}^{\infty} \frac{c^n}{n!} z^n e^{in\mu} = \sum_{n=0}^{\infty} \frac{c^n}{n!} r^n e^{in\theta} e^{in\mu},$$

we see that the inner integral in (15) equals

$$2\pi \sum_{n=0}^{\infty} \frac{c^n}{n!} e^{in\mu} \int_0^\infty \hat{\varphi}(-n, r)\, r^n e^{-r^2/2} r\, dr. \tag{16}$$

Now choose $t = t_N$ in (15) so that $\ln t / \ln(1/\delta) = N + 1/2$, where N is a positive integer at our disposal. Using the well-known identity

$$1 + 2\sum_{n=1}^{N} \cos n\mu = \frac{\sin(N+1/2)\mu}{\sin(\mu/2)},$$

we get inserting (16) into (15)

$$\begin{aligned} &\int_{\mathbb{C}} j(cz, t_N, \delta)\, \varphi(z)\, e^{-|z|^2/2}\, dm(z) \\ &\qquad = (2\pi)^2 \sum_{n=0}^{N} \frac{c^n}{n!} \int_0^\infty \hat{\varphi}(-n, r)\, r^n e^{-r^2/2} r\, dr. \end{aligned} \tag{17}$$

Specializing to the case $\varphi(e^{i\theta})$, that is, $\hat{\varphi}(n) = \hat{\varphi}(n, r)$ independent of r, we get

$$\begin{aligned} \int_0^\infty \hat{\varphi}(-n, r)\, r^n e^{-r^2/2} r\, dr &= \hat{\varphi}(-n) \int_0^\infty r^n e^{-r^2/2} r\, dr \\ &\underset{r^2=2s}{=} \hat{\varphi}(-n) \int_0^\infty (2s)^{n/2} e^{-s}\, ds \\ &= \hat{\varphi}(-n)\, 2^{n/2} \int_0^\infty s^{n/2} e^{-s}\, ds \\ &= \hat{\varphi}(-n)\, 2^{n/2}\, \Gamma(\tfrac{n}{2} + 1). \end{aligned}$$

Summing up we obtain in this case in view of (17) the desired formula, *viz.* (14).

Finally, we prove the following result.

Proposition 5. *There exist trigonometric polynomials $\varphi = \varphi_N$ of degree N uniformly bounded in z and N, such that*

$$\int_{\mathbb{C}} j(cz, t_N, \delta)\, \varphi_N(z)\, e^{-|z|^2/2}\, dm(z) \geq C\, e^{c^2/2} \log N\,, \tag{18}$$

where $C > 0$.

PROOF. We choose φ as a pure cosine series, *i.e.* $\hat{\varphi}(n) = \hat{\varphi}(-n)$:

$$\varphi = a_0 + 2 \sum_{n=1}^{\infty} c_n \cos n\theta\,. \tag{19}$$

Even more, we shall take φ to be the Fejér function

$$\begin{aligned} \varphi_N(z) = {} & \frac{\cos\theta}{2N-1} + \frac{\cos 2\theta}{2N-3} + \cdots + \frac{\cos N\theta}{1} \\ & - \frac{\cos(N+1)\theta}{1} - \frac{\cos(N+2)\theta}{2} - \cdots - \frac{\cos 2N\theta}{2N-1}\,; \end{aligned} \tag{20}$$

cf. [15, p. 416, 13.41], where it is proved that $\sup_{\mathbb{C}} |\varphi_N(z)| \leq C$. Comparing (19) and (20) we see that

$$\hat{\varphi}_N(0) = 0\,, \quad 2\,\hat{\varphi}_N(n) = \frac{1}{2N-(2n-1)}\,, \qquad (1 \leq n \leq N)\,,$$

i.e. (18) becomes

$$\begin{aligned} & \int_{\mathbb{C}} j(cz, t_N, \delta)\, \varphi_N(z)\, e^{-|z|^2/2}\, dm(z) \\ & \qquad = \frac{(2\pi)^2}{2} \sum_{n=1}^{N} \frac{c^n}{n!}\, \frac{\Gamma(\frac{n}{2}+1)\, 2^{n/2}}{2N-(2n-1)}\,. \end{aligned} \tag{21}$$

By Stirling's formula we have

$$\Gamma(x+1) = \left(\frac{x}{e}\right)^x \sqrt{2\pi x}\,(1+o(1))\,, \qquad \text{as } x \to \infty\,.$$

For large n we therefore get

$$\Gamma\left(\frac{n}{2}+1\right) = \left(\frac{n}{2e}\right)^{n/2} \sqrt{2\pi \frac{n}{2}}\,(1+o(1))\,;$$
$$n! = \Gamma(n+1) = \left(\frac{n}{e}\right)^{n} \sqrt{2\pi n}\,(1+o(1))\,,$$

so that

$$\frac{\Gamma(\frac{n}{2}+1)\,2^{n/2}}{n!} = \frac{\left(\frac{n}{e}\right)^{n/2}\sqrt{\pi n}}{\left(\frac{n}{e}\right)^{n}\sqrt{2\pi n}}\,(1+o(1)) = \left(\frac{e}{n}\right)^{n/2}\frac{1}{\sqrt{2}}\,(1+o(1))\,.$$

Hence, for n large,

$$\frac{c^n}{n!}\,\Gamma\left(\frac{n}{2}+1\right)2^{n/2} = \left(\frac{c^2 e}{n}\right)^{n/2}\frac{1}{\sqrt{2}}\,(1+o(1))\,. \tag{22}$$

For $N > M$ both large we now get by (21) and (22) (if we use $1+o(1) \geq 1/2$)

$$\begin{aligned} &\int_{\mathbb{C}} j(cz, t_N, \delta)\,\varphi_N(z)\,e^{-|z|^2/2}\,dm(z) \\ &\qquad \geq (2\pi)^2\,\frac{1}{4\sqrt{2}}\sum_{n=M}^{N}\left(\frac{c^2 e}{n}\right)^{n/2}\frac{1}{2N-(2n-1)}\,. \end{aligned} \tag{23}$$

Now, take c as a large positive integer and choose $N = c^2$, $M = c^2 - c$ in (23). Then the sum in the right hand side of (23) becomes

$$\begin{aligned} \sum_{n=M}^{N}\cdots &= \sum_{n=c^2-c}^{c^2}\left(\frac{c^2 e}{n}\right)^{n/2}\frac{1}{2c^2-(2n-1)} \\ &\underset{\substack{n=c^2-j\\ 0\leq j\leq c}}{=} \sum_{j=0}^{c}\left(\frac{c^2 e}{c^2-j}\right)^{(c^2-j)/2}\frac{1}{2j+1} \\ &= \sum_{j=0}^{c} e^{(c^2-j)/2}\underbrace{\left(1+\frac{j}{c^2-j}\right)^{(c^2-j)/2}}_{\approx e^{j/2}}\frac{1}{2j+1} \end{aligned}$$

$$\geq C\, e^{c^2/2} \sum_{j=0}^{c} \frac{1}{2j+1} \geq C_1\, e^{c^2/2} \log c\,,$$

proving (18).

It is clear that from Proposition 4 we get a contradiction to the hypothesis (6) (or (4)). Thereby we have proved also Theorem 3.

Next we treat the parabolic case. As this case is rather parallel to the elliptic one, we shall not be so detailed.

2. *Parabolic case.* We can again put ourselves in a model situation, namely, when we have the two spaces $G^p = G_1^p = F^p_{(1,1)}$ and $G_k^p = F^p_{(k,k)}$, where we without loss of generality can assume that $k < 1$ (or $k > 1$, whatever we like). Thus this situation corresponds to the pencil of circles tangent to the imaginary axis at the origin. We have to put the corresponding operators T_{g_ζ} on diagonal form. To this end we first take $p = 2$ so that we are dealing with Hilbert spaces. (As usual, we then drop the superscript p in the notation for the spaces.) Then we have the norms

$$\|f\|_G^2 = \int_{\mathbb{C}} |f(z)|^2\, e^{-2y^2}\, dm(z)\,,$$
$$\|f\|_{G_k}^2 = \int_{\mathbb{C}} |f(z)|^2\, e^{-2ky^2}\, dm(z)\,.$$

Introducing the Fourier transform

$$\hat{f}(\lambda) = \int_{-\infty}^{\infty} e^{-i\lambda x} f(x)\, dx\,,$$

then the right hand sides of the previous formulae become

$$C \int_{-\infty}^{\infty} |\hat{f}(\lambda)|^2\, e^{\lambda^2/2}\, d\lambda\,,$$
$$C \int_{-\infty}^{\infty} |\hat{f}(\lambda)|^2\, e^{\lambda^2/2k}\, d\lambda\,.$$

From here it is readily seen that

$$\widehat{g_\zeta f}(\lambda) = \hat{f}(\lambda)\, e^{\lambda^2 \zeta (1/k-1)/4}\,. \tag{24}$$

In terms of Moebius transformations this corresponds to the following. First it is seen that to $f(z) \mapsto \hat{f}(\lambda)\, e^{\lambda^2/4}$ there corresponds the map

$$a \mapsto b = \frac{a+2}{2a} = \frac{1}{a} + \frac{1}{2}\,.$$

(It is the correspondence $a \mapsto 1/a$ that accounts for the Fourier transform; as before (see Section 1) the symbol a is used the designate a generic point of $\mathbb{C}$.) This gives the isomorphism $G \approx S_0$ (Schrödinger space). In the same way, to $f(z) \mapsto \hat{f}(\lambda)\, e^{\lambda^2/4k}$ there corresponds the map

$$a \mapsto b_1 = \frac{a+2k}{2ka} = \frac{1}{a} + \frac{1}{2k}\,.$$

Elimination of a between the two equations yields

$$\frac{1}{b_1} = \frac{1}{b} + \frac{1}{2}\Big(\frac{1}{k} - 1\Big)\,.$$

REMARK 3. Note that in particular this implies that for $\theta \in (0,1)$

$$\|g_\theta\|_G^2 = \int_{\mathbb{C}} |\hat{f}(k)|^2\, e^{-2ky^2}\, dm(z)\,,$$

where $1/k_\theta = 1-\theta+\theta/k$. This is the result from [JPR], which we wrote down already in the Introduction (see also Theorem 2 in Section 2).

From (24) we see now that our question is about the multiplier

$$\frac{1}{1+t\,e^{(1/k-1)\lambda^2/2}}\,.$$

Imitating what we have done already in the elliptic case (this Section *infra*, we associate, quite generally, with any suitable locally integrable function $\Omega(\lambda)$ on the real line $\mathbb{R}$ a multiplier transform P_Ω (it is the analogue of the previous R_ω) defined on the space G^p by the formula

$$\widehat{P_\Omega f}(\lambda) = \Omega(\lambda)\, \hat{f}(\lambda)\,.$$

We denote by $\tilde{P}_\Omega$ the same transformation when consider on the Lebesgue space $L_p(\mathbb{R})$. Then we have the following analogue of Proposition 1.

Proposition 6. *The operator P_Ω is bounded on G^p whenever $\tilde{P}_\Omega$ is bounded on $L^p(\mathbb{R})$.*

PROOF. The proof parallels the proof of Proposition 1. Assuming that $\tilde{P}_\Omega$ is bounded on $L^p(\mathbb{R})$ we obtain

$$\begin{aligned}\int_{\mathbb{C}} |P_\Omega f(z)|^p e^{-2y^2}\, dm(z) &= \int_{-\infty}^{\infty} e^{-2y^2} \left(\int_{-\infty}^{\infty} |P_\Omega f_y(x)|^p\, dx\right) dy \\ &\leq C \int_{-\infty}^{\infty} e^{-2y^2} \left(\int_{-\infty}^{\infty} |f(x+iy)|^p\, dx\right) dy \\ &= C \int_{\mathbb{C}} |f(z)|^p e^{-2y^2}\, dm(z).\end{aligned}$$

Thus P_Ω is bounded on G^p.

Similarly, it is easy to carry over Proposition 2 and from there one derives the expected analogue of Theorem 1 (in the statement of the theorem replace everywhere F^p and F_p^k by G^p and G_p^k respectively), but according to our above promise we omit the details.

There remains the hyperbolic case. But there is a difficulty hidden which we have not been able to overcome ...

3. *Hyperbolic case.* In a model situation the pencil might consist of lines through a point, say, the origin -the configuration that perhaps first comes to our mind. So if $p = 2$ we are in a situation when the spaces to be interpolated are Schrödinger spaces S_θ. The trouble is that in the general case $p \neq 2$ we do not possess a workable analogue of these spaces; in particular, we know of no counterpart of the Corollary to Theorem 1 in Section 1.

So let us instead take as model the case of the spaces F^p and $F^p_{(k,l)}$ with $k = \sec\theta$, $l = i\tan\theta$. The pencil then consists of the circles through the points 0 and $-2 - 2i\tan\theta$. (It is easy to see that this is basically the set up of Section 2, Example 3 shifted the amount 2 to the left.)

It is convenient to introduce, quite generally, the notation $H_k^p = F^p_{(k,l)}$ if k and l are related by the previous relation $k = \sec\theta$, $l = i\tan\theta$. The corresponding circle with center at the point $-1 - i\tan\theta$ may be written C_θ. We are going to follow our usual abuses putting $H = H_1$ and also dropping the superscript p if $p = 2$.

So we want to interpolate between H^p and a fixed space H^p_k. The first thing is to determine the corresponding diagonalizing map (taking temporarily $p = 2$). In the case of H we have $\theta = 0$ and the circle $\mathcal{C}_0$ is the unit circle so transforming the space H into a Schrödinger space S_0 goes as before (in the parabolic case) via the Moebius map

$$a \mapsto b = \frac{a+2}{2a} = \frac{1}{a} + \frac{1}{2}\,.$$

In the case of H_k and the circle $\mathcal{C}_\theta$ we first observe that the parameter θ has a geometric meaning: it is the angle between this circle and $\mathcal{C}_0$. Hence the previous map must essentially be composed by a rotation by an angle θ. This leads to the map

$$a \mapsto b' = \frac{e^{i\theta}a + 2\sec\theta}{2\,e^{i\theta}\sec\theta a} = \frac{1}{a} + \frac{1}{2}\,.$$

Elimination between the last two identities, as in the parabolic case, yields

$$b' = \frac{1}{2}\sin\theta + \left(b - \frac{1}{2}\right)e^{-i\theta} = \frac{e^{-i\theta}}{a} + \frac{1}{2}\,i\sin\theta\,,$$

where we have used Euler's formula $e^{-i\theta} = \cos\theta - i\sin\theta$. From this we get the map

$$f(z) \mapsto \hat{f}(e^{i\theta}\lambda)\,e^{i\sin\theta\,\lambda^2/4}\,. \tag{25}$$

(This would correspond to $\hat{f}(\lambda) \mapsto \hat{f}(\lambda)\,e^{\zeta(1/k-1)\lambda^2/4}$ in the parabolic case; see formula (24).) But, in view of the appearence, of the factor $e^{i\theta}$ in front of the variable λ in the second half of (25), the map given by this formula does not give a simultanous diagonalization of the group operators $\zeta \mapsto g_\zeta$. To get a *bona fide* diagonalization we must first apply a Mellin type transformation to the Fourier transform $\hat{f}(\lambda)$. But then we do not have anymore any such simple transfer results as the above Propositions 1 and 6. Therefore we stop here hoping to be able to resume this thread on a future occasion. Concluding let us only remark (as a conjecture!) that perhaps it is the case that the presence of a hyperbolic pencil does not imply quasi-linearizability.

4. Concluding remarks.

In this section we consider some left-overs from the previous sections, also complementing some points in [9]. We begin by some easy observations on the Orlicz case.

4.1. On Orlicz-Fock spaces.

Recall that, generally speaking, a measurable function f on some measure space X endowed with a measure μ is said to belong to the Orlicz space $L^\Phi = L^\Phi(X,\mu)$, where Φ is an Orlicz function (in particular, increasing), if

$$\int_X \Phi\left(\frac{|f(x)|}{\alpha}\right) d\mu(x) < \infty \tag{1}$$

for some number $\alpha > 0$. It is well-known that L^Φ is a (quasi-)Banach space with the (quasi-)norm of f defined by

$$\|f\|_{L^\Phi} = \inf \alpha\,, \tag{2}$$

where α ranges over all numbers satisfying (1). If Φ is convex, we can drop the affix "quasi" everywhere. If $\Phi(u) = u^p$, $p > 0$, then we get back the Lebesgue space L^p.

This suggests (*cf.* Section 1) to introduce in our case the Orlicz-Fock spaces $F^\Phi_{(k,l)}$ as the space of entire analytic functions f in $\mathbb{C}$ such that the function $|f(z)|\,e^{-(k|z|^2-\mathrm{Re}(lz^2))/2}$ belongs to L^Φ when $X = \mathbb{C}$ and $\mu = m$ (Euclidean measure). Again it is clear that if $\Phi(u) = u^p$, $p > 0$, they reduce to the spaces $F^p_{(k,l)}$. Also in the general case they should have similar properties as the spaces $F^\Phi_{(k,l)}$. The norm of f in $F^\Phi_{(k,l)}$ is the induced norm and will be written $\|f\|_{(k,l);\Phi}$. (In the special case $k = 1$, $l = 0$ we allow us to drop these indices in the notation.)

We shall limit ourselves to calculating the norm of the corresponding Gauss-Weierstrass functions e_{ac} in one simple case, namely when Φ is of the form

$$\Phi(u) = \sum_{p=1}^{\infty} A_p u^p \qquad \text{with} \;\; A_p \geq 0\,,$$

for simplicity's sake also taking $k = 1$, $l = 0$.

REMARK 1. More generally we could have allowed functions admitting an integral representation (Mellin transform):

$$\Phi(u) = \int_0^\infty A(p)\, u^p\, dp \qquad \text{with } A(p) \geq 0\,.$$

Using a formula established in [9, Section 2] for the norm of e_{ac} in the space F^p, we then find

$$\begin{aligned}
&\int_{\mathbb{C}} \Phi\left(\frac{|e_{ac}(z)|\, e^{-|z|^2/2}}{\alpha}\right)^p dm(z) \\
&\quad = \sum_{p=1}^{\infty} A_p \int_{\mathbb{C}} \frac{(|e_{a,c}(z)\,|e^{-|z|^2/2})^p}{\alpha^p}\, dm(z) \\
(3)\qquad &\quad = \sum_{p=1}^{\infty} A_p \left(\frac{p}{2}\right)^{-1} \frac{\exp\left(\frac{p}{2}\,\frac{\operatorname{Re} a\bar{c}^2 + |c|^2}{1-|a|^2}\right)}{\alpha^p}\, \frac{1}{(1-|a|^2)^{1/2}} \\
&\quad = \Phi_1\left(\frac{\exp\left(\frac{1}{2}\,\frac{\operatorname{Re} a\bar{c}^2 + |c|^2}{1-|a|^2}\right)}{\alpha^p}\right) \frac{1}{(1-|a|^2)^{1/2}}\,,
\end{aligned}$$

where we in the last step have introduced the notation

$$\Phi_1(u) = \sum_{p=1}^{\infty} \left(\frac{p}{2}\right)^{-1} A_p u^p\,.$$

We now take account of (2) letting α tend to $\|e_{ac}\|_\Phi$. Then we end up with the formula

$$\|e_{ac}\|_\Phi = \exp\left(\frac{1}{2}\,\frac{\operatorname{Re} a\bar{c}^2 + |c|^2}{1-|a|^2}\right) \frac{1}{\Phi_1^{-1}((1-|a|^2)^{1/2})} \tag{4}$$

In particular, we draw the conclusion that $e_{a,c} \in F^\Phi_{(k,l)}$ if and only if $|a| < 1$. Obviously, a similar result must hold true for general k and l as well.

4.2. Fock space in the case $0 < p < 1$.

Now we return to the case of Lebesgue spaces L^p but for a change take $p < 1$. (Thus we are leaving the realm of Banach spaces.) Also we take again $k = 1$, $l = 0$. In this situation we have the following obvious generalizing of a corresponding result in [9, Section 8] for $p \geq 1$.[11]

Theorem 1. *The metaplectic group $\tilde{G} = \mathrm{Mp}(2, \mathbb{R})$ acts on the space F^p, $0 < p < 1$. Indeed, if $g \in \tilde{G}$ then we have the following estimate for the opearator norm of T_g in F^p:*

$$\|T_g\|_p \approx \frac{1}{(1 - |g0|^2)^{(1/p - 1/2)/2}} \,. \tag{5}$$

PROOF. Let f be in F^p. Let us write $\|f\|_p = \|f\|_{F_p}$ for the (quasi-) norm in F_p; we use this notation even for $p \geq 1$. Let us further put $e_c(z) = e_{0,c}(z) = e^{cz}$ (exponential function). In view of Wallstén's theorem [17] (atomic decomposition in F_p, $0 < p < 1$; generalization of the corresponding result for F_1 in [6, Theorem 8.1]) we can write

$$f = \sum_k \lambda_k \frac{e_{c_k}}{\|e_{c_k}\|_2} \qquad \text{with} \quad \sum_k |\lambda_k|^p < \infty \,,$$

for some sequence of complex numbers $\{c_k\}$. For $g \in \tilde{G}$ we now obtain

$$T_g f = \sum_k \lambda_k \frac{T_g e_{c_k}}{\|e_{c_k}\|_2} \,.$$

For each index k we have ([9], Section 3)

$$T_g e_{c_k} = e_{g0, c_k/\delta}\, \delta^{-1/2}\, e^{-\gamma c_k^2/2\delta} \,.$$

In particular, we have

$$\|T_g e_{c_k}\|_p = \|e_{g0,c_k/\delta}\|_p\, |\delta|^{-1/2} \exp\left(- \mathrm{Re}\, \frac{\gamma c_k^2}{2\delta} \right)$$
$$= \left(\frac{p}{2}\right)^{-1/p} \exp\left(\frac{1}{2} \, \frac{\mathrm{Re}\, \dfrac{\beta\, \bar{c}_k^2}{\delta\, |\delta|^2} + \dfrac{|c_k|^2}{\delta^2}}{1 - \dfrac{|\beta|^2}{|\delta|^2}} \right)$$

[11] This theorem again was used already in Section 2.

$$\cdot \left(1 - \frac{|\beta|^2}{|\delta|^2}\right)^{1/2p} |\delta|^{-1/2} \exp\left(-\frac{1}{2}\operatorname{Re}\frac{\gamma c_k^2}{\delta}\right)$$

$$= \left(\frac{p}{2}\right)^{-1/p} |\delta|^{1/p-1/2} e^{|c_k|^2/2},$$

while

$$\|e_{c_k}\|_2 = e^{|c_k|^2/2}.$$

It follows that

$$\|T_g f\|_p^p \le 2\sum_k |\lambda_k|^p \left(\frac{\|T_g e_{c_k}\|_p}{\|e_{c_k}\|_2}\right)^p \le 2\left(\frac{p}{2}\right)^{-1} |\delta|^{1-p/2} \sum_k |\lambda_k|^p < \infty,$$

whence $T_g f \in F_p$. As

$$g0 = \frac{\beta}{\delta}, \qquad 1 - |g0|^2 = 1 - \frac{|\beta|^2}{|\delta|^2} = \frac{1}{|\delta|^2},$$

it likewise follows that inequality (5) is true.

References.

[1] Blaschke, W., *Vorlesungen über Differentialgeometrie und geometrische Grundlagen von Einsteins Relativitätstheorie. III. Differentialgeometrie der Kreise und Kugeln.* Grundlehren **29**. Springer, 1929.

[2] Bergh, J., Löfström, J., *Interpolation spaces. An Introduction.* Grundlehren **223**. Springer-Verlag, 1976.

[3] Butzer, P. L., Berens, H., *Semi-groups of operators and approximation.* Grundlehren **145**. Springer-Verlag, 1967.

[4] Davis, C., Chang, A., *Lectures on Bochner-Riesz means.* London Math. Soc. Lecture Notes **114**. Cambridge University Press, 1987.

[5] Howe, R., The oscillator semigroup. In *Mathematical heritage of Hermann Weyl, Proc. Symp. Pure Math.* **43** (1988), 61-132.

[6] Janson, S., Peetre, J., Rochberg, R., Hankel forms and the Fock space. *Revista Mat. Iberoamericana* **3** (1987), 61-138.

[7] Klein, F., *Vorlesungen über nicht-euklidische Geometrie.* Grundlehren **26**. Springer-Verlag, 1928. (Nachdruck, 1968.)

[8] Kobayashi, S., Nomizu, K., *Foundations of differential geometry* 1-2. Interscience, 1963/69.

[9] Peetre, J., Some calculations related to Fock space and the Shale-Weil representation. *Int. Equations Operator Theory* **12** (1989), 343-359.

[10] Peetre, J., A class of kernels related to the inequalities of Beckner and Nelson. In *A tribute to Åke Pleijel*, 171-210. Uppsala, 1980.

[11] Peetre, J., The periodic Fock bundle. *Algebra i Analiz* **3** (1992), 134-154. *St. Petersburg Math. J.* **3** (1992), 1069-1088.

[12] Peetre, J., The Bellaterra connection. *Pub. Mat.* **37** (1993), 165-176.

[13] Peetre, J., Sjölin, A quasi-linearizable pair connected with 3-line theorems. *Complex Variables* **19** (1992), 143-149.

[14] Stein, E., Weiss, G., *Introduction to Fourier analysis on Euclidean spaces.* Princeton Univ. Press, 1971.

[15] Titchmarsh, E. C., *The theory of functions.* (Second edition.) Oxford Univ. Press, 1952.

[16] Triebel, H., *Interpolation theory. Function spaces. Differential operators.* Deutscher Verlag der Wissenschaften, 1978.

[17] Wallstén, R., The S^p-criterion for Hankel forms on the Fock space. *Math. Scand.* **64** (1989), 123-152.

Recibido: 23 de diciembre de 1.994
Revisado: 2 de junio de 1.995

Jaak Peetre
Matematiska institutionen
Lunds Universitet
Box 118,
S-221 00 Lund, SWEDEN
jaak@maths.lth.se

Sundaram Thangavelu†
Stat-Math Division
Indian Statistical Institute
R. V. College Post
Bangalore-560 059, INDIA
veluma@isibang.ernet.in

Nils-Olof Wallin
Matematiska institutionen
Lunds universitet
Box 118,
S-221 00 Lund, SWEDEN
nils-olof,wallin@maths.lth.se

† Sponsored by the "Civilingenjör Gustaf Sigurd Magnusons fond för främjande av vetenskapen inom ämnet matematik" of the Royal Swedish Academy of Sciences (Kungl. Vetenskapsakademien). He also wishes to thank Eila and Jaak Peetre for the immense hospitality showered upon him during his visit to Lund.

Revista Matemática Iberoamericana
Vol. 12, N.° 1, 1996

On fractional differentiation and integration on spaces of homogeneous type

A. Eduardo Gatto, Carlos Segovia and Stephen Vági

Abstract. In this paper we define derivatives of fractional order on spaces of homogeneous type by generalizing a classical formula for the fractional powers of the Laplacean [S1], [S2], [SZ] and introducing suitable quasidistances related to an approximation of the identity. We define integration of fractional order as in [GV] but using quasidistances related to the approximation of the identity mentioned before.

We show that these operators act on Lipschitz spaces as in the classical cases. We prove that the composition T_α of a fractional integral I_α and a fractional derivative D_α of the same order and its transpose (a fractional derivative composed with a fractional integral of the same order) are Calderón-Zygmund operators. We also prove that for small order α, T_α is an invertible operator in L^2. In order to prove that T_α is invertible we obtain Nahmod type representations for I_α and D_α and then we follow the method of her thesis [N1], [N2].

1. Definitions and statement of the main results.

In this paper (X, δ, μ) will be a space of homogeneous type which is normal and of order γ, $0 < \gamma \leq 1$, and such that $\mu(\{x\}) = 0$ for all x in X, and $\mu(X) = \infty$.

We recall that a *space of homogeneous type* consists of a set X, a quasidistance δ, *i.e.* a function $\delta : X \times X \to [0,\infty)$ that satisfies

$$(1.1) \qquad \delta(x,y) = 0 \quad \text{if and only if} \quad x = y\,,$$
$$(1.2) \qquad \delta(x,y) = \delta(y,x)\,, \quad \text{for every } x \text{ and } y \text{ in } X\,,$$

there is a positive constant κ such that

$$(1.3) \qquad \delta(x,y) \le \kappa\,(\delta(x,z) + \delta(z,y))$$

for every x, y and z in X, and a measure μ defined on a σ-algebra of subsets of X which contains the open sets of X and the balls $B_r(x) = \{y : \delta(x,y) < r\}$ and satisfies the doubling condition: there exists a positive constant A such that for every x in X and every $r > 0$, $0 < \mu(B_{2r}(x)) \le A\,\mu(B_r(x))$. If X has more than one element, as in this paper, the constant κ in (1.3) cannot be less than 1.

A space of homogeneous type is *normal* if there are positive constants A_1 and A_2 such that for all x in X

$$(1.4) \qquad A_1\, r \le \mu(B_r(x)) \le A_2\, r\,, \qquad \text{for all } r > 0\,.$$

Two quasidistances δ and ρ are said to be *equivalent*, $\rho \cong \delta$, if there exist positive constants c_1 and c_2 such that for all x, y in X

$$(1.5) \qquad c_1\, \delta(x,y) \le \rho(x,y) \le c_2\, \delta(x,y)\,.$$

It is easy to see that if (X, δ, μ) satisfies (1.4) then so does (X, ρ, μ).

A space of homogeneous type is of *order* γ, $0 < \gamma \le 1$ if there is a positive constant M such that for all x, x', y in X

$$(1.6) \qquad |\delta(x,y) - \delta(x',y)| \le M\, \delta^\gamma(x,x')\,(\delta(x,y) + \delta(x',y))^{1-\gamma}\,.$$

It is shown in [MS] that in any space of homogeneous type there is a topologically equivalent quasidistance δ that satisfies (1.4) and (1.6).

For $0 < \beta \le \gamma$, $\mathrm{Lip}\,(\beta)$ will denote the space of complex valued functions f such that for all x and y in X

$$(1.7) \qquad |f(x) - f(y)| \le C\, \delta^\beta(x,y)$$

holds with a constant C independent of x and y. The norm of an element f of $\mathrm{Lip}\,(\beta)$ is the infimum of the constants C in (1.7). Given a ball B, $C_0^\beta(B)$ will denote the space of functions f in $\mathrm{Lip}\,(\beta)$ with compact support in B. We shall say that f belongs to C_0^β if f belongs to $C_0^\beta(B)$ for some B. The space C_0^β is the inductive limit of the Banach spaces $C_0^\beta(B)$ with the inductive limit topology and $(C_0^\beta)'$ will denote the space of all continuous linear functionals on C_0^β.

Let $s(x,y,t)$ be a symmetric approximation to the identity of the type introduced by Coifman, see Section 2. Let $-\infty < \alpha < 1$,we define $\delta_\alpha : X \times X \to [0,\infty)$ by

$$\delta_\alpha(x,y) = \Big(\int_0^\infty t^{\alpha-1}\, s(x,y,t)\, dt\Big)^{1/(\alpha-1)}, \qquad \text{for } x \neq y, \tag{1.8}$$

and

$$\delta_\alpha(x,y) = 0, \qquad \text{for } x = y.$$

We shall see in Section 2, Lemma 2.2, that for each α, δ_α is a quasidistance equivalent to δ, and it satisfies (1.6). Note that (X, δ_α, μ) is a normal space of order γ.

For $0 < \alpha < \gamma$ the fractional derivative of order α of f in $\mathrm{Lip}\,(\beta) \cap L^\infty$, $\alpha < \beta \leq \gamma$ is defined by

$$D_\alpha f(x) = \int_X \frac{f(y) - f(x)}{\delta_{-\alpha}^{\alpha+1}(x,y)}\, d\mu(y). \tag{1.9}$$

The above definition extends the classical formula for functions on $\mathbb{R}^n$,

$$D_\alpha f(x) = \lim_{\varepsilon\to 0} \int_{|y|\geq\varepsilon} \frac{f(x+y) - f(x)}{|y|^{n+\alpha}}\, dy.$$

For f sufficiently restricted and $0 < \alpha < 2$, one has $D_\alpha f = c_\alpha\,(-\Delta)^{\alpha/2} f$, where Δ is the Laplacean [S1], [S2], [S3].

For $0 < \alpha < 1$, the fractional integral of order α of f in $\mathrm{Lip}\,(\beta) \cap L^1$ is defined by

$$I_\alpha f(x) = \int_X \frac{f(y)}{\delta_\alpha^{1-\alpha}(x,y)}\, d\mu(y). \tag{1.10}$$

The definitions of D_α and I_α can be extended to $\mathrm{Lip}\,(\beta)$ for the same β as above. This requires the following modification similar to the one needed to define singular integrals on L^∞:

$$\widetilde{D}_\alpha f(x) = \int_X \Big(\frac{f(y) - f(x)}{\delta_{-\alpha}^{1+\alpha}(x,y)} - \frac{f(y) - f(x_0)}{\delta_{-\alpha}^{1+\alpha}(x_0,y)}\Big)\, d\mu(y), \tag{1.11}$$

and

$$(1.12) \qquad \widetilde{I}_\alpha f(x) = \int_X f(y) \Big(\frac{1}{\delta_\alpha^{1-\alpha}(x,y)} - \frac{1}{\delta_\alpha^{1-\alpha}(x_0,y)} \Big) \, d\mu(y) \,,$$

where x_0 is a fixed but arbitrary point of X. It will be shown in Theorem 1.1 and Theorem 1.2 that $\widetilde{D}_\alpha f(x)$ and $\widetilde{I}_\alpha f(x)$ converge absolutely for all x and therefore changing x_0 in the definitions above results in adding a constant. We show in Section 2, Lemma 2.4, that δ_α , for $0 < \alpha < \gamma$, has the cancellation property:

$$\int_X (\delta_\alpha^{\alpha-1}(x,y) - \delta_\alpha^{\alpha-1}(x',y)) \, d\mu(y) = 0 \,.$$

In [GV] it was shown that for fractional integrals defined with a quasidistance which has the above properties the classical theorems on boundedness on L^p, BMO, Lip$\,(\beta)$ and H^p hold. For the sake of completeness we prove the result for Lipschitz spaces in Theorem 1.1. See [GGW], [GV].

We recall the definition of a singular integral operator as given in [DJS] and [S3]. Let $\Omega = X \times X \setminus \Delta$ where Δ is the diagonal of $X \times X$. A continuous function $K : \Omega \to \mathbb{C}$ is a standard kernel if there exist a number η, $0 < \eta \leq 1$, and constants $\nu > 1$ and $c > 0$ such that

$$(1.13) \qquad |K(x,y)| \leq \frac{c}{\delta(x,y)} \qquad \text{for } (x,y) \text{ in } \Omega \,,$$

and for $\nu \, \delta(x,y) < \delta(x,z)$ we have

$$(1.14) \qquad |K(x,z) - K(y,z)| \leq c \, \frac{\delta^\eta(x,y)}{\delta^{1+\eta}(x,z)} \,,$$

and

$$(1.15) \qquad |K(z,x) - K(z,y)| \leq c \, \frac{\delta^\eta(x,y)}{\delta^{1+\eta}(x,z)} \,.$$

A *singular integral operator* is a continuous linear operator $T : C_0^\beta \to (C_0^\beta)'$ associated with a standard kernel K in the following sense:

$$\langle Tf, g \rangle = \int_X \int_X K(x,y) \, g(x) \, f(y) \, d\mu(x) \, d\mu(y) \,,$$

for all $f, g \in C_0^\beta$ with disjoint supports, and where $\langle Tf, g\rangle$ denotes the evaluation of Tf on g.

A singular integral operator is called a *Calderón-Zygmund operator* if it can be extended to a continuous operator from L^2 to L^2.

The *transpose* ${}^tT$ of a singular integral operator T is defined by

$$\langle {}^tTf, g\rangle = \langle Tg, f\rangle\,,$$

for all $f, g \in C_0^\beta$, $0 < \beta \leq \gamma$.

The function $s(x,y,t)$ introduced before, is continuously differentiable in t. Let

$$q(x,y,t) = t\,\frac{\partial}{\partial t} s(x,y,t) \tag{1.16}$$

and set

$$-Q_t f(x) = \int_X q(x,y,t)\, f(y)\, d\mu(y)\,. \tag{1.17}$$

In this paper the letter c will denote a constant, not necessarily the same in different occurrences.

We can now state our main results.

Theorem 1.1. *Let* $0 < \alpha < \beta \leq \gamma$.

a) *If* $f \in \mathrm{Lip}(\beta) \cap L^1$ *then* $I_\alpha f(x)$ *converges absolutely for all* x *and there is a constant* c *independent of* f *such that*

$$\|I_\alpha f\|_{\mathrm{Lip}(\alpha+\beta)} \leq c\,\|f\|_{\mathrm{Lip}(\beta)}\,.$$

b) *If* $f \in \mathrm{Lip}(\beta)$, *then* $\widetilde{I}_\alpha f(x)$ *converges absolutely for all* x, *and there is a constant* c *independent of* f *such that*

$$\|\widetilde{I}_\alpha f\|_{\mathrm{Lip}(\alpha+\beta)} \leq c\,\|f\|_{\mathrm{Lip}(\beta)}\,.$$

c) *If* $f \in \mathrm{Lip}(\beta) \cap L^1$ *then* $\widetilde{I}_\alpha f$ *defines the same class as* $I_\alpha f$ *in* $\mathrm{Lip}(\alpha+\beta)$.

Theorem 1.2. *Let* $0 < \alpha < \beta \leq \gamma$.

a) *If* $f \in \mathrm{Lip}(\beta) \cap L^\infty$ *then* $D_\alpha f(x)$ *converges absolutely for all* x *and there is a constant* c *independent of* f *such that*

$$\|D_\alpha f\|_{\mathrm{Lip}(\beta-\alpha)} \leq c\,\|f\|_{\mathrm{Lip}(\beta)}\,.$$

b) *If* $f \in \mathrm{Lip}(\beta)$ *then* $\widetilde{D}_\alpha f(x)$ *converges absolutely for all* x *and there is a constant* c *independent of* f *such that*

$$\left\|\widetilde{D}_\alpha f\right\|_{\mathrm{Lip}(\beta-\alpha)} \leq c\,\|f\|_{\mathrm{Lip}(\beta)}\,.$$

c) *If* $f \in \mathrm{Lip}(\beta) \cap L^\infty$ *then* $\widetilde{D}_\alpha f$ *defines the same class as* $D_\alpha f$ *in* $\mathrm{Lip}(\beta - \alpha)$.

For similar classical results see [Z, Chapter XII].

Theorem 1.3. *Let* $0 < \alpha < \gamma$, *then* $T_\alpha = D_\alpha I_\alpha$ *is a singular integral operator with associated kernel*

$$K(x,y) = \int_X \frac{1}{\delta_{-\alpha}^{1+\alpha}(x,t)} \left(\frac{1}{\delta_\alpha^{1-\alpha}(y,t)} - \frac{1}{\delta_\alpha^{1-\alpha}(x,y)}\right) d\mu(t)\,. \tag{1.18}$$

Theorem 1.4. *Let* $0 < \alpha < \gamma$, *then* $T_\alpha = D_\alpha I_\alpha$ *is a Calderón-Zygmund operator.*

Theorem 1.5. *Let* $S_\alpha = I_\alpha D_\alpha$, *then* $S_\alpha f = {}^tT_\alpha f$ *for every* f *in* C_o^β, *with* $0 < \alpha < \beta \leq \gamma$, *and* S_α *is a Calderón-Zygmund operator.*

Theorem 1.6. *If* $Q_t(f)$ *is the operator defined by* (1.17) *then the following representation formulas hold pointwise everywhere and in the weak sense:*

$$\alpha\, I_\alpha f = \int_0^\infty t^\alpha\, Q_t(f)\,\frac{dt}{t}\,, \tag{1.19}$$

for f *in* $\mathrm{Lip}(\beta) \cap L^1$, $0 < \alpha$, $\alpha + \beta \leq \gamma$, *and*

$$-\alpha\, D_\alpha f = \int_0^\infty t^{-\alpha}\, Q_t(f)\,\frac{dt}{t}\,, \tag{1.20}$$

for f *in* $\mathrm{Lip}(\beta) \cap L^\infty$, $0 < \alpha < \beta \leq \gamma$.

The following theorem extends a result obtained by A. R. Nahmod in her Thesis [N2].

Theorem 1.7 *There exists* α_0, $0 < \alpha_0 < \gamma$, *such that for* $0 < \alpha < \alpha_0$ *the operator* T_α *as defined in Theorem* 1.3 *has a bounded inverse in* L^2.

2. Lemmas needed for the proofs of Theorems 1.1 through 1.5.

The first lemma states the properties of a Coifman type approximation to the identity. These properties are well known, see [DJS], and therefore the proofs will be omitted.

Let $h \geq 0$ be a C^∞ function on $[0,\infty)$ such that $h(r) = 1$ for $0 \leq r \leq 1/2$, and $h(r) = 0$ for $r \geq 2$. For $f \in L^1_{\text{loc}}(X)$ and $t > 0$ set

$$T_t f(x) = \frac{1}{t} \int_X h\Big(\frac{\delta(x,y)}{t}\Big) f(y)\, d\mu(y),$$

$$M_t f(x) = \frac{1}{(T_t 1)(x)} f(x) = \varphi(x,t)\, f(x),$$

$$V_t f(x) = \frac{1}{T_t\Big(\dfrac{1}{T_t 1}\Big)(x)} f(x) = \psi(x,t)\, f(x).$$

Now define S_t by

$$S_t = M_t\, T_t\, V_t\, T_t\, M_t\ ,$$

then

$$S_t f(x) = \int_X s(x,y,t)\, f(y)\, d\mu(y),$$

where

$$s(x,y,t) = \frac{\varphi(x,t)\,\varphi(y,t)}{t^2} \int_X h\Big(\frac{\delta(x,u)}{t}\Big)\, h\Big(\frac{\delta(y,u)}{t}\Big)\, \psi(u,t)\, d\mu(u).$$

Lemma 2.1. *There exist positive constants b_1, b_2, c_1, c_2, and c_3 independent of x, y, and t such that*

i) $s(x,y,t) = s(y,x,t)$ *for all x,y in X and $t > 0$,*

ii) $|s(x,y,t)| \leq c_1/t$ *for all x,y in X and $t > 0$, $s(x,y,t) = 0$ if $\delta(x,y) > b_1\, t$, and $c_2/t < s(x,y,t)$ if $\delta(x,y) < b_2\, t$,*

iii) $|s(x,y,t) - s(x',y,t)| < c_3\, \delta^\gamma(x,x')/t^{1+\gamma}$ *for all x, x' and y in X, and $t > 0$,*

iv) $\displaystyle\int s(x,y,t)\, d\mu(y) = 1$ *for all x in X, and $t > 0$,*

v) $s(x,y,t)$ *is continuously differentiable with respect to* t.

Lemma 2.2. *For each* α, $-\infty < \alpha < 1$, *the function* δ_α, *defined in* (1.8) *is a quasidistance equivalent to* δ *and it satisfies* (1.6).

PROOF. We shall prove first that there are positive constants c'_α and c''_α such that for all x, y in X

$$c'_\alpha\, \delta(x,y) \le \delta_\alpha(x,y) \le c''_\alpha\, \delta(x,y)\,.$$

Using the properties of $s(x,y,t)$ stated in Lemma 2.1, we have that $s(x,y,t) = 0$ if $\delta(x,y) > b_1\, t$. Then

$$\delta_\alpha^{\alpha-1}(x,y) = \int_{\delta(x,y)/b_1}^{\infty} t^{\alpha-1}\, s(x,y,t)\, dt\,.$$

On the other hand $\|s(\cdot,\cdot,t)\|_\infty \le c_1/t$, and therefore

$$\delta_\alpha^{\alpha-1}(x,y) \le c_1 \int_{\delta(x,y)/b_1}^{\infty} t^{\alpha-2}\, dt = \frac{c_1}{1-\alpha}\, b_1^{1-\alpha}\, \delta^{\alpha-1}(x,y)\,.$$

Raising this inequality to the power $1/(\alpha-1)$ we obtain the first inequality of (2.1).

To obtain the second inequality of (2.1) note that $s(x,y,t) \ge c_2/t$ if $\delta(x,y) < b_2\, t$, hence by (2.2)

$$\delta_\alpha^{\alpha-1}(x,y) \ge \int_{\delta(x,y)/b_2}^{\infty} t^{\alpha-1}\, \frac{c_2}{t}\, dt = \frac{c_2}{1-\alpha}\, b_2^{1-\alpha}\, \delta^{\alpha-1}(x,y)\,.$$

Raising this inequality to the power $1/(\alpha-1)$ we conclude the proof of (2.1).

The fact that $\delta_\alpha(x,y)$ is a quasidistance follows from the definition, property i) of $s(x,y,t)$ and (2.1). We will denote by κ_α the constant in the inequality (1.3) for δ_α.

We will show now that δ_α satisfies (1.6). If $\delta_\alpha(x,y) = 0$ then $x = y$ and $\delta_\alpha(x',y) = \delta_\alpha(x,x')$ and

$$|\delta_\alpha(x,y) - \delta_\alpha(x',y)| = \delta_\alpha(x,x') = \delta_\alpha^\gamma(x,x')\,(\delta_\alpha(x,y) + \delta_\alpha(x',y))^{1-\gamma}\,.$$

Similarly when $\delta_\alpha(x',y) = 0$ we get the estimate above.

Assume now that $\delta_\alpha(x,y) \neq 0$ and $\delta_\alpha(x',y) \neq 0$. Let

$$a = \frac{1}{b_1} \min\{\delta_\alpha(x,y), \delta_\alpha(x',y)\},$$

then by property ii) of Lemma 2.1

$$\begin{aligned}
&|\delta_\alpha(x,y) - \delta_\alpha(x',y)| \\
&= \left| \Big(\int_a^\infty t^{\alpha-1} s(x,y,t)\, dt \Big)^{1/(\alpha-1)} - \Big(\int_a^\infty t^{\alpha-1} s(x',y,t)\, dt \Big)^{1/(\alpha-1)} \right| \\
&\quad \leq \Big(\int_a^\infty t^{\alpha-1} \, |s(x',y,t) + \theta(s(x,y,t) - s(x',y,t))| \, dt \Big)^{(2-\alpha)/(\alpha-1)} \\
&\qquad \cdot \Big(\int_a^\infty t^{\alpha-1} \, |s(x,y,t) - s(x',y,t)| \, dt \Big),
\end{aligned}$$

with $0 < \theta < 1$. Using ii) and iii) of Lemma 2.1 we can majorize the last estimate by

$$\begin{aligned}
\Big(c \int_a^\infty t^{\alpha-2}\, dt \Big)^{(2-\alpha)/(\alpha-1)} & \Big(\int_a^\infty t^{\alpha-\gamma-2}\, c\, \delta_\alpha^\gamma(x,x')\, dt \Big) \\
& \leq c\, \delta_\alpha^\gamma(x,x')\, a^{1-\gamma} \\
& \leq c\, \delta_\alpha^\gamma(x,x')\, (\delta_\alpha(x,y) + \delta_\alpha(x',y))^{1-\gamma}.
\end{aligned}$$

This concludes the proof of the lemma.

Lemma 2.3. *Let $\alpha < 1$ and $k_\alpha > \kappa_\alpha$. There exists a positive constant C_{k_α} such that*

$$|\delta_\alpha^{\alpha-1}(x,y) - \delta_\alpha^{\alpha-1}(x',y)| \leq C_{k_\alpha}\, \delta_\alpha^\gamma(x,x')^\gamma\, \delta_\alpha^{\alpha-1-\gamma}(x,y),$$

for all x, x', y in X such that

$$k_\alpha\, \delta_\alpha(x,x') \leq \delta_\alpha(x,y).$$

The exponent γ is the order of the space.

This result follows from property (1.6), it was proved in [GV] for $k_\alpha = 2\,\kappa_\alpha$, the proof for $k_\alpha > \kappa_\alpha$ is similar.

Lemma 2.4. (Cancellation property of order $\alpha - 1$). *Let $0 < \alpha < \gamma$, then*

$$\int_X \left(\delta_\alpha^{\alpha-1}(x,y) - \delta_\alpha^{\alpha-1}(x',y)\right) d\mu(y) = 0\,,$$

for any x, x' in X.

PROOF. We show first that

$$\int_X \int_0^\infty t^{\alpha-1} \left|s(x,y,t) - s(x'y,t)\right| d\mu(y)\, dt < \infty\,.$$

We have

$$\int_X \int_0^1 t^{\alpha-1} \left|s(x,y,t) - s(x'y,t)\right| d\mu(y)\, dt \leq 2 \int_0^1 t^{\alpha-1}\, dt < \infty\,.$$

To estimate $\int_X \int_1^\infty t^\alpha \left|s(x,y,t) - s(x',y,t)\right| d\mu(y)\, dt$, observe that the functions $s(x,\cdot,t)$ are supported in balls of radius $b_1 t$, also by iii) of Lemma 2.1 we have

$$\left|s(x,y,t) - s(x',y,t)\right| \leq c_3 \frac{\delta^\gamma(x,x')}{t^{1+\gamma}}\,.$$

Therefore using normality the double integral is majorized by

$$\int_1^\infty t^{\alpha-1} \frac{\delta^\gamma(x,x')\, c\, t}{t^{1+\gamma}}\, dt \leq c\, \delta^\gamma(x,x') \int_1^\infty \frac{dt}{t^{1+\gamma-\alpha}} < \infty\,.$$

Since

$$\int_X \left(\delta_\alpha^{\alpha-1}(x,y) - \delta_\alpha^{\alpha-1}(x',y)\right) d\mu(y) = \int_X \int_0^\infty t^{\alpha-1} \left(s(x,y,t) - s(x',y,t)\right) dt\, d\mu(y)\,,$$

by changing the order of integration and using v) of Lemma 2.1 we obtain that the integral is zero.

Lemma 2.5. *Let $x \in X$ and $r > 0$. Then*

$$\int_{\delta(x,y)<r} \frac{1}{\delta^\lambda(x,y)}\, d\mu(y) \leq c\, r^{-\lambda+1}\,, \qquad \textit{for } \lambda < 1\,,$$

and

$$\int_{\delta(x,y)\geq r} \frac{1}{\delta^{\lambda}(x,y)}\, d\mu(y) \leq c\, r^{-\lambda+1}\,, \qquad \textit{for } \lambda > 1\,,$$

where c is a constant independent of x.

Note that this lemma is valid for any quasidistance equivalent to δ. This lemma is well known. See for instance [GV].

3. Proofs of Theorems 1.1 through 1.5.

In the next proofs we will use without notice Lemma 2.2, Lemma 2.5 and normality.

PROOF OF THEOREM 1.1. To prove part a) observe that, since $f \in \mathrm{Lip}\,(\beta) \cap L^1$, the integral

$$I_\alpha f(x) = \int \frac{f(y)}{\delta_\alpha^{1-\alpha}(x,y)}\, d\mu(y)$$

converges absolutely for any x.

Now consider $x_1 \neq x_2$ and let $r = \delta_\alpha(x_1,x_2)$, $B = B_{2\kappa_\alpha r}(x_2)$ and B^c the complement of B. Since δ_α has the cancellation property stated in Lemma 2.4, we have

$$\begin{aligned} &I_\alpha f(x_2) - I_\alpha f(x_1) \\ &\qquad = \int_X (f(y) - f(x_2)) \Big(\frac{1}{\delta_\alpha^{1-\alpha}(x_2,y)} - \frac{1}{\delta_\alpha^{1-\alpha}(x_1,y)} \Big)\, d\mu(y)\,. \end{aligned}$$

Then

$$\begin{aligned} &\big| I_\alpha f(x_2) - I_\alpha f(x_1) \big| \\ &\qquad \leq \int_B \frac{|f(y) - f(x_2)|}{\delta_\alpha^{1-\alpha}(x_2,y)}\, d\mu(y) + \int_B \frac{|f(y) - f(x_2)|}{\delta_\alpha^{1-\alpha}(x_1,y)}\, d\mu(y) \\ &\qquad\quad + \int_{B^c} |f(y) - f(x_2)| \left| \frac{1}{\delta_\alpha^{1-\alpha}(x_2,y)} - \frac{1}{\delta_\alpha^{1-\alpha}(x_1,y)} \right| d\mu(y)\,. \end{aligned}$$

Since $|f(y) - f(x_2)| \leq c\, \|f\|_{\mathrm{Lip}(\beta)}\, r^\beta$ for $y \in B$, the first integral is less than or equal to $c\, \|f\|_{\mathrm{Lip}(\beta)}\, r^{\beta+\alpha}$.

To estimate the second integral observe that $B \subset B_{\kappa_\alpha(2\kappa_\alpha+1)r}(x_1)$ then using the previous argument and integrating over this ball we obtain that this integral is less than or equal to $c\,\|f\|_{\mathrm{Lip}(\beta)}\, r^{\beta+\alpha}$.

To estimate the third integral observe that for $y \in B^c$ we can apply Lemma 2.3, and using that $f \in \mathrm{Lip}\,(\beta)$ this integral is majorized by

$$c\,\|f\|_{\mathrm{Lip}(\beta)} r^\gamma \int_{B^c} \delta_\alpha^{\beta+\alpha-\gamma-1}(x_2, y)\, d\mu(y) \leq c\,\|f\|_{\mathrm{Lip}(\beta)}\, r^{\beta+\alpha}.$$

Since $r = \delta_\alpha(x_1, x_2)$ the proof of part a) is complete.

To prove part b) we show first that $\widetilde{I}_\alpha f(x)$ converges absolutely for every x. Since δ_α has the cancellation property stated in Lemma 2.4 we can write

$$\widetilde{I}_\alpha f(x) = \int (f(y) - f(x)) \left(\frac{1}{\delta_\alpha^{1-\alpha}(x, y)} - \frac{1}{\delta_\alpha^{1-\alpha}(x_0, y)} \right) d\mu(y).$$

Now it is clear that the function inside this integral is integrable over the ball $B_{2\kappa_\alpha \delta_\alpha(x,x_o)}(x)$. To see that it is also integrable in the complement of this ball we apply Lemma 2.3 and use the fact that $f \in \mathrm{Lip}\,(\beta)$. The proof that $\|\widetilde{I}_\alpha f\|_{\mathrm{Lip}(\alpha+\beta)} \leq c\,\|f\|_{\mathrm{Lip}(\beta)}$ proceeds exactly as in part a).

Finally the fact that for $f \in \mathrm{Lip}\,(\beta) \cap L^1$, $\widetilde{I}_\alpha f$ coincides with $I_\alpha f$ as an element of $\mathrm{Lip}\,(\alpha+\beta)$, follows from the fact that for such a function

$$\widetilde{I}_\alpha f(x) = I_\alpha f(x) - I_\alpha f(x_0).$$

PROOF OF THEOREM 1.2. Since $f \in \mathrm{Lip}\,(\beta) \cap L^\infty$ with $\alpha < \beta \leq \gamma$, the integral

$$D_\alpha f(x) = \int_X \frac{f(y) - f(x)}{\delta_{-\alpha}^{1+\alpha}(x, y)}\, d\mu(y)$$

converges absolutely for every x.

Now consider $x_1 \neq x_2$ and let $r = \delta_{-\alpha}(x_1, x_2), B = B_{2\kappa_{-\alpha} r}(x_2)$ and B^c the complement of B. We have

$$|D_\alpha f(x_2) - D_\alpha f(x_1)| \leq \int_B \frac{|f(y) - f(x_2)|}{\delta_{-\alpha}^{1+\alpha}(x_2, y)}\, d\mu(y) + \int_B \frac{|f(y) - f(x_1)|}{\delta_{-\alpha}^{1+\alpha}(x_1, y)}\, d\mu(y)$$

$$+\int_{B^c}\Big|\frac{f(y)-f(x_2)}{\delta_{-\alpha}^{1+\alpha}(x_2,y)}-\frac{f(y)-f(x_1)}{\delta_{-\alpha}^{1+\alpha}(x_1,y)}\Big|\,d\mu(y)\,.$$

Since $f\in \mathrm{Lip}\,(\beta)$ the first integral is majorized by $c\,\|f\|_{\mathrm{Lip}(\beta)}\,r^{\beta-\alpha}$. To estimate the second integral observe that $B\subset B_{\kappa_{-\alpha}(2\kappa_{-\alpha}+1)r}(x_1)$ then integrating over this ball and arguing as before this integral is majorized by $c\,\|f\|_{\mathrm{Lip}(\beta)}\,r^{\beta-\alpha}$.

To estimate the last integral we first rewrite the integrand as follows

$$\Big|\frac{f(x_1)-f(x_2)}{\delta_{-\alpha}^{1+\alpha}(x_2,y)}+(f(y)-f(x_1))\Big(\frac{1}{\delta_{-\alpha}^{1+\alpha}(x_2,y)}-\frac{1}{\delta_{-\alpha}^{1+\alpha}(x_1,y)}\Big)\Big|\,;$$

then this integral is less than or equal to

$$\int_{B^c}\frac{|f(x_1)-f(x_2)|}{\delta_{-\alpha}^{1+\alpha}(x_2,y)}\,d\mu(y) + \int_{B^c}|f(y)-f(x_1)|\Big|\frac{1}{\delta_{-\alpha}^{1+\alpha}(x_2,y)}-\frac{1}{\delta_{-\alpha}^{1+\alpha}(x_1,y)}\Big|\,d\mu(y)\,.$$

The first term is majorized by $c\,\|f\|_{\mathrm{Lip}(\beta)}\,r^{\beta-\alpha}$. Using Lemma 2.3 we can majorize the second term by

$$\|f\|_{\mathrm{Lip}(\beta)}\,\delta_{-\alpha}^{\gamma}(x_1,x_2)\int_{B^c}\delta_{-\alpha}^{\beta-\alpha-1-\gamma}(x_2,y)\,d\mu(y)\le c\,\|f\|_{\mathrm{Lip}(\beta)}\,r^{\beta-\alpha}\,.$$

Since $r=\delta_{-\alpha}(x_1,x_2)$ the proof of part a) is complete.

To prove part b), we show first that $\widetilde{D}_\alpha f(x)$ converges absolutely for every x. For x fixed, since $f\in\mathrm{Lip}(\beta)$, the integral converges absolutely over the ball $B_{2\kappa_{-\alpha}\delta_{-\alpha}(x,x_0)}(x)$. To prove that it also converges absolutely in the complement of this ball rewrite the integrand as follows

$$\frac{f(x_0)-f(x)}{\delta_{-\alpha}^{1+\alpha}(x,y)}+(f(y)-f(x_0))\Big(\frac{1}{\delta_{-\alpha}^{1+\alpha}(x,y)}-\frac{1}{\delta_{-\alpha}^{1+\alpha}(x_0,y)}\Big)\,.$$

The first term is clearly integrable. The fact that the second term is integrable is a consequence of Lemma 2.3. The proof that $\|\widetilde{D}_\alpha f\|_{\mathrm{Lip}(\beta-\alpha)}\le c\,\|f\|_{\mathrm{Lip}(\beta)}$ proceeds exactly as in part a).

Finally, part c) follows from the fact that for $f\in\mathrm{Lip}(\beta)\cap L^\infty$,

$$\widetilde{D}_\alpha f(x)=D_\alpha f(x)-D_\alpha f(x_0)\,.$$

PROOF OF THEOREM 1.3. Let

$$\Psi(x,y) = \int_X \frac{1}{\delta_{-\alpha}^{1+\alpha}(x,t)} \left| \frac{1}{\delta_\alpha^{1-\alpha}(y,t)} - \frac{1}{\delta_\alpha^{1-\alpha}(x,y)} \right| d\mu(t)$$

for $x \neq y$. Observe that $|K(x,y)| \leq \Psi(x,y)$. We will show now that

$$\Psi(x,y) \leq \frac{c}{\delta(x,y)} . \tag{3.1}$$

For fixed $x \neq y$ we break up X into three regions:

$$D_1 = \{t : 2\,\kappa_\alpha\, \delta_\alpha(x,y) \leq \delta_\alpha(x,t)\} ,$$

$$D_2 = \left\{t : \frac{1}{2\kappa_\alpha}\, \delta_\alpha(x,y) \leq \delta_\alpha(x,t) < 2\,\kappa_\alpha\, \delta_\alpha(x,y)\right\} ,$$

and

$$D_3 = \left\{t : \delta_\alpha(x,t) \leq \frac{1}{2\kappa_\alpha}\, \delta_\alpha(x,y)\right\} .$$

In D_1 we have that $\delta_\alpha(y,t) \geq \kappa_\alpha^{-1}\, \delta_\alpha(x,t) - \delta_\alpha(x,y) \geq \delta_\alpha(x,y)$, and therefore the integral over D_1 is less than or equal to

$$\frac{c}{\delta_{-\alpha}^{1-\alpha}(x,y)} \int_{D_1} \frac{1}{\delta_\alpha^{1+\alpha}(x,t)}\, d\mu(t) \leq c\, \frac{\delta_\alpha^{-\alpha}(x,y)}{\delta_\alpha^{1-\alpha}(x,y)} \leq \frac{c}{\delta(x,y)} .$$

The integral over D_2 is majorized by

$$\frac{c}{\delta_{-\alpha}^{1+\alpha}(x,y)} \left(\int_{D_1} \frac{d\mu(t)}{\delta_\alpha^{1-\alpha}(y,t)} + \frac{1}{\delta_\alpha^{1-\alpha}(x,y)} \int_{D_2} d\mu(t) \right) .$$

For t in D_2 we have that

$$\delta_\alpha(y,t) \leq \kappa_\alpha\, (\delta_\alpha(t,x) + \delta_\alpha(x,y) \leq \kappa_\alpha\, (2\,\kappa_\alpha + 1)\, \delta_\alpha(x,y) ,$$

Enlarging D_2 to the ball of center y and radius $\kappa_\alpha(2\,\kappa_\alpha + 1)\, \delta_\alpha(x,y)$ and integrating we have that the expression above is less than or equal to $c/\delta(x,y)$.

Since in D_3 , $2\,\kappa_\alpha\, \delta_\alpha(x,t) < \delta_\alpha(x,y)$, Lemma 2.3 can be applied, and the integral over D_3 is majorized by

$$c \int_{D_3} \frac{1}{\delta_\alpha^{1+\alpha}(x,t)}\, \delta_\alpha^{\gamma}(x,t)\, \delta_\alpha^{\alpha-1-\gamma}(x,y)\, d\mu(t)$$

$$\le \delta_\alpha^{\alpha-1-\gamma}(x,y)\int_{D_3}\delta_\alpha^{\gamma-\alpha-1}(x,t)\,d\mu(t)$$
$$\le c\,\delta_\alpha^{\alpha-1-\gamma}(x,y)\,\delta_\alpha^{\gamma-\alpha}(x,y)$$
$$\le \frac{c}{\delta(x,y)}\,.$$

We are now going to show that T_α is associated with the kernel K, as defined in Section 1. Let f and g be in C_0^β, $0<\alpha+\beta\le\gamma$, with disjoint supports.

$$T_\alpha f(x) = D_\alpha I_\alpha f(x)$$
$$= \int_X \frac{I_\alpha f(t) - I_\alpha f(x)}{\delta_{-\alpha}^{1+\alpha}(x,t)}\,d\mu(t)$$
$$= \int_X \frac{1}{\delta_{-\alpha}^{1+\alpha}(x,t)}\left(\int_X\left(\frac{1}{\delta_\alpha^{1-\alpha}(t,y)} - \frac{1}{\delta_\alpha^{1-\alpha}(x,y)}\right)f(y)\,d\mu(y)\right)d\mu(t)\,.$$

For $x\notin \operatorname{supp} f$ using the estimate obtained above for $\Psi(x,y)$ the last integral converges absolutely. Then changing the order of integration we have

$$T_\alpha f(x) = \int_X K(x,y)\,f(y)\,d\mu(y)\,,$$

where $K(x,y)$ is the kernel defined in (1.18). Furthermore for $x\in \operatorname{supp} g$, $\int_X |K(x,y)|\,|f(y)|\,d\mu(y)$ is bounded, and therefore

$$\langle T_\alpha f, g\rangle = \int_X T_\alpha f(x)\,g(x)\,d\mu(x) = \iint K(x,y)\,f(y)\,g(x)\,d\mu(x)\,d\mu(y)\,.$$

We will now prove condition (1.14). Note that $\delta_{-\alpha}\simeq\delta$ and therefore is suffices to prove (1.14) with $\delta_{-\alpha}$; *i.e.* we will show that there are positive constants ν, M, η, $1<\nu, 0<M$ and $0<\eta\le 1$ such that if

$$\nu\,\delta_{-\alpha}(x,y) < \delta_{-\alpha}(x,z) \tag{3.2}$$

then

$$|K(x,z)-K(y,z)| \le M\,\frac{\delta_{-\alpha}^{\eta}(x,y)}{\delta_{-\alpha}^{1+\eta}(x,z)}\,.$$

Let $c_\alpha\ge 1$ be a constant such that

$$c_\alpha^{-1}\,\delta_\alpha(x,y)\le\delta_{-\alpha}(x,y)\le c_\alpha\,\delta_\alpha(x,y)\,.$$

Denote by κ_α and $\kappa_{-\alpha}$ the constants in the triangle inequalities for δ_α and $\delta_{-\alpha}$, respectively. Now we choose constants ν and k so that

$$c_\alpha^2\,\kappa_\alpha\,\kappa_{-\alpha}^2 < k < \frac{\nu}{\kappa_{-\alpha}\,c_\alpha^2\,\kappa_\alpha}\,. \tag{3.3}$$

Let x, y, z be fixed points satisfying (3.2). To estimate $|K(x,z)-K(y,z)|$ observe first that

$$\begin{aligned} &|K(x,z)-K(y,z)| \\ &\leq \int_X \Big|\frac{1}{\delta_{-\alpha}^{1+\alpha}(x,t)}\Big(\frac{1}{\delta_\alpha^{1-\alpha}(t,z)}-\frac{1}{\delta_\alpha^{1-\alpha}(x,z)}\Big) \\ &\qquad - \frac{1}{\delta_{-\alpha}^{1+\alpha}(y,t)}\Big(\frac{1}{\delta_\alpha^{1-\alpha}(t,z)}-\frac{1}{\delta_\alpha^{1-\alpha}(y,z)}\Big)\Big|\,d\mu(t)\,. \end{aligned} \tag{3.4}$$

To estimate the integral in (3.4) we divide X into two regions:

$$A = \Big\{t:\ \frac{1}{k}\,\delta_{-\alpha}(x,z) < \delta_{-\alpha}(x,t)\Big\}$$

and its complement A^c.

To estimate the integral on A we rewrite the integrand as follows:

$$\begin{aligned} &\Big|\Big(\frac{1}{\delta_{-\alpha}^{1+\alpha}(x,t)}-\frac{1}{\delta_{-\alpha}^{1+\alpha}(y,t)}\Big)\frac{1}{\delta_\alpha^{1-\alpha}(t,z)} \\ &+\frac{1}{\delta_{-\alpha}^{1+\alpha}(y,t)}\Big(\frac{1}{\delta_\alpha^{1-\alpha}(y,z)}-\frac{1}{\delta_\alpha^{1-\alpha}(x,z)}\Big) \\ &+\frac{1}{\delta_\alpha^{1-\alpha}(y,z)}\Big(\frac{1}{\delta_{-\alpha}^{1+\alpha}(y,t)}-\frac{1}{\delta_{-\alpha}^{1+\alpha}(x,t)}\Big)\Big| \leq |\mathrm{I}_1|+|\mathrm{I}_2|+|\mathrm{I}_3|\,. \end{aligned}$$

We first estimate $\int_A |\mathrm{I}_3|\,d\mu$. Observe that for t in A

$$\frac{\nu}{k}\,\delta_{-\alpha}(x,y) < \delta_{-\alpha}(x,t)\,.$$

On the other hand, by (3.3), $\nu/k > \kappa_{-\alpha}$ and therefore we can apply Lemma 2.3 to obtain

$$\begin{aligned} \int_A |\mathrm{I}_3|\,d\mu(t) &\leq c\,\frac{\delta_{-\alpha}^{\gamma}(x,y)}{\delta_\alpha^{1-\alpha}(y,z)}\int_A \delta_{-\alpha}^{-\alpha-1-\gamma}(x,t)\,d\mu(t) \\ &\leq c\,\frac{\delta_{-\alpha}^{\gamma}(x,y)}{\delta_{-\alpha}^{1-\alpha}(y,z)}\,\delta_{-\alpha}^{-\alpha-\gamma}(x,z)\,. \end{aligned}$$

Note that

$$\delta_{-\alpha}(x,z)\left(1-\frac{\kappa_{-\alpha}}{\nu}\right) \le \kappa_{-\alpha}\,\delta_{-\alpha}(y,z)$$

and hence

$$\int_A |\mathrm{I}_3|\,d\mu(t) \le c\,\frac{\delta^{\gamma}_{-\alpha}(x,y)}{\delta^{1+\gamma}_{-\alpha}(x,z)}\,.$$

Since

$$\frac{\nu}{c_\alpha^2}\,\delta_\alpha(x,y) \le \delta_\alpha(x,z)$$

and since, by (3.3), $\nu/c_\alpha^2 > \kappa_\alpha$ we can apply Lemma 2.3 to I_2 to obtain

$$\int_A |\mathrm{I}_2|\,d\mu(t) \le c\,\delta^{\gamma}_{\alpha}(x,y)\,\delta_\alpha^{\alpha-1-\gamma}(x,z)\int_A \frac{1}{\delta^{1+\alpha}_{-\alpha}(x,t)}\,d\mu(t)\,.$$

The last expression is majorized by

$$c\,\frac{\delta^{\gamma}_{-\alpha}(x,y)}{\delta^{1+\gamma}_{-\alpha}(x,z)}\,.$$

To estimate $\int_A |\mathrm{I}_1|\,d\mu$ we will further subdivide A into

$$D_1 = \{t:\ \delta_{-\alpha}(x,t) > k\,c_\alpha\,\delta_{-\alpha}(x,z)\}$$

and

$$D_2 = \left\{t:\ \frac{1}{k}\,\delta_{-\alpha}(x,z) < \delta_{-\alpha}(x,t) \le k\,c_\alpha\,\delta_{-\alpha}(x,z)\right\}.$$

For t in D_1 we have

$$\delta_{-\alpha}(x,t) > c_\alpha\,k\,\delta_{-\alpha}(x,z) > c_\alpha\,k\,\nu\,\delta_{-\alpha}(x,y)$$

and, by (3.3), $c_\alpha k\,\nu > \kappa_{-\alpha}$ and therefore we can apply Lemma 2.3 to obtain

$$\int_{D_1} |\mathrm{I}_1|\,d\mu(t) \le c\,\delta^{\gamma}_{-\alpha}(x,y)\int_{D_1} \delta_{-\alpha}^{-\alpha-\gamma-1}(x,t)\,\delta_\alpha^{\alpha-1}(t,z)\,d\mu(t)\,.$$

Now note that for t in D_1

$$\left(1-\frac{\kappa_{-\alpha}}{c_\alpha\,k}\right)\delta_{-\alpha}(x,t) \le c_\alpha\,\kappa_{-\alpha}\,\delta_\alpha(t,z)\,. \tag{3.5}$$

By (3.3), $1-\nu_{-\alpha}/c_\alpha k>0$, and hence by (3.5) we obtain

$$\int_{D_1}|\mathrm{I}_1|\,d\mu(t)\le c\,\frac{\delta^{\gamma}_{-\alpha}(x,y)}{\delta^{1+\gamma}_{-\alpha}(x,z)}\,.$$

To estimate $\int_{D_2}|\mathrm{I}_1|\,d\mu(t)$ observe that for t in D_2, by (3.2),

$$\nu\,\delta_{-\alpha}(x,y)<\delta_{-\alpha}(x,z)<k\,\delta_{-\alpha}(x,t)\,.$$

By (3.3) $\kappa_{-\alpha}<\nu/k$, and therefore we can apply Lemma 2.3 to the integral to get

$$|\mathrm{I}_1|\le c\,\frac{\delta^{\gamma}_{-\alpha}(x,y)}{\delta^{1+\alpha+\gamma}_{-\alpha}(x,t)}\,\frac{1}{\delta^{1-\alpha}_{-\alpha}(t,z)}\le c\,\frac{\delta^{\gamma}_{-\alpha}(x,y)}{\delta^{1+\alpha+\gamma}_{-\alpha}(x,z)}\,\frac{1}{\delta^{1-\alpha}_{-\alpha}(t,z)}\,.$$

Since D_2 is contained in the ball

$$B=\{t:\ \delta_{-\alpha}(t,z)\le(\kappa_{-\alpha}\,c_\alpha\,k+\kappa_{-\alpha})\,\delta_{-\alpha}(x,z)\}$$

we get

$$\int_{D_2}|\mathrm{I}_1|\,d\mu(t)\le\int_B|\mathrm{I}_1|\,d\mu(t)\le c\,\frac{\delta^{\gamma}_{-\alpha}(x,y)}{\delta^{1+\gamma}_{-\alpha}(x,z)}\,.$$

Now we estimate integral in (3.4) on $A^c=\{t\colon k^{-1}\,\delta_{-\alpha}(x,z)\ge\delta_{-\alpha}(x,t)\}$. We divide this region into two subregions:

$$B_1=\{t:\ \delta_{-\alpha}(x,t)<k\,\delta_{-\alpha}(x,y)\}\,,$$
$$B_2=A^c\backslash B_1=\Big\{t:\ k\,\delta_{-\alpha}(x,y)\le\delta_{-\alpha}(x,t)\le\frac{1}{k}\,\delta_{-\alpha}(x,z)\Big\}\,.$$

We estimate first on B_1 :

$$\begin{aligned}\int_{B_1}\Big|&\frac{1}{\delta^{1+\alpha}_{-\alpha}(x,t)}\Big(\frac{1}{\delta^{1-\alpha}_{\alpha}(t,z)}-\frac{1}{\delta^{1-\alpha}_{\alpha}(x,z)}\Big)\\&-\frac{1}{\delta^{1+\alpha}_{-\alpha}(y,t)}\Big(\frac{1}{\delta^{1-\alpha}_{\alpha}(t,z)}-\frac{1}{\delta^{1-\alpha}_{\alpha}(y,z)}\Big)\Big|\,d\mu(t)\\&\le\int_{B_1}\frac{1}{\delta^{1+\alpha}_{-\alpha}(x,t)}\Big|\frac{1}{\delta^{1-\alpha}_{\alpha}(t,z)}-\frac{1}{\delta^{1-\alpha}_{\alpha}(x,z)}\Big|\,d\mu(t)\\&\quad+\int_{B_1}\frac{1}{\delta^{1+\alpha}_{-\alpha}(y,t)}\Big|\frac{1}{\delta^{1-\alpha}_{\alpha}(t,z)}-\frac{1}{\delta^{1-\alpha}_{\alpha}(y,z)}\Big|\,d\mu(t)\,.\end{aligned}$$

Now observe that for t in A^c by (3.3) there are constants $c, c' > \kappa_\alpha$,

$$c = \frac{k}{c_\alpha^2}\,, \qquad c' = c_\alpha^2\,\kappa_{-\alpha}^2\,\frac{\dfrac{1}{k}+\dfrac{1}{\nu}}{1-\dfrac{\kappa_{-\alpha}}{\nu}}\,,$$

such that $c\,\delta_\alpha(x,t) \le \delta_\alpha(x,z)$ and $c'\,\delta_\alpha(y,t) \le \delta_\alpha(y,z)$; therefore we can apply Lemma 2.3 to the integrands of both terms to obtain, for the first term:

$$\begin{aligned}\int_{B_1} \frac{1}{\delta_{-\alpha}^{1+\alpha}(x,t)}\,\frac{c\,\delta_\alpha^\gamma(t,x)}{\delta_\alpha^{1-\alpha+\gamma}(x,z)}\,d\mu(t)&\\ &\le \frac{c}{\delta_\alpha^{1-\alpha+\gamma}(x,z)}\int_{B_1}\frac{1}{\delta_{-\alpha}^{1-\gamma+\alpha}(x,t)}\,d\mu(t)\\ &\le c\,\frac{\delta_{-\alpha}^{\gamma-\alpha}(x,y)}{\delta_{-\alpha}^{1+\gamma-\alpha}(x,z)}\,,\end{aligned}$$

and for the second term:

$$\begin{aligned}\int_{B_1} \frac{1}{\delta_{-\alpha}^{1+\alpha}(y,t)}\,\frac{c\,\delta_\alpha^\gamma(t,y)}{\delta_\alpha^{1-\alpha+\gamma}(y,z)}\,d\mu(t)&\\ &\le \frac{1}{\delta_{-\alpha}^{1+\gamma-\alpha}(y,z)}\int_{B_1^*}\frac{1}{\delta_{-\alpha}^{1-\gamma+\alpha}(t,y)}\,d\mu(t)\,,\end{aligned}$$

where $B_1^* = \{t : \delta_{-\alpha}(y,t) \le \kappa_{-\alpha}\,(k+1)\,\delta_{-\alpha}(x,y)\}$.

Integrating and using the fact that

$$\frac{1}{\kappa_{-\alpha}}\left(1-\frac{\kappa_{-\alpha}}{\nu}\right)\delta_{-\alpha}(x,z) \le \delta_{-\alpha}(y,z)\,,$$

the last integral is majorized by

$$c\,\frac{\delta_{-\alpha}^{\gamma-\alpha}(x,y)}{\delta_{-\alpha}^{1+\gamma-\alpha}(x,z)}\,.$$

Now we estimate

$$\int_{B_2}\left|\frac{1}{\delta_{-\alpha}^{1+\alpha}(x,t)}\left(\frac{1}{\delta_\alpha^{1-\alpha}(t,z)}-\frac{1}{\delta_\alpha^{1-\alpha}(x,z)}\right)\right.$$

$$
\begin{aligned}
&- \frac{1}{\delta_{-\alpha}^{1+\alpha}(y,t)}\Big(\frac{1}{\delta_{\alpha}^{1-\alpha}(t,z)} - \frac{1}{\delta_{\alpha}^{1-\alpha}(y,z)}\Big)\Big| \, d\mu(t) \\
&\le \int_{B_2} \Big|\frac{1}{\delta_{-\alpha}^{1+\alpha}(x,t)} - \frac{1}{\delta_{-\alpha}^{1+\alpha}(y,t)}\Big| \\
&\qquad \cdot \Big|\frac{1}{\delta_{\alpha}^{1-\alpha}(t,z)} - \frac{1}{\delta_{\alpha}^{1-\alpha}(x,z)}\Big| \, d\mu(t) \\
&\quad + \int_{B_2} \frac{1}{\delta_{-\alpha}^{1+\alpha}(y,t)} \Big|\frac{1}{\delta_{\alpha}^{1-\alpha}(y,z)} - \frac{1}{\delta_{\alpha}^{1-\alpha}(x,z)}\Big| \, d\mu(t) \\
&= \mathrm{J}_1 + \mathrm{J}_2 \, .
\end{aligned}
$$

To estimate J_1 observe that $k\,\delta_{-\alpha}(x,y) \le \delta_{-\alpha}(x,t)$ and that

$$
\frac{k}{c_\alpha^2}\,\delta_\alpha(x,t) \le \delta_\alpha(x,z)\,,
$$

therefore we apply Lemma 2.3 to both brackets in absolute value to obtain

$$
\begin{aligned}
\mathrm{J}_1 &\le \int_{B_2} \frac{c\,\delta_{-\alpha}^{\gamma}(x,y)}{\delta_{-\alpha}^{1+\alpha+\gamma}(x,t)} \, \frac{\delta_\alpha^{\gamma}(x,t)}{\delta_\alpha^{1-\alpha+\gamma}(x,z)} \, d\mu(t) \\
&\le \frac{c\,\delta_{-\alpha}^{\gamma}(x,y)}{\delta_\alpha^{1+\gamma-\alpha}(x,z)} \int_{B_2} \frac{1}{\delta_{-\alpha}^{1+\alpha}(x,t)} \, d\mu(t)\,.
\end{aligned}
$$

Note that $B_2 \subset \{t:\ k\,\delta_{-\alpha}(x,y) \le \delta_{-\alpha}(x,t)\}$ then integrating over this set we have that the last expression is majorized by

$$
c\,\frac{\delta_{-\alpha}^{\gamma-\alpha}(x,y)}{\delta_{-\alpha}^{1+\gamma-\alpha}(x,z)}\,.
$$

This concludes the proof of (1.14) with exponent $\eta = \gamma - \alpha$.

We will now prove the condition (1.15). Since $\delta \approx \delta_\alpha$ it suffices to prove condition (1.15) for δ_α, we will show that there are positive constants ν', M', η', $1 < \nu'$, $0 < M'$ and $0 < \eta' \le 1$, such that if

$$
\nu'\,\delta_\alpha(x,y) < \delta_\alpha(x,z) \tag{3.6}
$$

then

$$
|K(z,x) - K(z,y)| \le M'\,\frac{\delta_\alpha^{\eta'}(x,y)}{\delta_\alpha^{1+\eta'}(x,z)}\,.
$$

We choose $\kappa_\alpha < \nu'$ and $\kappa_\alpha < k'$ such that

$$\frac{\kappa_\alpha^2}{\nu'} < 1 - \frac{\kappa_\alpha}{k'} . \tag{3.7}$$

Let x, y, z be fixed points satisfying (3.6). To estimate $|K(z,x) - K(z,y)|$ observe first that

$$\begin{aligned} |K(z,x) - K(z,y)| \le \int_X \frac{1}{\delta_{-\alpha}^{1+\alpha}(z,t)} \Big| &\frac{1}{\delta_\alpha^{1-\alpha}(t,x)} - \frac{1}{\delta_\alpha^{1-\alpha}(z,x)} \\ &- \frac{1}{\delta_\alpha^{1-\alpha}(t,y)} + \frac{1}{\delta_\alpha^{1-\alpha}(z,y)} \Big| \, d\mu(t) . \end{aligned} \tag{3.8}$$

To estimate we divide X into three regions:

$$A = \Big\{ t : \ \delta_\alpha(z,t) < \frac{1}{k'} \min\{\delta_\alpha(y,z), \delta_\alpha(x,z)\} \Big\} ,$$

$$B = \Big\{ t : \ \frac{1}{k'} \min\{\delta_\alpha(y,z), \delta_\alpha(x,z)\} \le \delta_\alpha(z,t) < k' \, \delta_\alpha(x,z) \Big\} ,$$

and

$$C = \{ t : \ k' \, \delta_\alpha(x,z) \le \delta_\alpha(z,t) \} .$$

To estimate the integral on A we further subdivide A into two subregions:

$$A_1 = \{ t : \delta_\alpha(z,t) \le k' \, \delta_\alpha(x,y) \} , \qquad \text{and} \qquad A_2 = A \backslash A_1 .$$

The integral over A_1 is less than or equal to

$$\begin{aligned} &\int_{A_1} \frac{1}{\delta_{-\alpha}^{1+\alpha}(z,t)} \Big| \frac{1}{\delta_\alpha^{1-\alpha}(t,x)} - \frac{1}{\delta_\alpha^{1-\alpha}(z,x)} \Big| \, d\mu(t) \\ &\qquad + \int_{A_1} \frac{1}{\delta_{-\alpha}^{1+\alpha}(z,t)} \Big| \frac{1}{\delta_\alpha^{1-\alpha}(z,y)} - \frac{1}{\delta_\alpha^{1-\alpha}(t,y)} \Big| \, d\mu(t) . \end{aligned}$$

Note that for t in A, $k' \, \delta_\alpha(z,t) < \delta_\alpha(x,z)$ and $k' \, \delta_\alpha(z,t) < \delta_\alpha(y,z)$ by (3.7), $\kappa_\alpha < k'$. Therefore we can apply Lemma 2.3 in both integrands. The first term is majorized by

$$\frac{c}{\delta_\alpha^{1-\alpha+\gamma}(z,x)} \int_{A_1} \frac{1}{\delta_{-\alpha}^{1+\alpha}(z,t)} \, \delta_\alpha^\gamma(t,z) \, d\mu(t) \le c \, \frac{\delta_\alpha^{\gamma-\alpha}(x,y)}{\delta_\alpha^{1+\gamma-\alpha}(z,x)} .$$

The second term is majorized by

$$\frac{c}{\delta_\alpha^{1-\alpha+\gamma}(z,y)}\int_{A_1}\frac{1}{\delta_{-\alpha}^{1+\alpha}(z,t)}\,\delta_\alpha^\gamma(z,t)\,d\mu(t)\le c\,\frac{\delta_\alpha^{\gamma-\alpha}(x,y)}{\delta_\alpha^{1+\gamma-\alpha}(z,y)}\,,$$

since by (3.6)

$$\frac{1}{\kappa_\alpha}\left(1-\frac{\kappa_\alpha}{\nu'}\right)\delta_\alpha(x,z)\le\delta_\alpha(y,z)\,,$$

the last expression is less than or equal to

$$c\,\frac{\delta_\alpha^{\gamma-\alpha}(x,y)}{\delta_\alpha^{1+\gamma-\alpha}(z,x)}\,.$$

The integral over A_2 is less than or equal to

$$\int_{A_2}\frac{1}{\delta_{-\alpha}^{1+\alpha}(z,t)}\left|\frac{1}{\delta_\alpha^{1-\alpha}(t,x)}-\frac{1}{\delta_\alpha^{1-\alpha}(t,y)}\right|d\mu(t)$$
$$+\int_{A_2}\frac{1}{\delta_{-\alpha}^{1+\alpha}(z,t)}\left|\frac{1}{\delta_\alpha^{1-\alpha}(z,y)}-\frac{1}{\delta_\alpha^{1-\alpha}(z,x)}\right|d\mu(t)$$

and for t in A_2, by (3.7) there is $a>\kappa_\alpha$,

$$\nu'\left(1-\frac{\kappa_\alpha}{k'}\right)\frac{1}{\kappa_\alpha}=c\,,$$

such that $a\,\delta_\alpha(x,y)<\delta_\alpha(x,t)$ also by (3.6) $\nu'\,\delta_\alpha(x,y)<\delta_\alpha(x,z)$ and by (3.7), $\kappa_\alpha<\nu'$ therefore we can apply Lemma 2.3 in both terms.

The first term is less than or equal to

$$\delta_\alpha^\gamma(x,y)\int_{A_2}\frac{1}{\delta_{-\alpha}^{1+\alpha}(z,t)}\,\frac{1}{\delta_\alpha^{1-\alpha+\gamma}(x,t)}\,d\mu(t)$$
$$\le c\,\frac{\delta_\alpha^\gamma(x,y)}{\delta_\alpha^{1+\gamma-\alpha}(x,z)}\int_{A_2}\frac{1}{\delta_{-\alpha}^{1+\alpha}(z,t)}\,d\mu(t)$$

because for t in A_2,

$$\delta_\alpha(x,z)\left(1-\frac{\kappa_\alpha}{k'}\right)\le\kappa_\alpha\,\delta_\alpha(t,x)\,.$$

Now integrating over $\{t:\delta_\alpha(z,t)>k'\,\delta_\alpha(x,y)\}$ we get that the last expression is majorized by

$$c\,\frac{\delta_\alpha^{\gamma-\alpha}(x,y)}{\delta_\alpha^{1+\gamma-\alpha}(x,z)}\,.$$

The second term is less than or equal to

$$\frac{\delta_\alpha^\gamma(x,y)}{\delta_\alpha^{1+\gamma-\alpha}(x,z)} \int_{A_2} \frac{1}{\delta_\alpha^{1+\alpha}(z,t)}\, d\mu(t)\,.$$

Integrating as before over $\{t : \delta_\alpha(z,t) > k'\,\delta_\alpha(x,y)\}$ the last expression is majorized by

$$c\,\frac{\delta_\alpha^{\gamma-\alpha}(x,y)}{\delta_\alpha^{1+\gamma-\alpha}(x,z)}\;.$$

The integral over B is less than or equal to

$$\int_B \frac{1}{\delta_{-\alpha}^{1+\alpha}(z,t)} \left|\frac{1}{\delta_\alpha^{1-\alpha}(z,y)} - \frac{1}{\delta_\alpha^{1-\alpha}(z,x)}\right| d\mu(t) + \int_B \frac{1}{\delta_{-\alpha}^{1+\alpha}(z,t)} \left|\frac{1}{\delta_\alpha^{1-\alpha}(t,x)} - \frac{1}{\delta_\alpha^{1-\alpha}(t,y)}\right| d\mu(t)\,.$$

To estimate the first integral observe that by (3.6) and (3.7) we can apply Lemma 2.3 to the integrand, and majorize this integral by

$$\frac{\delta_\alpha^\gamma(x,y)}{\delta_\alpha^{1-\alpha+\gamma}(x,z)} \int_B \frac{d\mu(t)}{\delta_{-\alpha}^{1+\alpha}(z,t)}\;.$$

Now integrating over

$$\left\{t:\ \delta_\alpha(z,t) > \frac{1}{k'}\,\min\{\delta_\alpha(y,z),\delta_\alpha(x,z)\}\right\}$$

and using

$$\frac{1}{\kappa_\alpha}\left(1-\frac{\kappa_\alpha}{\nu'}\right)\delta_\alpha(x,z) \le \delta_\alpha(y,z)$$

we get that the last expression is majorized by

$$c\,\frac{\delta_\alpha^\gamma(x,y)}{\delta_\alpha^{1+\gamma}(x,z)}\;.$$

To estimate the second term we consider two regions

$$D_1 = B \cap \{t:\ \delta_\alpha(t,x) < k'\,\delta_\alpha(x,y)\}\;,$$

and

$$D_2 = B \cap \{t:\ \delta_\alpha(t,x) \ge k'\,\delta_\alpha(x,y)\}\;.$$

By (3.6) and (3.7) there are constants a_1 and a_2 such that

$$a_1\,\delta_\alpha(z,y) \le \delta_\alpha(z,x) \le a_2\,\delta_\alpha(z,y)\,,$$

therefore the integral over D_1 is less than or equal to

$$\frac{c}{\delta_\alpha^{1+\alpha}(z,x)} \int_{D_1} \frac{d\mu(t)}{\delta_\alpha^{1-\alpha}(t,x)} + \frac{c}{\delta_\alpha^{1+\alpha}(z,x)} \int_{D_1} \frac{d\mu(t)}{\delta_\alpha^{1-\alpha}(t,y)}\,.$$

Since $D_1 \subset \{t : \delta_\alpha(t,x) < k'\,\delta_\alpha(x,y)\}$ integrating over this ball we get that the first term is majorized by $c\,\delta_\alpha^\alpha(x,y)/\delta_\alpha^{1+\alpha}(x,z)$. On the other hand

$$D_1 \subset \{t :\ \delta_\alpha(t,y) \le \kappa_\alpha\,(k'+1)\,\delta_\alpha(x,y)\}\,.$$

Therefore integrating over this ball the second integral is majorized by $c\,\delta_\alpha^\alpha(x,y)/\delta_\alpha^{1+\alpha}(x,z)$. For t in D_2, $\delta_\alpha(t,x) \ge k'\,\delta_\alpha(x,y)$, and therefore we can apply Lemma 2.3 to the integrand, and majorize the integral by

$$c\,\delta_\alpha^\gamma(x,y) \int_{D_2} \frac{d\mu(t)}{\delta_\alpha^{1+\alpha}(z,t)\,\delta_\alpha(x,t)^{1+\gamma-\alpha}}\,.$$

On the other hand there is $a_3 > 0$ such that $a_3\,\delta_\alpha(x,z) \le \delta_\alpha(z,t)$, and therefore the last expression is less than or equal to

$$c\,\frac{\delta_\alpha^\gamma(x,y)}{\delta_\alpha^{1+\alpha}(z,x)} \int_{D_2} \frac{d\mu(t)}{\delta_\alpha^{1+\gamma-\alpha}(x,t)}\,.$$

Now integrating over $\{t : \delta_\alpha(t,x) \ge k'\,\delta(x,y)\}$ the last expression is majorized by

$$c\,\frac{\delta_\alpha^\alpha(x,y)}{\delta_\alpha^{1+\alpha}(x,z)}\,.$$

Finally, we will estimate the integral over C. This integral is less than or equal to

$$\int_C \frac{1}{\delta_{-\alpha}^{1+\alpha}(z,t)} \left| \frac{1}{\delta_\alpha^{1-\alpha}(t,x)} - \frac{1}{\delta_\alpha^{1-\alpha}(t,y)} \right| d\mu(t) + \int_C \frac{1}{\delta_{-\alpha}^{1+\alpha}(z,t)} \left| \frac{1}{\delta_\alpha^{1-\alpha}(z,y)} - \frac{1}{\delta_\alpha^{1-\alpha}(z,x)} \right| d\mu(t)\,.$$

For t in C by (3.6) and (3.7) there are constants a_4, a_5

$$a_4 = \frac{k'-\kappa}{\kappa}\,\nu'\,, \qquad a_5 = \frac{\kappa_\alpha}{k'\left(1-\frac{\kappa_\alpha}{k'}\right)}\,,$$

such that $a_4\,\delta_\alpha(x,y) \le \delta_\alpha(x,t)$ and $\delta_\alpha(x,z) \le a_5\,\delta_\alpha(x,t)$. Therefore we can apply Lemma 2.3 to the integrand of the first term, and majorize the integral by

$$c\,\delta_\alpha^\gamma(x,y)\int_C \frac{d\mu(t)}{\delta_\alpha^{1+\alpha}(t,z)\,\delta_\alpha^{1+\gamma-\alpha}(x,t)} \le c\,\frac{\delta_\alpha^\gamma(x,y)}{\delta_\alpha^{1+\gamma-\alpha}(x,z)}\int_C \frac{d\mu(t)}{\delta_\alpha^{1+\alpha}(t,z)} \le c\,\frac{\delta_\alpha^\gamma(x,y)}{\delta_\alpha^{1+\gamma}(x,z)}\,.$$

By (3.6), $\nu'\,\delta_\alpha(x,y) \le \delta_\alpha(x,z)$, and by (3.7), $\kappa_\alpha < \nu'$, therefore we can apply Lemma 2.3 to the integrand of the second term and majorize the integral by

$$c\,\frac{\delta_\alpha^\gamma(x,y)}{\delta_\alpha^{1+\gamma-\alpha}(x,z)}\int_C \frac{d\mu(t)}{\delta_\alpha^{1+\alpha}(t,z)} \le c\,\frac{\delta_\alpha^\gamma(x,y)}{\delta_\alpha^{1+\gamma}(x,z)}\,.$$

To conclude the proof choose η' to be α.

PROOF OF THEOREM 1.4. To prove Theorem 1.4 we will use the "$T1$ theorem" (see [Ch]), *i.e.* "A singular integral operator T is a Calderón-Zygmund operator if and only if

1) T is weakly bounded,

2) $T1 \in \mathrm{BMO}$,

3) ${}^tT1 \in \mathrm{BMO}$."

We recall that an operator $T : C_0^\eta \to (C_0^\eta)'$ is *weakly bounded* if there exists a constant c such that

$$|\langle Tf, g\rangle| \le c\,\mu(B)^{1+2\eta}\,\|f\|_\eta\,\|g\|_\eta \tag{3.9}$$

for every f, g in $C_0^\eta(B)$ and for every ball B.

We will show that

i) T_α is weakly bounded,

ii) $T_\alpha 1 = 0$,

iii) ${}^tT_\alpha 1 = 0$,

To prove i) we will show first the following estimate for $f \in C_0^\eta(B)$, $0 < \eta + \alpha < \gamma$,

$$\|T_\alpha f\|_\infty \leq c\,\mu(B)^\eta \|f\|_\eta \,. \tag{3.10}$$

Consider $f \in C_0^\eta(B)$, $B = B_r(x_0)$. Observe that

$$|I_\alpha f(x)| \leq \int_B \frac{|f(y)|}{\delta_\alpha^{1-\alpha}(x,y)}\, d\mu(y) \leq c\,\|f\|_\infty\,(\mu(B))^\alpha \,.$$

Now

$$\begin{aligned}|T_\alpha f(x)| &\leq \int \frac{|(I_\alpha f)(t) - (I_\alpha f)(x)|}{\delta_{-\alpha}^{1+\alpha}(x,t)}\, d\mu(t) \\ &\leq \int\limits_{\delta(x,t)<r} \frac{|I_\alpha f(t) - I_\alpha f(x)|}{\delta_{-\alpha}^{1+\alpha}(x,t)}\, d\mu(t) \\ &\quad + \int\limits_{\delta(x,t)\geq r} \frac{|I_\alpha f(t) - I_\alpha f(x)|}{\delta_{-\alpha}^{1+\alpha}(x,t)}\, d\mu(t) \,.\end{aligned}$$

To estimate the first integral we use the fact proved in Theorem 1.1 that $|I_\alpha f(t) - I_\alpha f(x)| \leq c\,\|f\|_\eta\, \delta_{-\alpha}^{\eta+\alpha}(t,x)$ then integrating this integral we see that it is less than or equal to $c\,\|f\|_\eta (\mu(B))^\eta$. For the second integral we use the estimate for $I_\alpha f$ obtained above and integrating we obtain that this integral is less than or equal $c\,\|f\|_\infty$.

Note that for $f \in C_0^\eta(B)$, $\|f\|_\infty \leq c\,\|f\|_\eta\, \mu(B)^\eta$; this concludes the proof of (3.10). Let f and g be in $C_0^\eta(B)$, then

$$\begin{aligned}|\langle T_\alpha f, g\rangle| &\leq \int_B |T_\alpha f(x)|\ |g(x)|\, d\mu(x) \\ &\leq \|T_\alpha f\|_\infty\ \|g\|_\infty\ \mu(B) \\ &\leq c\,\mu(B)^{1+2\eta}\ \|f\|_\eta\ \|g\|_\eta \,.\end{aligned}$$

The last inequality follows from (3.10) and the fact that $g \in C_0^\eta(B)$.

To prove ii) we observe that the extension of T_α to $L^\infty \cap \operatorname{Lip}(\eta)$ coincides with the operator $\widetilde{T}_\alpha = \widetilde{D}_\alpha \widetilde{I}_\alpha$. Since, by Lemma 2.4, $\widetilde{I}_\alpha 1 = 0$ we have $T_\alpha 1 = \widetilde{T}_\alpha 1 = 0$.

To prove iii) we use the following fact

$$^tT_\alpha = I_\alpha D_\alpha \,, \tag{3.11}$$

which will be proved in Theorem 1.5. Now, since $D_\alpha 1 = 0$ we have $^tT_\alpha 1 = 0$. This concludes the proof of the Theorem.

PROOF OF THEOREM 1.5. Let $S_\alpha = I_\alpha D_\alpha$ and consider f and g in C_o^β, $0 < \alpha + \beta \leq \gamma$. We want to show that

$$\langle T_\alpha f, g\rangle = \langle f, S_\alpha g\rangle \ . \tag{3.12}$$

We will show first that for $f \in L^\infty \cap \mathrm{Lip}\,(\eta)$, $\alpha < \eta \leq \gamma$ and $g \in C_0^\beta$,

$$\langle D_\alpha f, g\rangle = \langle f, D_\alpha g\rangle \ . \tag{3.13}$$

For every $f \in L^\infty \cap \mathrm{Lip}\,(\eta)$, note that

$$\int \frac{|f(t) - f(x)|}{\delta_{-\alpha}^{1+\alpha}(x,t)}\, d\mu(t)$$

is bounded as a function of x and therefore

$$\langle D_\alpha f, g\rangle = \iint \frac{f(t) - f(x)}{\delta_{-\alpha}^{1+\alpha}(x,t)}\, g(x)\, d\mu(t)\, d\mu(x)$$

because the double integral converges absolutely. Now rewrite this integral as follows

$$\iint \frac{f(t)\,g(x) - f(x)\,g(t)}{\delta_{-\alpha}^{1+\alpha}(x,t)}\, d\mu(t)\, d\mu(x) + \iint \frac{f(x)\,g(t) - f(x)\,g(x)}{\delta_{-\alpha}^{1+\alpha}(x,t)}\, d\mu(t)\, d\mu(x) .$$

The second integral converges absolutely since for $g \in C_0^\beta(B_r(x_0))$

$$\int \frac{|g(t) - g(x)|}{\delta_{-\alpha}^{1+\alpha}(x,t)}\, d\mu(t) \leq \frac{c}{1 + \delta_{-\alpha}^{1+\alpha}(x_0, x)} ,$$

and it is equal to $\langle f, D_\alpha g\rangle$. Since the second integral is absolutely convergent so is the first one, and it is equal to zero because its integrand, $h(x,t)$, satisfies $h(x,t) = -h(t,x)$.

Now consider f and g in C_o^β. It was shown before (see Theorem 1.1 and (3.10)) that $I_\alpha f \in L^\infty \cap \mathrm{Lip}\,(\eta)$, therefore

$$\langle D_\alpha I_\alpha f, g\rangle = \langle I_\alpha f, D_\alpha g\rangle = \iint D_\alpha g(x)\, \frac{f(t)}{\delta_\alpha^{1-\alpha}(x,t)}\, d\mu(t)\, d\mu(x) .$$

Since $I_\alpha|f| \in L^\infty$ and

$$|D_\alpha g(x)| \leq \frac{c}{1+\delta_{-\alpha}^{1+\alpha}(x,x_0)},$$

the double integral converges absolutely and by Fubini's theorem is equal to $\langle f, I_\alpha D_\alpha g\rangle$.

The fact that $S_\alpha = {}^tT_\alpha$ is a Calderón-Zygmund operator follows from the fact that T_α is a Calderón-Zygmund operator.

4. Lemmas needed for the proofs of Theorem 1.6 and Theorem 1.7

Lemma 4.1. *The kernel $q(x,y,t)$ defined in* (1.16) *has the following properties*

i) $q(x,y,t) = q(y,x,t)$ *for all x, y in X, and $t > 0$,*

ii) $q(x,y,t) = 0$ *if $\delta(x,y) > b_1 t$,*

iii) $|q(x,y,t)| \leq c_4/t$ *for all x,y in X and $t>0$,*

iv) $|q(x,y,t) - q(x',y,t)| \leq c_5 \dfrac{\delta^\gamma(x,x')}{t^{1+\gamma}}$ *for all x, x', y in X and $t > 0$,*

v) $\displaystyle\int q(x,y,t)\, d\mu(y) = 0$ *for all x in X and $t>0$.*

Lemma 4.1 is the continuous version of known results [N2], [DJS].

Lemma 4.2. *The integral operator Q_t introduced in* (1.17) *satisfy the following estimates*

$$\|Q_tQ_s\| \leq c \begin{cases} \left(\dfrac{s}{t}\right)^\gamma, & \text{if } 0 < s \leq t, \\ \left(\dfrac{t}{s}\right)^\gamma, & \text{if } 0 < t \leq s, \end{cases}$$

where the norm is the operator norm in L^2, and c is a constant independent of s and t.

From this result one obtains the next lemma.

Lemma 4.3. *For positive r, s, t define a function $h_t(s,r)$ as follows*

$$h_t(s,r) = \begin{cases} \min\left\{t^{\gamma}, \left(\frac{s}{r}\right)^{\gamma/2}\right\}, & \text{if } \frac{s}{r} \le 1 \text{ and } 0 < t \le 1, \\ \min\left\{t^{\gamma}, \left(\frac{s}{r}\right)^{-\gamma/2}\right\}, & \text{if } \frac{s}{r} > 1 \text{ and } 0 < t \le 1, \end{cases}$$

$$h_t(s,r) = \begin{cases} \min\left\{t^{-\gamma}, \left(\frac{s}{r}\right)^{\gamma/2}\right\}, & \text{if } \frac{s}{r} \le 1 \text{ and } t > 1, \\ \min\left\{t^{-\gamma}, \left(\frac{s}{r}\right)^{-\gamma/2}\right\}, & \text{if } \frac{s}{r} > 1 \text{ and } t > 1. \end{cases}$$

Then the operators Q_t introduced in (1.17) satisfy

$$\|Q_s Q_{ts} (Q_r Q_{tr})^*\| \le h_t^2(s,r)$$

and

$$\|(Q_s Q_{ts})^* Q_r Q_{tr}\| \le h_t^2(s,r).$$

Moreover, setting

$$c(t) = \sup_s \int_0^\infty h_t(s,r) \frac{dr}{r}$$

one has the estimates

$$c(t) \le c \begin{cases} t^{\gamma} + t^{\gamma} \log \frac{1}{t}, & \text{for } 0 < t \le 1, \\ t^{-\gamma} + t^{-\gamma} \log t, & \text{for } t > 1. \end{cases} \tag{4.1}$$

Lemmas 4.2 and 4.3 are continuous versions of known results, see *e.g.* [DJS] and [N2].

Lemma 4.4. *Let Q_t be the operators defined by* (1.17), *and let $f \in L^2(X)$, then for every set E of finite measure of the measure space $([0,\infty), dt/t)$ one has*

$$\left\| \int_E Q_s Q_{st} f \frac{ds}{s} \right\|_2 \le c(t) \|f\|_2,$$

where $c(t)$ is the quantity introduced in Lemma 4.3. *Furthermore*

$$W_t f = \int_0^\infty Q_s Q_{ts} f \frac{ds}{s}$$

exists in the weak L^2 sense and satisfies

$$\|W_t f\|_2 \leq c(t)\, \|f\|_2 \,. \tag{4.2}$$

This result follows from Lemma 4.3 and the continuous version of the Cotlar-Knapp-Stein Lemma, see [CV] and [F].

5. Proofs of Theorems 1.6 and 1.7.

PROOF OF THEOREM 1.6. To prove (1.19) observe that for $f \in \operatorname{Lip}(\beta) \cap L^1$ the integral (1.10) converges absolutely for every x, therefore using (1.8) we have

$$\alpha\, I_\alpha f(x) = \alpha \int_X \int_0^\infty t^{\alpha-1}\, s(x,y,t)\, dt\;\, f(y)\, d\mu(y)$$

and the double integral converges absolutely for every x. Then by changing the order of integration we obtain

$$\alpha\, I_\alpha f(x) = \alpha \int_0^\infty t^{\alpha-1}\, u(x,t)\, dt\,, \tag{5.1}$$

where

$$u(x,t) = \int_X s(x,y,t)\, f(y)\, d\mu(y)\,. \tag{5.2}$$

Since

$$\frac{\partial}{\partial t}\, s(x,y,t) = \frac{1}{t}$$

and $q(x,y,t)$ has the properties ii) and iii) of Lemma 4.1, we can differentiate with respect to t under the integral sign of (5.2) to get

$$\frac{\partial}{\partial t} u(x,t) = \frac{1}{t}\, v(x,t)\,, \tag{5.3}$$

where

$$v(x,t) = \int_X q(x,y,t)\, f(y)\, d\mu(y) = -Q_t f(x)\,. \tag{5.4}$$

Now integrating the integral in (5.1) by parts, using (5.3) and the fact that $f \in \mathrm{Lip}(\beta) \cap L^1$ we obtain

$$\begin{aligned}\alpha\, I_\alpha f(x) &= \lim_{\substack{a\to 0\\ b\to\infty}} \alpha \int_a^b t^{\alpha-1}\, u(x,t)\, dt \\ &= \lim_{\substack{a\to 0\\ b\to\infty}} (t^\alpha\, u(x,t)|_a^b) - \lim_{\substack{a\to 0\\ b\to\infty}} \int_a^b t^\alpha\, v(x,t)\, \frac{dt}{t} \\ &= \int_0^\infty t^\alpha\, Q_t(f)(x)\, \frac{dt}{t}\,.\end{aligned}$$

To prove (1.20) observe that for $f \in \mathrm{Lip}(\beta) \cap L^\infty$ the integral (1.9) converges absolutely for every x, therefore using (1.8) we have

$$-\alpha\, D_\alpha f(x) = -\alpha \int_X \int_0^\infty t^{-\alpha-1}\, s(x,y,t)\, dt\, (f(y) - f(x))\, d\mu(y)$$

and the double integral converges absolutely. Then by changing the order of integration we have

$$-\alpha\, D_\alpha f(x) = -\alpha \int_0^\infty t^{-\alpha-1}\, (u(x,t) - f(x))\, dt\,. \tag{5.5}$$

Now integrating the integral in (5.5) by parts, using (5.3), and the fact that $f \in \mathrm{Lip}(\beta) \cap L^\infty$, $\alpha < \beta \le \gamma$, we obtain

$$\begin{aligned}-\alpha\, D_\alpha f(x) &= \lim_{\substack{a\to 0\\ b\to\infty}} -\alpha \int_a^b t^{-\alpha-1}\, (u(x,t) - f(x))\, dt \\ &= \lim_{\substack{a\to 0\\ b\to\infty}} (t^{-\alpha}\, (u(x,t) - f(x))|_a^b) - \lim_{\substack{a\to 0\\ b\to\infty}} \int_a^b t^{-\alpha}\, v(x,t)\, \frac{dt}{t} \\ &= \int_0^\infty t^{-\alpha}\, Q_t(f)(x)\, \frac{dt}{t}\,.\end{aligned}$$

The fact that representation formulas hold in the weak sense, *i.e.* that

$$\lim_{\substack{a\to 0\\ b\to\infty}} \int_a^b t^\alpha\, \langle Q_t f, \varphi\rangle\, \frac{dt}{t} = \langle \alpha\, I_\alpha f, \varphi\rangle$$

and

$$\lim_{\substack{a\to 0\\ b\to\infty}} \int_a^b t^{-\alpha}\, \langle Q_t f, \varphi\rangle\, \frac{dt}{t} = -\langle \alpha\, D_\alpha f, \varphi\rangle$$

for all $\varphi \in C_0^\mu$, $0 < \mu < \gamma$, follows from observing that the double integrals

$$\int_0^\infty \int_X t^\alpha \, Q_t f(x) \, \varphi(x) \, d\mu(x) \, \frac{dt}{t}$$

and

$$\int_0^\infty \int_X t^{-\alpha} \, Q_t f(x) \, \varphi(x) \, d\mu(x) \, , \frac{dt}{t}$$

are absolutely convergent.

This concludes the proof of Theorem 1.6.

PROOF OF THEOREM 1.7. The proof of Theorem 1.7 is a continuous version of the method of Nahmod [N2].

We first show that for $0 \leq \alpha < \gamma$ the integral

$$\int_0^\infty \int_0^\infty s^{-\alpha} \, t^\alpha \, \langle Q_s Q_t f, g \rangle \, \frac{ds}{s} \, \frac{dt}{t} \tag{5.6}$$

converges absolutely for f and g in $L^2(X)$. We make the following change of variables in (5.6)

$$s = u \, , \qquad t = u \, v \, ,$$

and obtain

$$\int_0^\infty \int_0^\infty v^\alpha \, \langle Q_u Q_{uv} f, g \rangle \, \frac{du}{u} \, \frac{dv}{v} \, . \tag{5.7}$$

By Cotlar's lemma (Lemma 4.3) it can be seen that

$$\int_0^\infty |\langle Q_u Q_{uv} f, g \rangle| \, \frac{du}{u} \leq 4 \, c(v) \, \|f\|_2 \, \|g\|_2 \, ,$$

where $c(v)$ is the constant of Lemma 8. Using the estimates (4.1) for $c(v)$ one sees that (5.7), and therefore (5.6) are absolutely convergent.

We will show next that for $f \in C_0^\beta$, $g \in C_0^\mu$, where $0 < \alpha + \beta < \gamma$, and $0 < \mu < \gamma$, (5.6) is equal to $-\alpha^2 \, \langle T_\alpha f, g \rangle$ for $0 < \alpha < \gamma$, and equal to $\langle f, g \rangle$ for $\alpha = 0$.

In other words

$$f = \int_0^\infty \int_0^\infty Q_t Q_s f \, \frac{ds}{s} \, \frac{dt}{t} \, , \tag{5.8}$$

and

$$(5.9) \qquad -\alpha^2 T_\alpha f = \int_0^\infty \int_0^\infty s^{-\alpha} t^\alpha Q_s Q_t f \frac{ds}{s} \frac{dt}{t},$$

where the integrals are in the weak L^2 sense. The equality (5.8) is a well known formula of Coifman, see [C]. Let, then, $f \in C_0^\beta$, $0 < \alpha + \beta < \gamma$ and $g \in C_0^\mu$, $0 < \mu < \gamma$. Since (5.6) is absolutely convergent it can be written as an iterated integral

$$\begin{aligned}
\int_0^\infty \int_0^\infty & s^{-\alpha} t^\alpha \langle Q_s Q_t f, g\rangle \frac{ds}{s} \frac{dt}{t} \\
&= \int_0^\infty s^{-\alpha} \left(\int_0^\infty t^\alpha \langle Q_s Q_t f, g\rangle \frac{dt}{t}\right) \frac{ds}{s} \\
&= \int_0^\infty s^{-\alpha} \left(\int_0^\infty t^\alpha \langle Q_t f, Q_s g\rangle \frac{dt}{t}\right) \frac{ds}{s} \\
&= \int_0^\infty s^{-\alpha} \left\langle \int_0^\infty t^\alpha Q_t f \frac{dt}{t}, Q_s g\right\rangle \frac{ds}{s} \\
&= \int_0^\infty s^{-\alpha} \langle \alpha I_\alpha f, Q_s g\rangle \frac{ds}{s} \\
&= \alpha \int_0^\infty s^{-\alpha} \langle Q_s(I_\alpha f), g\rangle \frac{ds}{s} \\
&= \alpha \left\langle \int_0^\infty s^{-\alpha} Q_s(I_\alpha f) \frac{ds}{s}, g\right\rangle \\
&= -\alpha^2 \langle D_\alpha I_\alpha f, g\rangle \\
&= -\alpha^2 \langle T_\alpha f, g\rangle .
\end{aligned}$$

The above chain of equalities is easily justified by using the properties of the kernel $q(x, y, t)$ and Theorem 1.6. For $\alpha = 0$ the calculation is quite similar except for the fact that instead of Theorem 1.6 one uses the known identity

$$\int_0^\infty Q_t f \frac{dt}{t} = f .$$

Let now $0 < \alpha$, $f \in C_0^\beta$, $\alpha + \beta < \gamma$. Using that (5.6) and (5.7) are the same we can write

$$(I + \alpha^2 T_\alpha) f = \int_0^\infty (1 - v^\alpha) W_v f \frac{dv}{v},$$

where W_v is the operator defined in Lemma 4.4. Applying the estimate (4.2) we obtain

$$\|(I + \alpha^2 T_\alpha) f\|_2 \le \int_0^\infty |1 - v^\alpha|\, c(v)\, \frac{dv}{v}\, \|f\|_2 \, .$$

To estimate the last integral we write it as the sum

$$\int_0^{1/N} |1 - v^\alpha|\, c(v)\, \frac{dv}{v} + \int_{1/N}^{N} |1 - v^\alpha|\, c(v)\, \frac{dv}{v} + \int_N^\infty |1 - v^\alpha|\, c(v)\, \frac{dv}{v} = \mathrm{I}_1 + \mathrm{I}_2 + \mathrm{I}_3 \, .$$

Using the estimate (4.1) for $c(v)$ we can find $N = N_0$ sufficiently large so that I_1 and I_3 are less than $1/4$ uniformly with respect to α with α in $(0, \gamma']$ for a fixed γ' less than γ. Having chosen N_0 we can find an α_0 such that for $0 < \alpha < \alpha_0$, I_2 is less than $1/2$. Therefore $\|I + \alpha^2 T_\alpha\| < 1$, and hence $-\alpha^2 T_\alpha$ is invertible and therefore so is T_α.

References.

[C] Coifman, R. R., Multiresolution analysis in non homogeneous media. In *Wavelets, Time Frequency Methods and Phase Space*, Proceedings of the International Conference, Marseille, December 14-18, 1987 (1990, 2nd ed.), J. M. Combes, A. Grossman, Ph. Tchamitchian, eds., Springer Verlag, 1990.

[Ch] Christ, M., A $T(b)$ theorem with remarks on analytic capacity and the Cauchy Integral. *Colloquium Math.* **60/61** (1990), 601-628.

[CV] Calderón, A. P., Vaillancourt, R., On boundedness of pseudodifferential operators. *J. Math. Soc. Japan* **23** (1971), 374-378.

[DJS] David, G., Journé, J. L., Semmes, S., Opérateurs de Caldéron-Zygmund, fonctions para-accrétives et interpolation. *Rev. Mat. Iberoamericana* **1** (1985), 1-56.

[F] Folland, G. B., *Harmonic Analysis in Phase Space.* Annals of Math. Studies **122**. Princeton Univ. Press, 1989.

[GGW] Gatto, A. E., Gutiérrez, C. E., Wheeden, R. L., On weighted fractional integrals. In *Conference on Harmonic Analysis in Honor of Antoni Zygmund*, vol. 1, 124-137. Wadsworth International Group, 1983.

[GV] Gatto, A. E., Vági, S., Fractional integrals on spaces of homogeneous type. In *Analysis and Partial Differential Equations*. Cora Sadosky ed., Marcel Dekker, 1990.

[MS] Macías, R. A., Segovia, C., Lipschitz functions on spaces of homogeneous type. *Advances in Math.* **33** (1976), 257-270.

[N1] Nahmod, A. R., Geometry of operators and spectral analysis on spaces of homogeneous type. *C. R. Acad. Sci. Paris, Serie I* **313** (1991), 721-725.

[N2] Nahmod, A. R., Geometry of operators and spectral analysis. Thesis, Yale University, 1991.

[S1] Stein, E. M., The Characterization of functions arising as potentials I. *Bull. Amer. Math. Soc.* **67** (1961), 102-104. II (ibid), **68** (1962), 577-582.

[S2] Stein E. M., *Singular Integrals and Differentiability Properties of Functions*. Princeton Univ. Press, 1970.

[S3] Stein E. M., *Harmonic Analysis: Real Variable Methods, Orthogonality and Oscillatory Integrals*. Princeton Univ. Press, 1993.

[SZ] Stein, E. M., Zygmund, A., On the differentiability of functions. *Studia Math.* **23** (1964), 247-283.

[Z] Zygmund, A., *Trigonometric Series*, 2nd ed. Cambridge Univ. Press, 1959.

Recibido: 6 de febrero de 1.995
Revisado: 25 de mayo de 1.995

A. Eduardo Gatto* and Stephen Vági
DePaul University
Chicago, IL 60614, U.S.A.
aegatto@condor.depaul.edu and svagi@condor.depaul.edu

Carlos Segovia
Universidad de Buenos Aires and
Instituto Argentino de Matemática (CONICET)
ARGENTINA
segovia@iamba.edu.ar

* Part of this paper was done while this author was on sabbatical leave from DePaul University, visiting the Instituto Argentino de Matematica (CONICET) whose hospitality is gratefully acknowledged. It is also a pleasure for this author to thank Professor Carlos Kenig for pointing out to him the results of A. Nahmod in her thesis.

Revista Matemática Iberoamericana
Vol. 12, N.° 1, 1996

The heat kernel on Lie groups

N. Th. Varopoulos

0. Introduction.

0.1. Background: Lie groups.

In this paper G will denote a real connected amenable Lie group; $N \subset Q \subset G$ will denote the radical and the nilradical. Amenability on G is equivalent to the fact that the (semisimple) group $G/Q = S$ is compact. The group S can then be identified locally to some closed subgroup of G (a Levi subgroup).

If G is simply connected then all the groups N, Q, S are simply connected and $G = Q \lambda S$ is a semidirect product. We shall suppose throughout until the final section that G is simply connected. The choice of S is not unique but $G/N = Q/N \times S$ where the product is now direct (and not just semidirect *cf.* [1]) and $Q/N = V (\cong \mathbb{R}^a)$ furthermore S identified to a subgroup of G/N is then unique. Indeed if $\pi : G/N \longrightarrow Q/N \cong \mathbb{R}^a$ is the projection that we obtain from such an identification then $\pi(S_1) = \{0\}$ for any other subgroup S_1 that is either compact or semisimple. Let now $\Delta = -\sum X_j^2$ be some subelliptic left invariant "Laplacian" on G ,*i.e.*, $X_1, \dots, X_k$ are left invariant vector fields (*i.e.*, $(Xf)_g = Xf_g$; $f_g(x) = f(gx)$) that satisfy the Hörmander condition and generate together with all their brackets the tangent space on G (*cf.* [2]). Using the natural (uniquely defined!) projections

$$G \longrightarrow V \times S \longrightarrow V \tag{0.1}$$

we can define $\dot{\Delta}$ an elliptic operator V by projecting Δ on V. There exists therefore one and only one (up to orthogonal transformation) choice of coordinates $V \cong \mathbb{R}^a$ (in other words the corresponding scalar product on V is uniquely defined!) for which

$$\dot{\Delta} = -\sum_{i=1}^{a} \frac{\partial^2}{\partial^2 x_i} . \tag{0.2}$$

I shall denote by $\langle \cdot , \cdot \rangle_\Delta = \langle \cdot , \cdot \rangle$ the above well determined scalar product on V.

If we make the additional assumption that the Lie algebra $\mathfrak{g}$ of G is algebraic (or equivalently that G coincides locally with an algebraic group) then the structure of G simplifies further since we can then find $\mathbb{R}^a \cong V \subset Q$ such that $Q = N \lambda V$, and S an appropriate Levi subgroup, such that V and S commute (*cf.* [3]). We have thus a representation

$$G = N \lambda (V \times S) . \tag{0.3}$$

This is the model that the reader should keep in mind in what follows. We shall explain the correct substitute of (0.3) for the general (*i.e.* not necessarily algebraic) groups.

0.2. Background: The heat diffusion semigroup and kernel.

The notations are as in Section 0.1. I shall denote by $T_t = e^{-t\Delta}$ the heat diffusion semigroup generated by Δ and I shall also denote by

$$d^l g = dg , \quad d^r g = dg^{-1} , \quad m(g) = d^r g / dg$$

the left and right Haar measure on G and the modular function. I shall also denote by

$$\tilde{T}_t = m^{1/2} \, T_t \, m^{-1/2} = \exp(-t\tilde{\Delta}) = \exp(-t \, m^{1/2} \, \Delta \, m^{-1/2}) , \quad t > 0 .$$

The semigroup T_t (respectively $\tilde{T}_t$) is symmetric with respect to $d^r g$ (respectively dg). I shall denote by $\phi_t(g)$ (respectively $\psi_t(g)$) the convolution kernel of T_t (respectively $\tilde{T}_t$) with respect to dg and by $\mu_t \in P(G)$ the corresponding convolution measures:

$$T_t f(x) = \int \phi_t(y^{-1}x) \, f(y) \, dy = f * \mu_t(x) = \int f(xy^{-1}) \, d\mu_t(y) ,$$
$$\tilde{T}_t f(x) = \int \psi_t(y^{-1}x) \, f(y) \, dy \qquad x \in G , \ t > 0 , \ f \in C_0^\infty(G) ,$$

where

$$\psi_t(g) = m^{1/2}(g)\,\phi_t(g) = \psi_t(g^{-1})$$

and

$$\begin{aligned} d\mu_t(g) &= d\mu_t(g^{-1}) = \phi_t(g)\,d^r g = m(g)\,\phi_t(g)\,dg \\ &= m^{1/2}(g)\,\psi_t(g)\,dg = m^{-1/2}(g)\,\psi_t(g)\,d^r g\,. \end{aligned}$$

The semigroup T_t defines a diffusion on G: $\Omega = \{z(t) \in G : \ t > 0\}$ and for the corresponding probabilities on the path space Ω we denote as usual

$$\mathbf{P}_x(z(t) \in dy) = P_t(x, dy) = \mathbf{P}(z(t) \in dy : \ z(0) = x)$$

and we have

$$T_t f(x) = \int f(y)\,P_t(x, dy)\,.$$

We therefore have

$$d\mu_t(g) = d\mu_t(g^{-1}) = P_t(e, dg)\,.$$

0.3. The disintegration of the Haar measure and the L^p-norms.

If we use the projection $G \longrightarrow G/N = A = V \times S$ we can disintegrate the Haar measure

$$\int f(g)\,d^r g = \int_A \Big(\int_N f(na)\,dn \Big)\,da$$

this we shall summarise by saying

$$g = n\,a\,, \quad d^r g = dn\,da\,, \quad n \in N\,, \quad a \in A\,,$$

similarly

$$\int f(g)\,dg = \int_A \Big(\int_N f(an)\,dn \Big)\,da\,,$$

where $g = a\,n$, $dg = dn\,da$, $n \in N$, and $a \in A$.

The notations in the above integrals are of course clear enough but somewhat abusive. Indeed A cannot necessarily be identified to a subgroup of G but there is always some section $A \longrightarrow G$ (of $G \longrightarrow A$) (*cf.* [4]) and this identifies A as a subset of G. (In fact the above iterated integrals make obvious sense even without the existence of the above section (*cf.* [5])). We shall now fix $A \subset G$ such a section and express the diffusion

$$\Omega = \{z(t) = n(t)\,a(t) : \; t > 0\}\,, \quad n(t) \in N\,,\; a(t) \in A\,,\; t > 0\,.$$
$$\Omega_A = \{a(t) \in A : \; t > 0\}\,.$$

Ω_A corresponds to the path space of the diffusion on $A = V \times S$ generated by Δ_A the image of Δ by $G \longrightarrow G/N = A$.

Let us denote by $G_t^A(x)$ ($x \in A$) the convolution kernel of $e^{-t\Delta_A}$ on A. Here the notation G_t^A has been deliberately chosen to invoque the Gaussian functions:

$$G_t^a(x) = \frac{1}{(4\pi t)^{a/2}} \exp\Big(-\frac{|x|^2}{4t}\Big)\,, \qquad x \in \mathbb{R}^a\,,\; t > 0\,.$$

because when $S = \{e\}$ reduces to the identity we have $G_t^A = G_t^a$. The case $S = \{e\}$ is the model that the reader should keep in mind.

The reader should also keep in mind that we always have

$$\int_S G_t^A(x,s)\,ds = G_t^a(x)\,, \qquad x \in V \cong \mathbb{R}^a\,,$$

this by the standard local Harnack estimate gives

$$C^{-1} G_{t-1}^a(x) \le G_t^A(x,s) \le C\, G_{t+1}^a(x)\,, \tag{0.4}$$

for all $t \geq 10$, and $(x,s) \in V \times S$. With the above notations we have

$$\phi_t(na)\,dn = \mathbf{P}_e(n(t) \in dn : \; a(t) = a)\, G_t^A(a)\,. \tag{0.5}$$

The above disintegration allows us now to write the following formulaes:

$$\begin{aligned}
\|\tilde{T}_t\|_{1\to p}^p = \|\psi_t\|_p^p &= \int \psi_t^p(g)\,dg = \int \psi_t^p(g)\,d^r g \\
&= \int \psi_t^p(g)\,m(g)\,dg = \int \psi_t^p(g)\,m^{-1}(g)\,d^r g \\
&= \int_A dx \int_N \psi_t^p(xn)\,dn = \int_A dx \int_N \psi^p(nx)\,dn \\
&= \int_A dx\; m(x) \int_N \psi_t^p(xn)\,dn = \int_A dx\; m^{-1}(x) \int_N \psi^p(nx)\,dn\,,
\end{aligned}$$

$$\int_N \phi_t(nx)\,dn = m(x)\int \phi_t(xn)\,dn = G_t^A(x)\,,$$

$$\int \psi_t(nx)\,dn = m^{1/2}(x)\,G_t^A(x)\,, \qquad \int \psi_t(xn)\,dn = m^{-1/2}(x)\,G_t^A(x)\,.$$

here and throughout we denote by $\|\cdot\|_p$ the norm in $L^p(G;dg)$ and for any operator we denote by $\|\cdot\|_{p\to q}$ the norm in $L^p(G;dg) \longrightarrow L^q(G;dg)$.

0.4. The roots and the modular function.

Let $\mathfrak{n} \subset \mathfrak{q} \subset \mathfrak{g}$ be the Lie algebra of G, its radical and its nilradical. We have of course $\mathfrak{g}/\mathfrak{n} = (\mathfrak{q}/\mathfrak{n}) \times \mathfrak{s}$ where $\mathfrak{s}$ is the Lie algebra of S. The *ad* action on $\mathfrak{g}$ induces

$$ad_{\mathfrak{n}_\mathbb{C}} q = \begin{pmatrix} \rho_1(q) & & * \\ & \ddots & \\ 0 & & \rho_n(q) \end{pmatrix}, \qquad q \in \mathfrak{q}\,.$$

linear endomorphisms on $\mathfrak{n}_\mathbb{C} = \mathfrak{n} \otimes \mathbb{C}$ the complexified of $\mathfrak{n}$ which for an appropriate basis of $\mathfrak{n}_\mathbb{C}$ can be simultaneously triangulated for every $q \in \mathfrak{q}$ (Lie's theorem, *cf.* [1], [6], *cf.* also [7] where the above roots are systematically examined from a point of view that is adapted to our needs). We can identify $\rho_j \in \mathrm{Hom}_\mathbb{R}(\mathfrak{q};\mathbb{C})$ and we shall denote $L_j = \mathrm{Re}\,\rho_j \in \mathfrak{q}^* = \mathrm{Hom}_\mathbb{R}(\mathfrak{q};\mathbb{R})$, $(j = 1,\ldots,n)$. It is clear that $\rho_j|_\mathfrak{n} \equiv 0$ and therefore $\rho_j \in \mathrm{Hom}_\mathbb{R}(\mathfrak{q}/\mathfrak{n};\mathbb{C})$.

The space $\mathfrak{q}/\mathfrak{n}$ can be canonically identified with $V = Q/N$ and therefore we can identify the ρ_j's with functions on $V \times S$ and therefore also on G. (What is more natural is to identify the e^{ρ_j}'s with functions on G).

The definition of *ad* that we have adopted above is the differential at the identity of $Ad : G \longrightarrow GL(\mathfrak{g})$. And

$$Ad(g) = d_h I_g(h)|_{h=e}\,, \qquad I_g : G \longrightarrow G\,, \quad I_g(h) = h^g = ghg^{-1},$$

in terms of the roots we have then

$$Ad(q)|_{\mathfrak{n}_\mathbb{C}} = \begin{pmatrix} e^{\rho_1(q)} & & * \\ & \ddots & \\ 0 & & e^{\rho_n(q)} \end{pmatrix}, \qquad q \in Q\,.$$

Indeed this clearly holds when $q \in \operatorname{Exp}(\mathfrak{q})$ and therefore also for all q.

An elementary verificati on that has to be carried out by the reader shows therefore that with the above identifications we have:

$$\exp\left(\sum \rho_j(g)\right) = \exp\left(\sum L_j(g)\right) = m(g), \qquad g \in G.$$

We shall finally recall the following.

Definition. *We shall say that the group G is* NC-(*respectively* WNC) *if there exists $x \in V$ such that $L_j(x) > 0$ for all $L_j \neq 0$, $j = 1, \dots, n$ (respectively, $L_j(x) \geq 0$, $j = 1, \dots, n$, $\sum L_j(x) > 0$). When G is semi-simple or $\{0\}$ and the roots are not defined we say that G is* NC *but not* WNC.

The C-condition was first introduced in [8] (C stands there for "condition". I guess I should have called it the V condition, now is too late to do anything about it). NC-stands for Non-C, WNC stands for weak NC which is slightly abusive. Observe that if we suppose that the L_i's span V^*, the negation of WNC is that for every $0 \neq x \in V$ there exist $1 \leq j, k \leq n$ such that $L_j(x) < 0$, $L_k(x) > 0$. A unimodular group is NC if and only if $L_j = 0$ ($j = 1, 2, \dots, n$) and such a group is not WNC.

0.5. The main estimate.

The upper estimate. All the notations are as before. We have for the supremum on $n \in N$ over each fiber:

$$\begin{aligned}\sup_{n\in N} \phi_t(nx) &= \sup_{n} \phi_t(xn) \\ &\leq C_\varepsilon \exp\left(-\sum L_j^+(x)\right) G^A_{(1+\varepsilon)t}(x),\end{aligned} \tag{0.6}$$

for all

$$t \geq t_\varepsilon, \qquad x \in G/N = A = V \times S,$$

and where we denote as usual $r^+ = \sup\{r, 0\}$, $r^- = \inf\{r, 0\}$, $(r \in \mathbb{R})$ and where $0 < \varepsilon < 1$ is arbitrary and C_ε, $t_\varepsilon > 0$ depends on ε.

Equivalently the above estimates can be formulated in terms of ψ_t:

$$\begin{aligned}\sup_{n} \psi_t(nx) &= \sup_{n} \psi_t(xn) \\ &\leq C_\varepsilon \exp\left(-\frac{1}{2}\sum |L_j(x)|\right) G^A_{(1+\varepsilon)t}(x).\end{aligned} \tag{0.7}$$

This estimate together with (0.5) implies at once

$$\left(\int \phi_t^p(nx)\,dn\right)^{1/p} \leq C_\varepsilon \exp\left(\left(1-\frac{1}{p}\right)\left(-\sum L_j^+(x)\right)\right) G^A_{(1+\varepsilon)t}(x) \tag{0.8}$$

and

$$\int \psi_t^p(nx)\,dn \leq C_\varepsilon \exp\left(\left(1-\frac{p}{2}\right)\sum L_j^+(x) + \frac{p}{2}\sum L_j^-(x)\right)(G^A_{(1+\varepsilon)t}(x))^p\,, \tag{0.9}$$

$$\int \psi_t^p(xn)\,dn \leq C_\varepsilon \exp\left(-\frac{p}{2}\sum L_j^+(x) + \left(\frac{p}{2}-1\right)\sum L_j^-(x)\right)(G^A_{(1+\varepsilon)t}(x))^p\,, \tag{0.10}$$

for all

$$t \geq t_\varepsilon\,, \qquad x \in A = V \times S\,, \qquad 1 \leq p \leq +\infty\,.$$

If we integrate the above on A (with respect to dx) we obtain at once upper estimates of $\|\psi_t\|_p$. Observe also that because of (0.4) we can replace $G_t^A(\cdot)$ by the standard Gaussian on V.

The lower estimate. The most convenient way to express the lower estimate is through probabilistic language. We shall show that for all $x \in V$ and all $t \geq 10$ there exists $P \subset N$ some subset (that depends on x and t) such that:

$$\begin{aligned}\text{Vol.} &= \text{Haar-mes}_N(P)\\ &\leq C \exp\left(c\,t^{1/3} + C\frac{|x|}{\sqrt{t}} + \sum L_j^+(x)\right),\end{aligned} \tag{0.11}$$

$$\text{Pr.} = \mathbf{P}(z(t) \in P \times B_x) \geq C \exp(-c\,t^{1/3})\,G_t^a(x)\,, \tag{0.12}$$

where $B_x = \{y \in V : |x-y| \leq 1\} \times S \subset A$ and where A has been identified to some fixed section of $G \longrightarrow G/N$ as in Section 0.3.

The above estimate gives of course information about the rapidity with which the mass goes to infinity under the diffusion $\{z(t) : t > 0\}$. An obvious application of Hölder's estimate gives at once

$$\text{Pr.} \leq \Big(\int_{N \times B_x} \phi_t^p(nx)\, dn\, dx \Big)^{1/p} (\text{Vol.})^{1-1/p}, \quad 1 \leq p \leq +\infty .$$

This together with the local Harnack estimate gives

$$\begin{aligned} &\Big(\int_N \phi_t^p(ny)\, dn \Big)^{1/p} \\ (0.13) \qquad &\qquad \geq C_\varepsilon \exp\Big(-c\,t^{1/3} - \Big(1 - \frac{1}{p}\Big) \sum L_j^+(x) \Big) G^a_{(1-\varepsilon)t}(x) , \end{aligned}$$

for all $1 \leq p \leq +\infty$, $t \geq 10$, $0 < \varepsilon < 1$, and $y = (x, \sigma) \in V \times S$.

It follows in particular that up to a factor $e^{-ct^{1/3}}$ (which in view of [4], [9] is perfectly natural) the upper and lower estimates that we have given are sharp.

NC-lower estimates. If we make the additional assumption that the group G is NC we can make a substantial improvement in the above lower estimates! In that case we can replace the factors $\exp(\pm c\,t^{1/3})$ in the right hand side of (0.11) and (0.12) by the polynomial factors $t^{\pm c}$ respectively. This allows us to improve the estimate (0.13) and obtain

$$\Big(\int_N \phi^p(ny)\, dn \Big)^{1/p} \geq C_\varepsilon\, t^{-c} \exp\Big(- \Big(1 - \frac{1}{p}\Big) \sum L_j^+(x) \Big)\, G^a_{(1-\varepsilon)t}$$

with the same notations as in (0.13).

An analogous refinement that refers to the upper estimate holds for C-groups (*cf.* end of Section 1.2).

0.6. The Hardy-Littlewood theory.

The first thing to observe is that

$$\liminf_{t\to\infty} \frac{1}{t} \log \|\tilde{T}_t\|_{p\to q} \geq 0 , \qquad 1 \leq p, q < +\infty .$$

Let us consider the spectral decomposition of $\tilde{T}_t$ in $L^2(G; dg)$:

$$\tilde{T}_t = \int_0^\infty e^{-\lambda t}\, dE_\lambda .$$

By the amenability of G it follows that for all $\lambda > 0$ there exists $\varphi \in C_0^\infty$ such that $\langle E_\lambda \varphi, \varphi \rangle \geq c > 0$ and therefore $\langle \tilde{T}_t \varphi, \varphi \rangle > c\, e^{-\lambda t}$ $(t > 0)$.

We have, on the other hand,

$$\|\tilde{T}_t\|_{p \to q} = O(1)\,, \qquad 1 \leq p \leq 2 \leq q \leq +\infty\,.$$

This is easy and was pointed out in [10] (*cf.* (4.1) below). Let us consider

$$\ell(q) = \limsup_{t \to \infty} \frac{1}{t} \log \|\tilde{T}_t\|_{1 \to q} = \limsup_{t \to \infty} \frac{1}{t} \log \|\psi_t\|_q$$

which is a convex function of $1/q \in [0,1]$ and identically 0 for $q \in [2, +\infty]$. It follows that $\ell(q)$ is continuous for $q \in (1, +\infty]$ and in fact it is also continuous for $q = 1$. This last point is of course a consequence of the explicit formulaes given below but can also be seen in a more "abstract" way. Indeed by the "general" Gaussian estimates for ψ_t we see that $\ell(q) = \limsup (1/t) \log \|\psi_t\|_q$ can be defined and is finite and convex in q for all $q \in (0, +\infty]$. The continuity follows. At any rate the estimates of Section 0.5 allows us to obtain the following more precise information.

Theorem 1. *Let $L_1, \ldots, L_n \in V^* = (\mathfrak{q}/\mathfrak{n})^*$ be the real parts of the roots as in Section* 0.4 *and let $\langle \cdot\,, \cdot \rangle$ the scalar product on $V = \mathfrak{q}/\mathfrak{n}$ defined in Section* 0.4. *We have then*

$$\begin{aligned} \ell(p) &= \lim_{t \to \infty} \frac{1}{t} \log \|\psi_t\|_p \\ &= \lim_{t \to \infty} \frac{1}{t} \log \int_V \exp\Big(\Big(\frac{1}{p} - \frac{1}{2}\Big) \sum L_j^+(x) \\ &\qquad\qquad + \frac{1}{2} \sum L_j^-(x) - \frac{|x|^2}{4t} \Big)\, dx\,. \end{aligned} \tag{0.14}$$

In (0.14) $|\cdot|^2 = \langle \cdot\,, \cdot \rangle$ denotes of course the corresponding euclidean norm on V. We choose to set $V = \mathfrak{q}/\mathfrak{n}$ rather than Q/N because formulated like this the theorem makes sense even for non simply connected groups. The point is that the above theorem holds as stated for general, not necessarily simply connected, groups. An explicit formula for the limit (0.14) can also be given by an elementary computation.

What is significant is the first point where $\ell(q)$ vanishes:

$$L = \inf\{q \geq 1 :\ \ell(q) = 0\} \leq 2\,.$$

We have $L = 1$ if and only if G is unimodular. To see this we can either use our theorem or we can use the value of $\|\tilde{T}_t\|_{1\to 1}$ (*cf.* Theorem 2.iv) and [10]) and the continuity of $\ell(q)$.

An important consequence of the theorem is the following:

Theorem 2. *Let $G, \tilde{T}_t$ and L be as above. Then*

i) *The constant $1 \le L \le 2$ depends only on G and is in particular independent of the particular choice of Δ.*

ii) *We have $L = 1$ if and only if G is unimodular. We have $L = 2$ if and only if G is* WNC.

iii) *If G is not* WNC, *for every $1 \le p < 2$ there exists $p < q < 2$ such that*

$$\|\tilde{T}_t\|_{p\to q} = O(1).$$

iv) *For any group and any $1 \le p \le +\infty$ we have*

$$\|\tilde{T}_t\|_{p\to p} = \exp\left(t\left(\frac{1}{2} - \frac{1}{p}\right)^2 \rho^2\right),$$

where ρ^2 is defined by $\Delta m = -\rho^2 m$. If G is WNC *we have*

$$\lim \frac{1}{t} \log \|\tilde{T}_t\|_{p\to q} = \left(\frac{1}{2} - \frac{1}{q}\right)^2 \rho^2, \qquad 1 \le p \le q \le 2.$$

v) *Conversely if G is non unimodular and if*

$$\limsup \frac{1}{t} \log \|\tilde{T}_t\|_{p\to q} \ge \left(\frac{1}{2} - \frac{1}{q}\right)^2 \rho^2 \tag{0.15}$$

for some $1 \le p < q < 2$, then G is a WNC *group.*

Once more the above theorem holds for general, not necessarily simply connected groups.

The main problem of Hardy-Littlewood theory in the above context is to find all the combinations $\alpha \in \mathbb{C}$, $\operatorname{Re}\alpha \ge 0$ for which the following mapping

$$\begin{aligned} &\tilde{\Delta}^{-\alpha/2} : L^p(G; dg) \longrightarrow L^q(G; dg), \\ &\tilde{\Delta}^{-\alpha/2} = c_\alpha \int_0^\infty t^{\alpha/2-1}\, \tilde{T}_t \, dt \end{aligned} \tag{0.16}$$

is bounded. From our previous results on $\tilde{T}_t$ we can obtain some partial but significant information in that direction. First of all, the operator

$$\tilde{\Delta}^{-\alpha/2} = c_\alpha \int_0^\infty t^{\alpha/2-1}\, \tilde{T}_t\, dt \geq 0\,, \qquad \alpha > 0\,, \tag{0.17}$$

is positive and therefore $\tilde{\Delta}^{-\alpha/2} f$ is defined (possibly $+\infty$) for every $f \geq 0$. We have then

Corollary 1. *If G is a* WNC*-group then $\|\tilde{\Delta}^{-\alpha/2}\|_{p\to q} = +\infty$ for every $1 \leq p < q < 2$, $\alpha > 0$.*

The notation $\|\cdot\|_{p\to q} = +\infty$ is abusive but in view of (0.17) it has an obvious meaning. This corollary says that the mapping (0.16) cannot be bounded unless we cross the L^2- level. What happens for $1 \leq p \leq 2 \leq q \leq +\infty$ is in general an open question. Partial results can be found in [10], [11], [12], [13].

The above corollary essentially characterises WNC groups. Recall that quite generally there exists $\delta = 1, 2, \ldots$ such that $\psi_t(e) \sim t^{-\delta/2}$ $(t \to 0)$. We have then

Corollary 2. *Assume that G is neither a* NC *nor a* WNC *group, and let $1 \leq p < 2$. Then there exists $p \leq L(p) < 2$ such that for all $q \in]L(p), +\infty]$ we have*

$$\|\tilde{T}_t\|_{p\to q} = o(e^{-ct^{1/3}})\,, \qquad as\ \ t \to \infty\,,$$

for some $c > 0$. Furthermore if $1 < p \leq L(p) < q < +\infty$ the mapping (0.16) is bounded if and only if $1/p - 1/q \leq \mathrm{Re}\,\alpha/\delta$.

1. The upper estimate.

1.1. The general set up.

Let G be a simply connected Lie group, let $N \subset Q \subset G$ be the radical and nilradical, let $H = Z(N) \subset N$ be the center of N, let also $\alpha : G \longrightarrow G/H$ be the canonical projection. Let us also recall that $G/N \cong V \times S$ where $V \cong \mathbb{R}^m$ and S is semisimple (and compact if G is amenable). In this section I shall rely very heavily on the ideas the methods and even the notations introduced in [8] (*e.g.*, I shall set

$V \cong \mathbb{R}^m$ rather than $\mathbb{R}^a$) and there is no way at all that the reader can read this section without constantly refering back to [8].

The crucial thing is that (*cf.* [8, Section 1] where the notations are the same as above) G/H acts by inner automorphisms on H because H is abelian. Furthermore the estimates of [8, Section 1] hold. The extra "twist" for us refers to [8, Section 2]: H being central in N it follows that the above action of $G/H \longrightarrow GL(H)$ is trivial on N/H so that we can factorise this action by

$$G/H \xrightarrow{\pi} G/N = V \times S \longrightarrow GL(H).$$

Furthermore the estimates on the determinants given in [8, Section 2] apply. We then move on to [8, Section 3] to observe that everything applies *verbatim* here except that $G/H \ncong V \times S$ and therefore in the last few lines [8, Section 3] we have to define

$$A_n(L_i) = \inf_{1 \leq j \leq n} \exp\left(c\,|g_j|^2_{G/H} - L_i(b_j)\right),$$
$$(b_j, \sigma_j) = \pi(s_j) \in V \times S, \qquad i = 1, 2, \ldots, p.$$

To simplify notations I have droped here and in what follows the cofactor d_i (the dimension of the corresponding root space). I shall denote by $L_1, \ldots, L_p$ the real parts of the roots of the action of G/H and identify them with a subset of the real parts of the roots of G.

With all the other notations being as in [8, Section 3] the estimate (3.3) of [8] therefore holds in our setting and if we follow closely [8, Section 3] we see that for every $r \in G/H$ we have

$$(1.1)\quad \begin{aligned}\Big(\sup_{g \in \alpha^{-1}(r)} \phi_n(g)\Big)\, d^r r &\leq C\,\mathbb{E}\big(A_n(L_1)\cdots A_n(L_p);\ z_{G/H}(n) \in dr\big)\\ &= C\,\mathbb{E}\big(A_n(L_1)\cdots A_n(L_p)\, I\{z_{G/H}(n) \in dr\}\big),\end{aligned}$$

where $n \geq 1$. This is an inequality between two measures on G/H where $\{z_{G/H}(t) \in G/H\ :\ t > 0\}$ is the diffusion naturally induced on G/H by our laplacian Δ. The next step is to project $\pi(r) = (x, \sigma) \in V \times K$ and to distinguish two cases:

Case i): The $x \in G/H$ (*cf.* (0.6)) is such that $L_i(x) \leq 0$.

In the infimum that defines $A_n(L_i)$ we then set $j = 1$ and obtain

$$A_n(L_i) \leq C\,\exp(c\,|g_1|^2 + c\,|g_1|) \leq C\,\exp(2\,c\,|g_1|^2).$$

Case ii): $L_i(x) > 0$.

In the infimum that defines $A_n(L_i)$ we then set $j = n$ and obtain

$$A_n(L_i) \le C \exp(c\,|g_n|^2 - L_i^+(x))\,, \qquad b_n = x\,.$$

For $b_n = x$ we can therefore estimate the right hand side of (1.1) by

$$\begin{aligned}(1.2)\qquad &\exp\Big(-\sum L_i^+(x)\Big)\\ &\qquad\cdot\,\mathbb{E}\big(\exp(2\,c\,p\,(|g_1|^2_{G/H} + |g_n|^2_{G/H})\,;\ z_{G/H}(n) \in dr\big)\,,\end{aligned}$$

for $i = 1, 2, \dots, p$.

In the case when $H = N$ (which is the case considered in [8]) the above estimate simplifies since, S being compact, we can replace

$$|g_j|_{G/H} \approx |X_j|_V\,, \qquad (X_j, \tilde{\sigma}_j) = \pi(g_j) \in V \times S\,.$$

In general however this is not possible.

When $H = N$ the above estimate allows us to conclude very easily. Indeed we have then $dr = dx\, d\sigma$ and if we use the local Harnack estimate on the left hand side of (1.1) we see that since S is compact, by replacing n by $n-1$, and with $\alpha(g) = r = (x, \sigma) \in V \times S$ ($g \in G$), we have

$$\begin{aligned}\phi_{n-1}(g)\,dx \le C &\exp\Big(-\sum L_i^+(x)\Big)\\ &\cdot\,\mathbb{E}\big(\exp 2\,c\,p\,(|X_1|^2 + |X_n|^2)\,;\ b(n) \in dx\big)\,,\end{aligned}$$

for $n > 1$. (The notations are of course as in [8, Section 4] and we use interchangeably $b_j = b(j)$ ($j = 1, 2, \dots$) which is Brownian motion at times $t = 1, 2, \dots$). The coefficient of dx in the above expectation is equal to ($m = \dim V$)

$$(1.3)\quad n^{-m/2} \iint \exp\Big(\Big(2\,p\,c - \frac{1}{4}\Big)\,(|\xi|^2 + |\varsigma|^2) - \frac{|x - (\xi + \varsigma)|^2}{4\,(n-2)}\Big)\,d\xi\, d\varsigma\,.$$

A direct computation on the Gaussian shows that the double integral is comparable to

$$\exp\Big(-\frac{|x|^2}{4}\Big(\frac{1}{n} + \frac{a}{n^2}\Big)\Big)\,.$$

Alternatively Peter-Paul gives:

$$\begin{aligned}|x-(\xi+\varsigma)^2| &\geq |x|^2+|\xi+\varsigma|^2-\varepsilon_0\,|x|^2-\varepsilon_0^{-1}\,|\xi+\varsigma|^2\\ &=(1-\varepsilon_0)\,|x|^2+(1-\varepsilon_0^{-1})\,|\xi+\varsigma|^2\,, \qquad 0<\varepsilon_0\ll 1\,,\end{aligned}$$

which means that we can estimate (1.3) by

$$c_\varepsilon\, n^{-m/2}\exp\Big(-\frac{|x|^2}{(4+\varepsilon)\,n}\Big)\,, \qquad n\geq n_\varepsilon\,.$$

Provided of course that the c appearing in the exponential (1.2) can be taken small enough. As explained in [8] this is always possible, but highly *non* trivial to see. For that reason an alternative method towards estimating expressions as above was given in [7] and no assumption as to the smallness of $c>0$ was needed in that alternative method. The same thing applies here, but to avoid diverting the argument, we shall not give the details. With the help of [7] (or/and Section 1.4 further down) the reader can work this out for himself if he so wishes.

1.2. The inductive step.

Let G, N be as before (simply connected) and let:

$$H=Z(N)=N_k\subset N_{k-1}\subset\cdots\subset N_0=N$$

a central series of N so that H is the center of N. The upper estimates will be proved by induction on k. The case $k=0$ was dealt with in Section 1.1. What we shall do here is to prove the inductive step and show that if the estimate (0.6) holds for the group G/H, where the k is one unit lower, it also hold for G. The issue is clearly to estimate the expectation $\mathbb{E}(\cdots)$ that appears in (1.2). We shall change slightly the notations and write this expectation in the form

$$(1.4)\qquad \mathbb{E}\big(\exp(c\,(|g_1|^2+|g_n|^2)):\ g_1g_2\cdots g_n\in dr\big)\,,$$

where now $g_1,\ldots,g_j,\cdots\in G/H$ are independent, equidistributed, G/H valued, random variables with distribution $\mu_1^{G/H}(\,\cdot\,)$ the heat diffusion convolution measure on G/H at time $=1$ (*cf.* Section 0.2) (the $g_1g_2\cdots g_n$ in the expectation is of course a group product) using this formulation and the fact that $\mu_t(g) = \mu_t(g^{-1})$ it is clear that (1.4) is equal to

$$\iint_{\zeta,\xi\in G/H} \mathbb{E}\big(\exp(c(|\xi|^2+|\zeta|^2)) : g_2\cdots g_{n-1}\in \xi\, dr\, \zeta\big) \cdot \mathbf{P}(g_1\in d\xi)\,\mathbf{P}(g_n\in d\zeta). \tag{1.5}$$

If we use the inductive hypothesis which ensures that the estimate (0.6) holds for the group G/H we see that we can estimate the expectation inside the integral as follows

$$\mathbb{E}(\cdots) \le C_\varepsilon \exp\Big(c(|\xi|^2_{G/H}+|\zeta|^2_{G/H}) - \sum_j \tilde{L}_j^+(\xi\, r\, \zeta) - \frac{|\pi(\xi\, r\, \zeta)|_V^2}{(4+\varepsilon)\,n}\Big)\, m(\xi)\, d^r r, \tag{1.6}$$

where $\pi : G/H \longrightarrow V$ is the canonical projection (as in Section 1.1) and $\tilde{L}_j$ are the real parts of the roots of G/H. What is important is to note that each $\tilde{L}_j$ factors through π and is additive. We can thus absorb the $\tilde{L}_j^+(\xi)$, $\tilde{L}_j^+(\zeta)$ with the $c(|\xi|^2+|\zeta|^2)$ and therefore, if we bear in mind that $|g|_{G/H} \ge C\,|\pi(g)|_V$ we can estimate the coefficient of $d^r r$ in (1.6) by

$$\exp\Big(-\sum \tilde{L}_j^+(x) + c(|\xi|^2+|\zeta|^2) - \frac{|x-\pi(\xi)-\pi(\zeta)|_V^2}{(4+\varepsilon)\,n}\Big),$$

where $x=\pi(r)$. Peter-Paul is used as in (1.3) and it yields the estimate

$$\exp\Big(-\frac{|x|_V^2}{(4+\varepsilon)\,n} - \sum \tilde{L}_j^+(x)\Big)\exp\Big(c(|\xi|^2_{G/H}+|\zeta|^2_{G/H})\Big),$$

provided that n is large enough (depending on ε). The c in (1.6) can again be chosen in advance and as small as we like. We shall insert this estimate in (1.5) and use the Gaussian decay of $\mathbf{P}(g_i\in d\xi)$ on G/H (and not just on V), *cf.* [11]. If we use the estimate in (1.2) and bear in mind that the real parts of the roots $L_1,\ldots,L_p$ of the action of G/H on H (*cf.* Section 4.1) together with the root $\tilde{L}_1, \tilde{L}_2,\ldots$ make up for all the roots of G we see that the inductive step follows. A slightly more subtle computation gives as before the estimate $\exp(-|x|^2(1/n - c/n^2)/4)$ for the Gaussian contribution (instead of $\exp(-|x|^2/(4+\varepsilon)n)$).

1.3. An alternative approach.

In this section I shall explain how we can simplify considerably the proof of the upper estimate if we are prepared to loose a little at the end result.

Let $X \subset G$ be some compact subset and let

$$Y = \alpha(X) \subset G/H , \qquad Z = \beta(X) \subset V ,$$

where β is the composed mapping $\beta : G \longrightarrow G/N = V \times K \longrightarrow V$. Then (*cf.* (1.2)) there exists $C = C_X$ such that

$$\mathbf{P}(z(n) \in X) \le C \sup_{x \in Z} \exp(-\sum L_i^+(x))\, T \tag{1.7}$$

$$T = T(c) = \mathbb{E}\big(\exp(c\,|g_1|_G^2 + c|g_n|_G^2);\ \alpha(z(n)) \in Y\big) , \tag{1.8}$$

where the expectation in (1.8) refers to diffusion on G and where we have replaced $|\alpha(g)|_{G/H}$ by $|g|_G$ which is larger. The c's appearing in the exponential of (1.8) can again be assumed as small as we like. We can now use Hölder to estimate $T(c)$. Indeed for a given $p > 1$ if c is sufficiently small we have $\mathbb{E}(\exp(p\,c\,(|g_1|_G^2 + |g_n|_G^2)) < +\infty$ by the Gaussian estimate (on G) of the g_i's (*cf.* [7]). It follows that we can replace T in (1.7) by

$$T^* = (\mathbf{P}(\alpha(z(n) \in Y))^{1/q} ,$$

where $1/p + 1/q = 1$ are conjugate indices.

The estimate (1.7) becomes thus an inductive step that reduced the estimate of $\mathbf{P}_G(z(n) \in X)$ to the estimate of $\mathbf{P}_{G/H}(\alpha(z(n)) \in \alpha(X))$ with an arbitrary small loss at the exponent: $1/q$. We shall then set X = "dg" = "an appropiate small element", as in Section 1.1 and we shall examine the dependence of C_X on X. The details of this computation are not trivial but they will not be given here; they can be found in [17].

The end result of the above induction is our upper estimate in Section 0.5 with the L_i's replaced by $(1-\varepsilon)\, L_i$'s for an arbitrary $\varepsilon > 0$. With this approach however we obtain the estimate for $t \ge 1$ and *not* only for $t \ge t_0(\varepsilon)$. For all practical purposes this estimate is as good as the one we have in Section 0.5.

1.4. A general overview of upper estimates.

First of all by the definition (*cf.* Section 1.1) we have

$$A_n(L_i) \le B_n \cdot D_n \cdot C_n(L_i), \qquad i = 1, 2, \dots, p. \tag{1.9}$$

where

$$C_n(L_i) = \inf_{1\le j\le n} \exp(-L_i(b_j)),$$
$$B_n = \exp\big(\sup_{1\le r\le n} c(\inf_{kr\le j<(k+1)r} |g_j|_G^2)\big),$$
$$D_n = \exp\big(C\,r \sup_{1\le j\le n} |g_j|_G\big).$$

A few comments are in order: the c that appears in the definition of B_n is the same as the one in the definition of $A_n(L_i)$ but in the estimates that follow I shall not assume that it can be chosen as small as we like. The r in the definition of B is a new parameter and will be chosen later. To see (1.9) one simply samples the j in the infimum over $1 \le j \le n$ on the succesive blocks $[kr, (k+1)r)$ so as to pick up the inf $|g_j|_G^2$ in that block. Then one makes the appropriate correction bearing in mind that $|b_j - b_{j-1}|_V \le C\,|g_j|_G$.

Let X, Y, Z and the other notations be as in Section 1.3 and let us condition with respect to the projected path on V

$$\underline{\beta} = \{\beta(z(j)) = b_j \in V : \ 1 \le j \le n\} \subset V^n.$$

For the corresponding conditional expectations, going back to [8, Section 3] and by the same reasoning as in sections 1.1 and 1.3, we obtain

$$\begin{aligned}\mathbf{P}(z_G(n) \in X \,|\, \underline{\beta}) \le C_X\, \mathbb{E}(B_n^C D_n^C;\ z_{G/H}(n) \in Y \,|\, \underline{\beta})\\ \cdot \prod_i C_n(L_i)\, I(b(n) \in Z),\end{aligned} \tag{1.10}$$

where the product $\prod_i$ in the right hand side is taken as before over the roots of the action of G/H on H (as in Section 1.1). This product factors out since it only depends on $\underline{\beta}$. $I(\cdots)$ is the indicator function.

Using Hölder we can estimate the conditional expectation $\mathbb{E}(\,\cdot\,|\,\underline{\beta})$ by

$$\mathbb{E}(B_n^{Cp} \,|\, \underline{\beta})^{1/p}\, \mathbb{E}(D_n^{Cp} \,|\, \beta)^{1/p} \,(\mathbf{P}(z_{G/H}(n) \in Y \,|\, \beta))^{1/q},$$

where $2/p+1/q = 1$ and where clearly the q can be made as close to 1 as we like provided that p is large enough. It follows that (1.10) can be used as a recurrence formula that allows us to estimate $\mathrm{P}(z_G(n) \in X \mid \underline{\beta})$ in terms of $\mathrm{P}(z_{G/H}(n) \in \alpha(X) \mid \underline{\beta})$ with a small error $(1/q < 1, 1/q \sim 1)$ in the exponent.

If we repeat this procedure we finally see that we can estimate the left hand side of (1.10) by

$$I(b(n) \in Z) \prod_i C_n((1-\varepsilon_i)\, L_i) \Big(\prod_\alpha \mathbb{E}(B_n^{Cp_\alpha} \mid \underline{\beta})^{1/p_\alpha}\, \mathbb{E}(D_n^{Cp_\alpha} \mid \underline{\beta})^{1/p_\alpha} \Big),$$

where the $0 \leq \varepsilon_i \ll 1$ can be made arbitrarily small and the p_α's are appropriately large and where all the roots of G are now involved in the first product. We shall integrate the above estimate over the path space and use Hölder once more.

It follows that we can estimate

$$\text{(1.11)}\quad \mathrm{P}(z_G(n) \in X) \leq C_X\, \theta\, \mathbb{E} \prod_i \big(C_n((1-\eta_i)\, L_i)\, I(b(n) \in Z)\big)^{1-\delta_i},$$

where $0 \leq \eta_i, \delta_i \ll 1$ can be made arbitrarily small and where the cofactor θ is some product of $(\mathbb{E}(B_n^\alpha))^\beta$ and $(\mathbb{E}(D_n^\gamma))^\delta$ for various values of α, β, γ, $\delta > 0$. The cofactor θ admits a polynomial bound. Indeed we have:

$$\text{(1.12)}\qquad \mathbb{E}(B_n^C),\ \mathbb{E}(D_n^C) = 0(n^A).$$

This is clear for D_n by the (non trivial *cf.* [11]) Gaussian estimate that we have for the g_j's on G

$$\text{(1.13)}\qquad \mathrm{P}(|g_j|_G > \lambda) \leq C \exp(-c_0\, \lambda^2)$$

for some fixed but positive $c_0 > 0$. This implies

$$\mathrm{P}(\sup_{1\leq j\leq n} |g_j| > \lambda) \leq Cn \exp(-C_0\, \lambda^2).$$

The estimate (1.13) gives also

$$\mathrm{P}\big(\inf_{kr\leq j<(k+1)r} |g_j|_G^2 = \zeta_k > \lambda\big) \leq C^r \exp(-c_0\, r\, \lambda), \quad r = 1, 2, \ldots,$$

and therefore:

$$\mathbf{P}\big(\sup_{1\leq k\leq n} \xi_k \geq \lambda\big) \leq n\, C^r \exp(-c_0\, r\, \lambda)\,.$$

If r is large enough we can clearly "absorb" any exponent M in $E(B_n^M)$ and (1.12) follows for B_n.

The final "move" in this general approach is to examine the principal term $\prod_i(\cdots)I[\cdots]$ in the right hand side of (1.11), which is a Brownian functional and which can therefore be estimated by purely probabilistic (Brownian or random-walks) methods. Two cases have been examined in details

1) X is some neighbourhood of 0 in G and the linear functionals satisfy the C-condition. This was done in [8] and the estimate obtained there is $O(e^{-cn^{1/3}})$

2) Z is some neighbourhood of $x \in V$, some point "far out" on V. This is what was done in this paper.

It is of course possible to incorporate the C-condition in the case 2) and obtain an extra factor $e^{-ct^{1/3}}$ in front of the upper estimates of Section 0. The details will be left to the interested reader.

2. Dilation structure on a Lie algebra.

2.1. Algebraic considerations.

Let $\Lambda^t \in GL(V)$ $(-\infty < t < +\infty)$ be a one parameter group of automorphisms of the real vector space $V(\cong \mathbb{R}^a,\ a \geq 1)$. Let $V_{\mathbb{C}} = V \otimes_{\mathbb{R}} \mathbb{C}$ be the complexified space and let

$$V_{\mathbb{C}} = V_1^{\mathbb{C}} \oplus \cdots \oplus V_k^{\mathbb{C}}$$

the corresponding root space decomposition of $V_{\mathbb{C}}$ under Λ^t. In other words if $\lambda_1, \ldots, \lambda_k \in \mathbb{C}$ are the corresponding roots we have $x \in V_j$ $(1 \leq j \leq k)$ if and only if

$$(\Lambda^t - e^{\lambda_j t})^N x = 0\,, \qquad t \in \mathbb{R}\,,$$

for some $N \geq 1$ large enough (independently of j).

We shall combine together the subspaces $V_{i_s}^{\mathbb{C}}$, $1 \leq s \leq r$, for which the corresponding roots have the same real part $\operatorname{Re}\lambda_{j_s} = \gamma$, $1 \leq s \leq r$, and write

$$(2.1) \qquad V = V_1 \oplus \cdots \oplus V_p \,, \qquad ; V_j \otimes \mathbb{C} = V_{i_1}^{\mathbb{C}} \oplus \cdots \oplus V_{i_r}^{\mathbb{C}} \,.$$

The operator norms of Λ^t restricted to V_s satisfy

$$\|\Lambda^t|_{V_s}\| = O(e^{\gamma t}|t|^A)\,, \qquad \text{as } t \to \pm\infty\,.$$

This is simply because

$$\Lambda^t|_{V_s \otimes \mathbb{C}} = e^{\gamma t}\, T(t)\,,$$

where $T(t)$ is an upper triangular complex matrix with unimodular diagonal coefficients and therefore satisfies (*cf.* [7])

$$T(t) = T^{-1}(-t)\,, \qquad |T(t)| = O(|t|^A)\,.$$

Let us now suppose that $V = \mathfrak{g}$ is some real Lie algebra and that Λ^t are algebra automorphisms. We have then $\Lambda^t = \exp(tD)$ where $D \in \partial(\mathfrak{g})$ is some derivation of $\mathfrak{g}$. Let

$$\begin{gathered}\mathfrak{g} = \mathfrak{g}_{-\beta_1} \oplus \cdots \oplus \mathfrak{g}_{-\beta_s} \oplus \mathfrak{g}_0 \oplus \mathfrak{g}_{\alpha_1} \oplus \cdots \oplus \mathfrak{g}_{\alpha_r}\,,\\ -\beta_1 < \cdots < -\beta_s < 0 < \alpha_1 < \cdots < \alpha_r\,,\end{gathered}$$

be the corresponding decomposition as in (2.1) with

$$\gamma = \operatorname{Re}\lambda_{i_s} = -\beta_1, \ldots, 0, \alpha_1, \ldots$$

It is easy to see (*cf.* [6]) that we have $[\mathfrak{g}_{\gamma_1}, \mathfrak{g}_{\gamma_2}] \subset \mathfrak{g}_{\gamma_1+\gamma_2}$ if $\gamma_1 + \gamma_2$ is a real part of a root and $[\mathfrak{g}_{\gamma_1}, \mathfrak{g}_{\gamma_2}] = 0$ if not. It follows in particular that

$$\begin{aligned}\mathfrak{g}_+ &= \mathfrak{g}_{\alpha_1} \oplus \cdots \oplus \mathfrak{g}_{\alpha_r}\,, \qquad & \mathfrak{g}_+^0 &= \mathfrak{g}_+ \oplus \mathfrak{g}_0\,,\\ \mathfrak{g}_- &= \mathfrak{g}_{-\beta_1} \oplus \cdots \oplus \mathfrak{g}_{-\beta_s}\,, \qquad & \mathfrak{g}_-^0 &= \mathfrak{g}_- \oplus \mathfrak{g}_0\,,\end{aligned}$$

are subalgebras. Let $|\cdot|$ be some norm on $\mathfrak{g}$, it is then clear from the above that we have

$$(2.2) \qquad \begin{aligned}|\Lambda^t x| &= O(|t|^A)\,, \quad \text{as } t \to +\infty \qquad &\Longleftrightarrow \qquad x \in \mathfrak{g}_-^0\,,\\ |\Lambda^t x| &= O(|t|^A)\,, \quad \text{as } t \to -\infty \qquad &\Longleftrightarrow \qquad x \in \mathfrak{g}_+^0\,.\end{aligned}$$

From this it follows in particular that if $\mathfrak{h} \subset \mathfrak{g}$ is a subalgebra, or even a subspace, stable by Λ (*i.e.*, $\Lambda\mathfrak{h} \subset \mathfrak{h}$) then

$$\mathfrak{g}_+^0 \cap \mathfrak{h} \subset \mathfrak{h}_+^0 , \qquad \mathfrak{g}_-^0 \cap \mathfrak{h} \subset \mathfrak{h}_-^0 . \tag{2.3}$$

We shall now describe a rather technical construction that will be essential in what follows.

Proposition. *Let us assume that* $\mathfrak{g}$ *is nilpotent and let* $\mathfrak{h} \subset \mathfrak{g}$ *be a subalgebra that satisfies*

$$\mathfrak{g}_+ \subset \mathfrak{h} \neq \mathfrak{g} , \ (\textit{respectively}, \ \mathfrak{g}_+^0 \subset \mathfrak{h} \neq \mathfrak{g}) , \qquad \Lambda^t\mathfrak{h} \subset \mathfrak{h} .$$

Then we can find $y, z \in \mathfrak{g}_-^0$ *(respectively,* $\mathfrak{g}_-$*) such that the subalgebra*

$$\mathfrak{h}_1 = \mathrm{Alg}\,\{\mathfrak{h}, y, z\}$$

is stable by the action of Λ^t *(i.e.,* $\Lambda^t\mathfrak{h}_1 \subset \mathfrak{h}_1$*) and furthermore*

$$\mathfrak{h} \neq \mathfrak{h}_1 , \qquad [\mathfrak{h}, \mathfrak{h}_1] \subset \mathfrak{h} .$$

PROOF. Let us denote

$$\tilde{\mathfrak{h}} = (\mathfrak{h} \otimes \mathbb{C}) \curlywedge \mathbb{C} D \subset \tilde{\mathfrak{g}} = (\mathfrak{g} \otimes \mathbb{C}) \curlywedge \mathbb{C} D ,$$

where the skew product is defined by the action of the derivation $D \in \partial(\mathfrak{g})$ on $\mathfrak{g}$. The algebra $\tilde{\mathfrak{g}}$ is a complex soluble algebra and let $V = \tilde{\mathfrak{g}}/\tilde{\mathfrak{h}}$ which is a complex vector space on which $\tilde{\mathfrak{h}}$ acts by adjoint action. By Lie's theorem there exists therefore

$$0 \neq X = y + iz + \tilde{\mathfrak{h}} \in V , \qquad y, z \in \mathfrak{g} \ \ (y + iz \notin \tilde{\mathfrak{h}}) ,$$

such that

$$\xi(X) = \lambda(\xi) \cdot X , \qquad \xi \in \tilde{\mathfrak{h}} , \qquad \lambda(\,\cdot\,) \in \mathrm{Hom}_{\mathbb{C}}(\tilde{\mathfrak{h}}; \mathbb{C}) . \tag{2.4}$$

The fact that $\mathfrak{g}$ is nilpotent implies that every $\zeta \in \mathfrak{h}$ gives rise to a nilpotent transformation on V and therefore that

$$\lambda(\zeta) = 0 , \qquad \zeta \in \mathfrak{h} . \tag{2.5}$$

The fact that $\mathfrak{h} \supset \mathfrak{g}_+$ (respectively, $\supset \mathfrak{g}_+^0$) implies furthermore that we can assume that $y, z \in \mathfrak{g}_-^0$ (respectively, $\mathfrak{g}_-$). Now (2.5) implies that

$$[X, \mathfrak{h}] = [y, \mathfrak{h}] + i\,[z, \mathfrak{h}] \subset \tilde{\mathfrak{h}}\,.$$

From this it follows that

$$[y, \mathfrak{h}]\,, [z, \mathfrak{h}] \subset \mathfrak{h}\,.$$

What we have shown is that the subalgebra $\mathfrak{h}_1$ generated by $\{\mathfrak{h}, y, z\}$ is strictly larger than $\mathfrak{h}$ and normalises $\mathfrak{h}$. By (2.4) it follows that

$$[D, y] + i\,[D, z] \in \lambda(D)(y + iz) + \tilde{\mathfrak{h}}\,.$$

Let $\lambda(D) = \alpha + i\beta$ ($\alpha, \beta \in \mathbb{R}$) we then have

$$[D, y] \in \alpha\, y - \beta\, z + \mathfrak{h}\,, \qquad [D, z] \in \beta\, y + \alpha\, z + \mathfrak{h}\,.$$

This shows that $[D, \mathfrak{h}_1] \subset \mathfrak{h}_1$ and therefore that $\Lambda^t \mathfrak{h}_1 \subset \mathfrak{h}_1$. The above algebra $\mathfrak{h}_1$ satisfies therefore all the conditions of our proposition.

From the above proposition it follows that we can construct

$$\mathfrak{g}_+^0 = \mathfrak{h}_0 \subset \mathfrak{h}_1 \subset \cdots \subset \mathfrak{h}_k = \mathfrak{g}\,, \tag{2.6}$$

a finite chain of subalgebras such that

$$\mathfrak{h}_{j+1} = \mathrm{Alg}\,\{\mathfrak{h}_j, y_j, z_j\}\,, \qquad j = 0, 1, \ldots, k-1\,,$$

with $y_j, z_j \in \mathfrak{g}_-$ and so that $\mathfrak{h}_{j+1}$ normalises $\mathfrak{h}_j$ ($j = 0, 1, \ldots, k-1$) and

$$\Lambda^t \mathfrak{h}_j \subset \mathfrak{h}_j\,, \qquad j = 0, 1, 2, \ldots, k\,.$$

We shall end up this section by recalling an important lemma due to Hörmander (*cf.* [2], [11]).

Lemma (L. Hörmander). *Let $\mathfrak{g}$ be a nilpotent Lie algebra and let G be some Lie group that corresponds to $\mathfrak{g}$. Then there exists $N \geq 0$, $c_\alpha \in \mathbb{R}$ ($0 \leq \alpha \leq N$), $k_\alpha = 0, 1, \ldots$ ($1 \leq j \leq N$) such that*

$$\mathrm{Exp}\,(x + y) = \mathrm{Exp}\,(c_0\, y) \prod_{\alpha=1}^{N} \mathrm{Exp}\,(c_\alpha\, x_\alpha) \in G\,, \qquad x, y \in \mathfrak{g}\,,$$

where

$$x_\alpha = (ad(y))^{k_\alpha} x = [y, [y[\cdots[y, x]\cdots]\,, \qquad 1 \leq \alpha \leq N\,.$$

2.2. Positive and negative roots.

All the notations of the previous section will be preserved. $\mathfrak{g}$ will be assumed throughout to be nilpotent and we shall denote by $|\cdot|$ some norm on $\mathfrak{g}$ (the exact value of $|\cdot|$ will be irrelevant here). We shall denote by

$$A(a) = \{x \in \mathfrak{g} : |x| \leq a\}.$$

We shall say that $\mathfrak{g}$ is a positive (respectively, negative) algebra if $\mathfrak{g} = \mathfrak{g}_+^0$ (respectively, $\mathfrak{g} = \mathfrak{g}_-^0$).

We shall also consider $\mathrm{Exp} : \mathfrak{g} \longrightarrow G$ the exponential mapping where G is the simply connected nilpotent group associated to $\mathfrak{g}$. The Haar measure of G will be denoted by m_G (not to be confused with the previous notation for the modular function that here is identically 1) which is the image by Exp of Lebesgue measure on $\mathfrak{g}$. The Jacobian of $\Lambda^t : \mathfrak{g} \longrightarrow \mathfrak{g}$ is

$$\mathrm{Jacb}\,(\Lambda^t) = \exp\left(\left(\sum \alpha_j - \sum \beta_j\right)t\right),$$

where here and in what follows we count the α_j's and the β_j's with multiplicity (*i.e.*, they are tacitly multiplied by the dimensionality of the corresponding root spaces $\mathfrak{g}_{\alpha_j}, \mathfrak{g}_{-\beta_j}$).

The group Λ^t induces a one parameter group, also denoted by $\Lambda^t : G \longrightarrow G$, of group automorphisms and Exp interwines Λ^t. The norm $|\cdot|$ on $\mathfrak{g}$ induces a left (or right) invariant distance on G (*cf.* [11]), where for $g \in G$ we set

$$d(e,g) = \inf\Big\{ \sum |t_j| : \ g = \mathrm{Exp}\,(t_1 x_1)\,\mathrm{Exp}\,(t_2 x_2)\cdots, \quad x_j \in \mathfrak{g}, \ |x_j| \leq 1 \Big\}$$

and clearly

$$d(\Lambda x, \Lambda y) \leq |\Lambda|\, d(x,y), \qquad x, y \in G,$$

where $|\Lambda|$ denotes the operator norm of $\Lambda : \mathfrak{g} \longrightarrow \mathfrak{g}$.

Observe finally that when $\Lambda = Ad(g)$, $g \in G$ on $\mathfrak{g}$, then $\Lambda = I_g : h \mapsto ghg^{-1} = h^g$ on G. It follows therefore that for any normal subgroup $H \subset G$ we have

$$d_H(x^g, y^g) \leq |Ad(g)|\, d_H(x,y), \qquad g \in G, \ x, y \in H,$$

where $d_H(\cdot,\cdot)$ denotes the "intrinsic " distance on H (and not the distance induced by the ambient G).

Observe also that since the algebra $\mathfrak{g}$ is nilpotent the operator norm of $Ad(g)$ satisfies

$$|Ad(g)| \leq C\,(|g|+1)^C .$$

I shall denote

$$\begin{aligned} B(a) &= \{g \in G :\ d(e,g) \leq a\} \\ D(a) &= D_0(a) = \mathrm{Exp}\,(A(a))\,, \qquad a > 0\,. \end{aligned} \tag{2.8}$$

We have then

$$D(a) \subset B(a)\,, \qquad B(a) \subset D(a^C)\,, \qquad a > 1\,, \tag{2.9}$$

for some $C > 0$ where the second inclusion followes from the Baker-Campbell-Hausdorff formula (*cf.* [1, Section 2.15]). I shall denote

$$D_t(a) = \mathrm{Exp}\,(\Lambda^t A(a)) = \Lambda^t \mathrm{Exp}\,(A(a)) = \Lambda^t D_0(a)\,, \tag{2.10}$$

for $t \in \mathbb{R}$, $a > 0$. From the above we have

$$m_G(D_t(a)) = C \exp\left(\left(\sum \alpha_j - \sum \beta_j\right)t\right) a^{\dim \mathfrak{g}}\,. \tag{2.11}$$

Let us first suppose that $\mathfrak{g}$ is a *negative* algebra. We have then

$$D_{t_1}(a_1)\cdots D_{t_n}(a_n) \subset D_0\left(C\left(\sum a_j + \sum t_j + n\right)^C\right), \tag{2.12}$$

for $a_j \geq 0$ and $t_j \geq 0$ and where $C > 0$ only depends on Λ^t (the left hand side of (2.12) is, of course, a group product in G). Two special cases of the inclusion (2.12) are easy to prove. The case $t_1 = t_2 = \cdots = t_n = 0$ is a consequence of the Baker-Campbell-Hausdorff formula. The case $n = 1$ is a consequence of (2.2). The inclusion (2.12) in general follows immediately from the above two special cases.

Let us now suppose that $\mathfrak{g}$ is a *positive* algebra, let $t_1, \ldots, t_n \in \mathbb{R}$ and let $t = \max_{1 \leq j \leq n} t_j$. Then

$$\begin{aligned} D_{t_1}(a_1)\cdots D_{t_n}(a_n) &= \Lambda^t\big(D_{t_1-t}(a_1)\cdots D_{t_n-t}(a_n)\big) \\ &\subset D_t\left(C\left(\sum a_j + \sum |t_j| + n\right)^C\right), \end{aligned} \tag{2.13}$$

where the second inclusion follows from (2.12) and the fact that $t_j - t \leq 0$ $(1 \leq j \leq n)$. Indeed $\Lambda^{t_j - t} = \tilde{\Lambda}^{t - t_j}$ and $\mathfrak{g}$ is a negative algebra for the dilation Λ^{-t}.

Together with the above observations we shall need the following

Lemma. *Let us assume that* $\mathfrak{g}$ *is a positive algebra, then for each* $t, s > 0$ *there exist*

$$z_\nu, \tilde{z}_\nu \in G, \qquad 1 \leq \nu \leq M \leq C\,(s + t + 1)^C \exp\left(t \sum \alpha_j\right)$$

that satisfy

$$D_t(s) \subset \Big(\bigcup_\nu z_\nu\, D_0(1)\Big) \bigcap \Big(\bigcup_\nu D_0(1)\, \tilde{z}_\nu\Big) = R(M). \tag{2.14}$$

This lemma will be combined with (2.13) to yield

$$\begin{aligned} D_{t_1}(a_1) \cdots D_{t_n}(a_n) \subset R\Big(C\Big(\sum a_j + \sum |t_j| + n\Big)^C\Big) \\ \cdot \exp\left(t \sum \alpha_j\right). \end{aligned} \tag{2.15}$$

where we suppose that $t = \max(t_j) \geq 0$.

PROOF. Let us fix $d(\,\cdot\,,\,\cdot\,)$ some left invariant Riemannian distance on G and let z_ν, $1 \leq \nu \leq M$, be some maximal ε-net in $D_t(s)$ (for $\varepsilon > 0$ appropriately small but fixed), *i.e.*, we choose

$$z_\nu \in D_t(s)\,, \ (1 \leq \nu \leq M)\,, \qquad d(z_\nu, z_\mu) \geq \varepsilon\,, \ \nu \neq \mu\,,$$

and the set $z_1, \ldots, z_M$ is maximal under the above two conditions. It is clear that (2.14) is then verified. What remains is to give a bound of M. To achieve this it suffices to observe that

$$\bigcup_\nu z_\nu\, D_0(\varepsilon_1) \subset D_t(s) \cdot D_0(\varepsilon_1) \subset D_t\big(C\,(s + t + 1)^C\big)\,, \tag{2.16}$$

where the union in (2.16) is disjoined if $\varepsilon_1 > 0$ is appropriately small. The volume estimate (2.11) of the right hand side of (2.16) (we have $\beta_j = 0$ now) gives then the required bound for M.

2.3. General nilpotent algebras.

All our previous notations will be preserved but here $\mathfrak{g}$ will not be assumed to be either a positive or a negative algebra. What will be proved in this section is that for all $s > 0$, $N = 1, 2, \ldots$, we can find

$$(2.17) \qquad z_\nu \in G\,, \quad \nu = 1,2,\ldots M \le C(s+N+1)^C \exp\Big(N \sum \alpha_j\Big)$$

such that

$$(2.18) \qquad D_0(s)\cdots D_N(s) \subset \bigcup_\nu D_0(1)\, z_\nu\,.$$

The consequence that we shall draw from this is that

$$(2.19) \qquad m_G(D_0(s)\cdots D_N(s)) \le C(s+N+1)^C\; exp\Big(N\sum \alpha_j\Big)\,.$$

At this point the reader is *strongly* encouraged to give for himself a proof of (2.19) when $G = \mathbb{R}^d$. A direct proof of (2.19) is highly not trivial even for the Heisenberg group with the dilation structure $X \longrightarrow \lambda X$, $Y \longrightarrow \lambda^{-1}Y$, $Z \longrightarrow Z$. ($X, Y, Z = [X,Y]$ is here the standard basis of the corresponding algebra).

We shall consider the chain of subalgebras (2.6) and we shall prove (2.18) by induction on the k of (2.6). For $k = 0$ (2.18) is contained in the Lemma of Section 2.2.

We shall use of the notations of sections 2.1 and 2.2. The norm in $\mathfrak{g}$ will be chosen so that

$$A_j(s) = \{x \in \mathfrak{h}_j :\ |x| \le s\}$$

satisfy

$$(2.20) \qquad A_{j+1}(s) = A_j(s) + \theta\, y_j + \theta\, z_j\,.$$

The above notation is abusive: the two θ's are *not* identical. They are θ_1 and θ_2 and the right hand side of (2.20) is to be understood as the union over all the $|\theta_i| \le s$, $i = 1, 2$. This kind of abusive but convenient notation will be adopted throughout the rest of the proof. We shall also denote by H_j the subgroup that corresponds to $\mathfrak{h}_j$. The inductive hypothesis is that for some $j \ge 1$ in the algebra $\mathfrak{h}_j$ we have

$$(L_j) \qquad \operatorname{Exp}(A_j(s))\operatorname{Exp}(\Lambda A_j(s))\cdots\operatorname{Exp}(\Lambda^N A_j(s)) \subset \bigcup_\nu D_0(1)\, z_\nu\,.$$

$D_0(1) \subset H_j$ of course here corresponds to $\mathfrak{h}_j$. Using (L_j) we shall procede to prove (L_{j+1}) in the algebra $\mathfrak{h}_{j+1}$.

The first step is to apply Hörmander's lemma (*cf.* Section 2.1) on each factor

$$\begin{aligned}\operatorname{Exp}(\Lambda^m A_{j+1}(s)) &= \operatorname{Exp}(\Lambda^m A_j(s) + \theta\,\Lambda^m y_j + \theta\,\Lambda^m z_j) \\ &= \operatorname{Exp}(e_m) \prod_\alpha \operatorname{Exp}(\Lambda^m Z_\alpha) \\ &= \operatorname{Exp}(e_m)\, M_m = E_m\, M_m\,,\end{aligned}$$

where

$$e_m = c\,\theta\,\Lambda^m y_j + c\,\theta\,\Lambda^m z_j\,, \qquad Z_\alpha = \big(c_\alpha\,\theta\, ad(y_j) + c_\alpha\,\theta\, ad(z_j)\big)^{k_\alpha} A_j(s)\,.$$

Here we make the same abuse of notation over the θ's, and to simplify notations we have droped the j's. The left hand side of (L_{j+1}) is therefore

$$\prod_{m=1}^{N} E_m\, M_m\,. \tag{2.21}$$

The next step is to "commute backwards" all the E_k's through the M_p's that precede it ($i.e., p < k$) so as to put (2.21) in the form

$$E_1\, E_2 \cdots E_N\, \tilde{M}_1 \cdots \tilde{M}_N\,,$$

where

$$\begin{aligned}\tilde{M}_p &\subset Ad(E_N) \cdots Ad(E_{p+1})\,\Lambda^p B_j(c(s+1)) \\ &= \Lambda^p(Ad(\Lambda^{-p} E_N\, \Lambda^{-p} E_{N-1} \cdots \Lambda^{-p} E_{p+1}) B_j(c(s+1)))\,,\end{aligned}$$

$$B_j(s) = \{h \in H_j :\ d_j(e,h) = |h|_j \le s\}\,,$$

and where d_j is the distance on H_j.

The fact that $y_j, z_j \in \mathfrak{g}_-$ implies (*cf.* (2.2), (2.3)) that $y_j, z_j \in (\mathfrak{h}_{j+1})^0_-$ and that $|e_m|_{j+1} \le C\,(s+N)^C$ and, more generally, that

$$|\Lambda^{-p} e_k|_{j+1} \le C\,(s+N)^C\,, \qquad k \ge p\,,$$

$$|\Lambda^{-p} E_N \cdots \Lambda^{-p} E_{p+1}|_{j+1} \le C\,(s+N)^C\,, \qquad p \ge 0\,.$$

We obtain therefore that

$$
\begin{aligned}
&E_1 \cdot E_2 \cdots E_N \subset \operatorname{Exp}(A_{j+1}(C\,(s+N)^C)), \\
(2.22)\qquad &\tilde{M}_p \subset \Lambda^p B_j(C\,(s+N)^C), \\
&\tilde{M}_1 \cdot \tilde{M}_2 \cdots \tilde{M}_N \subset \operatorname{Exp}(A_j(s')) \cdots \operatorname{Exp}(\Lambda^N A_j(s')),
\end{aligned}
$$

with $s' = C\,(s+N)^C$. If we use (L_j) on right hand side of (2.22) we conclude therefore that the right hand side of of (L_{j+1}) is contained in

$$
(2.23)\qquad \bigcup_\nu \operatorname{Exp}(A_{j+1}(C\,(s+N)^C))\, D_0(1)\, z_\nu\,.
$$

An obvious use of the Baker-Campbell-Hausdorff formula gives that (2.23) is contained in

$$
\bigcup_\nu \operatorname{Exp}(A_{j+1}(C\,(s+N)^C))\, z_\nu\,.
$$

To complete the inductive step it suffices therefore to observe that for obvious reasons (*cf.*, proof of the Lemma in Section 2.2) we have

$$
\operatorname{Exp}(A_{j+1}(C\,(s+N)^C)) \subset \bigcup_{\mu=1}^{C(s+N)^C} D_0(1)\, u_\mu
$$

for an appropiate choice of $u_\mu \in H_{j+1}$.

A simple use of the involution $g \longrightarrow g^{-1}$ in G shows that from (2.18) we have the symmetric result

$$
(2.24)\qquad D_N(s)\, D_{N-1}(s) \cdots D_0(s) \subset \bigcup_\nu z_\nu^{-1}\, D_0(1)\,.
$$

It is worth observing also that the left hand side of (2.24) is

$$
\Lambda^N(\tilde{D}_0(s) \cdots \tilde{D}_N(s)),
$$

where $\tilde{D}_t(s) = \operatorname{Exp}(\tilde{\Lambda}^t A(s))$ with $\tilde{\Lambda}^t = \Lambda^{-t}$. The effect of replacing Λ by $\tilde{\Lambda}$ is to swap the positive roots with the negative ones. Using the above observations one can obtain several variants of (2.18) that are relevant in different contexts.

3. The lower estimate.

3.1. Algebraic groups.

In this section I shall follow very closely [4], including the notations. The proof in [4] simplifies considerably if I make the additional assumption that G (*cf.* Section 0) has the form

$$G = Q\lambda M = N\lambda(V \times M), \qquad Q = N\lambda V, \tag{3.1}$$

where $V \cong \mathbb{R}^a$. This assumption is in particular verified for all algebraic groups (or more generally when $\mathfrak{g}$ the Lie algebra is algebraic). Indeed in that case I shall take $\Sigma = V$ for the section constructed in [4] and many of the geometric and algebraic difficulties disappear in one stroke. For the convenience of the reader I shall here first give the proof under the additional assumption (3.1) and then proceed to consider the general case.

All the other notations of [4, Section 3] are preserved:

$$Z_n = \gamma_1 \gamma_2 \cdots \gamma_n = \dot{Z}_n \Lambda_n, \qquad \dot{Z}_n \in S, \ \Lambda_n \in N,$$

is the random walk on G controlled by $d\mu_1 = \phi_1 \, d^r g$ (where however in [4] the above product $Z_s = \dot{Z}_s \Lambda_s$ was written in the other way round). We shall now fix $x \in V$, $s = 1, 2, \ldots$ and we shall find $A \subset \Omega$ a subset of the path space of this random walk on which the following conditions are verified

$$|\dot{Z}_s - x| \leq 1 \tag{3.2}$$

$$L_k\Big(\dot{Z}_j - \frac{j}{s}x\Big) \leq D, \quad j = 1, 2, \ldots s, \ k = 1, 2, \ldots, n, \tag{3.3}$$

$$|\gamma_j| \leq \delta\Big(cD + \frac{|x|^2}{4s}\Big)^{1/2} = \delta\lambda, \quad j = 1, 2, \ldots, s. \tag{3.4}$$

The $D \sim s^{1/3}$ and the δ will be chosen later. I will show that we can choose A so that

$$\mathbf{P}(A) \geq c \exp\Big(-c\,s^{1/3} - \frac{|x|^2}{4s}\Big). \tag{3.5}$$

Indeed let $\Omega_D \subset \Omega$ be the subset of the path space determined by (3.2) and (3.3) then since M is compact, for $D = c\,s^{1/3}$, we have (*cf.* Appendix):

$$\mathbf{P}(\Omega_D) \geq \exp\Big(-c\,s^{1/3} - \frac{|x|^2}{4s}\Big) = \exp(-\lambda^2).$$

Let $\Omega^\lambda \subset \Omega$ be the subset determined by (3.4) then

$$\mathrm{P}(\Omega^\lambda) \geq 1 - s\, e^{-C\delta^2\lambda^2}$$

because the variables $\gamma_j \in G$ satisfy a Gaussian estimate on G (*cf.* [11]). We shall fix $D \sim s^{1/3}$ and δ large enough. We have then

$$\mathrm{P}(\Omega_D \cap \Omega^\lambda) \geq C \exp(-\lambda^2)$$

and our assertion follows.

Let us now go back to [4] where we shall assume that $G = N\lambda(V \times M)$ and that $\Sigma = V$, $S = V \times M$. For any subset $E \subset N$ we shall denote $E^g = gEg^{-1}$ $(g \in G)$, we shall further denote B_α the ball in N of radius $\exp(C\alpha)$ centered at the identity. Analysing closely the argument in [4] it follows that on our subset $A \subset \Omega$ (that satisfies (3.2), (3.3), (3.4) and (3.5)) we have

$$Z_s \in (B_{\delta\lambda}^{\dot{Z}_1} B_{\delta\lambda}^{\dot{Z}_2} \cdots B_{\delta\lambda}^{\dot{Z}_s}) \dot{Z}_s\,. \tag{3.6}$$

On the other hand, from (3.3) it is clear that

$$B_{\delta\lambda}^{\dot{Z}_j} \subset (B_{\delta\lambda+D})^{jx/s}\,.$$

It follows therefore that

$$Z_s \in (B_{\delta\lambda+D}^{x/s} \cdots B_{\delta\lambda+D}^{jx/s} \cdots)\, \dot{Z}_s\,, \tag{3.7}$$

where because of (2.19) the m_N-measure of $(\cdots)$ is bounded above by

$$s^c \exp\Big(C\,(\delta\,\lambda + D) + \sum L_j^+(x)\Big)\,.$$

The estimates (0.11), (0.12) follow.

The only modification needed to obtain the improvement under the (NC)-condition is that we set $D \sim c \log s$ instead (*cf.* Appendix). The rest of the proof is identical.

3.2. The general case.

A thorough understanding of the geometric construction in [4, Section 1] is essential for this section. The notations here are those of the previous section and of [4] and we shall start the proof exactly as in the previous section. The point where we run into trouble is (3.6), (3.7). Indeed the section Σ is in general not a group and the elements of Σ do not commute between themselves.

The first step towards resolving these difficulties is to choose the generators $e_1, e_2, \dots, e_m$ of the nilpotent algebra $\mathfrak{a}$ (*cf.* proof [4, Lemma 1.2]) so that our preasigned point $x \in V$ (on which we want to prove the lower estimates) is $x = \pi(\exp(|x|e_1))$, *i.e.*, x lies in the image by $\pi : Q \longrightarrow V$ of the first one parameter subgroup $x_1(t)$ of Σ. With the obvious identification of Σ and V and with the basis $d\pi(e_j)$, $j = 1, 2, \dots, m$, on V we have then $x = (|x|, 0, 0, \dots, 0)$.

Let us assume for simplicity that $G = Q$ is soluble, *i.e.*, that $M = \{e\}$. We have then as in [4, (3.5)]

$$\gamma_j = \dot\gamma_j n_j, \qquad \dot\gamma_j = \sigma(t_j) = x_1(t_j^{(1)}) \cdots x_m(t_j^{(m)}) \in \Sigma,$$
$$t_j = (t_j^{(1)}, \dots, t_j^{(m)}) \in V, \qquad j = 1, 2, \dots, s,$$

and because of (3.2), (3.3), (3.4) we have

$$|t_1 + t_2 + \cdots + t_s - x| \leq 1,$$
$$L_k(t_1 + t_2 + \cdots + t_j - jx/s) \leq D,$$
$$|t_j^{(i)}| \leq \delta\Big(D + \frac{|x|^2}{4s}\Big) = \delta\,\lambda,$$
$$j = 1, 2, \dots, s, \qquad k = 1, 2, \dots, n, \qquad i = 1, 2, \dots, m.$$

The critical step is to prove that

$$\begin{aligned} Z_s &= \dot\gamma_1 n_1 \dot\gamma_2 n_2 \cdots \dot\gamma_s n_s \\ &\in x_1(t_1^{(1)})\, B\, x_1(t_2^{(1)})\, B \cdots \\ &\quad \cdot x_1(t_s^{(1)})\, B\, x_2(t_1^{(2)} + \cdots + t_s^{(2)}) \cdots x_m(t_1^{(m)} + \cdots + t_s^{(m)}), \end{aligned} \tag{3.8}$$

where $B \subset N$ is the ball of radius

$$\exp(C\,\delta\,\lambda + c\,D)\, s^C\, |x|^C \leq s^C \Big(1 + \frac{|x|^2}{4s}\Big)^C \exp(C\,\delta\,\lambda + C\,D). \tag{3.9}$$

To see this we must shift each $x_2(t_j^{(2)})\cdots x_m(t_j^{(m)})$ (*i.e.*, the co-factors of $x_1(t_j^{(1)})$ in $\dot\gamma_j$) through to product. Every time we cross a ball $|n_k|_N \le \exp(C\,\delta\,\lambda)$ by

$$x_2\big(t_1^{(2)}+\cdots+t_k^{(2)}\big)\,x_3\big(t_1^{(3)}+\cdots+t_k^{(3)}\big)\cdots=\zeta_k\,. \tag{3.10}$$

We multiply the radius of the ball by $\exp(CD)$ hence the exponential term in the left hand side of (3.9). The extra factor $s^C\,|x|^C$ in (3.9) is the bound of the number of commutators of terms of the form $x_j(t)$ ($t \in \mathbb{R}$, $|t| \le 1$, $j = 1,2,\ldots,n$) (*cf.*, Nil-Gp Lemma in [4]). Each of these commutators lie in some fixed ball in N (for the $|\cdot|_N$ distance). The above commutators are placed between the $x_1(t_j^{(1)})$'s ($j = 1,2,\ldots,s$) and the B's and they arise because we have to put the factors $x_i(t_j^{(i)})$ in the "right order" to make a term ζ_k as in (3.10).

We are now in a position to finish the proof. Indeed one more set of commutations, with the $x_1(t_j^{(1)})$'s this time, brings the right hand side of (3.8) in the form

$$\big(B^{x/s}_{\delta\lambda+D}\cdots B^{jx/s}_{\delta\lambda+D}\cdots\big)\,\sigma(t_1+\cdots+t_s)\,.$$

This expression is essentially the same as (3.7) and the proof finishes as before.

Finally since the elements of Σ commute with M if we do not assume that $M=\{e\}$ nothing changes in the above argument (*cf.* [4, Section 1]).

4. The Hardy-Littlewood theory.

Theorem 1 is an obvious integration of the estimates (0.9), (0.13) on V which also proves the first estimate (4.1) below. The reader should observe that the change of variable $x \mapsto x^{-1}$ in G shifts left to the right measure and therefore the "left" and "right" $\|\cdot\|$-normes of ψ_t are the same. What is important in Theorem 1 is to understand the exact role played by Δ on G and by the induced scalar product $\langle\cdot\,,\,\cdot\rangle_\Delta$ on V (*cf.* Section 0.1).

The expression of $\ell(q)$ given by theorem does of course depend on that scalar product and therefore on the particular choice of the Laplacian Δ on G. One of the consequences of Theorem 2 is that L is a *linear invariant*, *i.e.*, that it is independent of the particular scalar product used in (0.2). This is very easy to see. Indeed let

$$L(p;x) = \Big(\frac{1}{p} - \frac{1}{2}\Big) \sum L_j^+(x) + \frac{1}{2} \sum L_j^-(x), \quad 1 \le p \le +\infty, \; x \in V.$$

We clearly have

$$L(1;x) = \frac{1}{2} \sum L_j(x), \qquad L(2;x) = \frac{1}{2} \sum L_j^-(x).$$

We also have

$$L = \inf\{1 \le p \le +\infty : \; L(p;x) \le 0 \text{ for all } x \in V\}.$$

Indeed assume that there exists $x_0 \in V$ such that $\lambda_0 = L(p, x_0) > 0$; by the homogeneity of degree 1 of $L(p, \cdot\,)$ it follows that

$$L(p,x) \ge C\,|x|\,\lambda_0\,; \qquad x \in V,\; |x - |x|\,x_0| \le 1,$$

provided that $|x|$ is large enough. And if we integrate the "shifted" Gaussian inside the tube $\{x : \; |x - |x|\,x_0| \le 1\}$ we see that $\ell(p) > 0$. Conversely of course $\ell(p) = 0$ if $L(p;x) \le 0$ $(x \in V)$.

Parts i) and ii) of Theorem 2 follow at once from this, we also have

$$(4.1) \qquad \begin{aligned} &\|\tilde{T}_t\|_{1 \to p} = \|\psi_t\|_p = O(1), \quad \text{as } t \to +\infty,\; L \le p \le +\infty. \\ &\|\tilde{T}_t\|_{r \to 2} = O(1), \quad \text{as } t \to +\infty,\; 1 \le r \le 2. \end{aligned}$$

For $r = 2$ the second estimate is the definition of amenability; for $r = 1$ it is very easy and has been verified in [10]; we interpolate to obtain the other values of r.

If we assume that $L < 2$ and interpolate between $\|\cdot\|_{r \to 2}$ and $\|\cdot\|_{1 - L}$ we obtain iii).

Part iv) of theorem is a trifle more subtle and it relies on the a priori knowledge of

$$(4.2) \qquad \|\tilde{T}_t\|_{p \to p} = \exp(\rho^2\,(1/2 - 1/p)^2\,t), \qquad 1 \le p \le +\infty,$$

where ρ^2 is defined by $\Delta m = -\rho^2 m$. (4.2) is an immediate consequence of amenability and of the formula (*cf.* [10])

$$\tilde{T}_t f = (f\,m^{-1/p} * \mu_t\, m^{1/2-1/p})\, m^{1/p}.$$

On the other hand we clearly have by the semigroup property:

(4.3) $\|\tilde{T}_{t+2}\|_{1\to p} \le C_2\,\|\tilde{T}_{t+1}\|_{r\to p} \le C_1\,\|\tilde{T}_t\|_{p\to p}\,, \quad 1 \le r \le p \le +\infty\,.$

The only thing that has to be done to complete the proof of iv) is to show that under the WNC-condition, up to negligible error, we have

$$\exp(\rho^2\,(1/2 - 1/p)^2\,t) \le \|\psi_t\|_p\,.$$

To see this let us choose $x_0 \in V$ so that

$$L_j(x) \ge 0\,; \qquad j = 1, 2, \dots, n\,, \ \ \langle x, x_0\rangle \ge 0\,,$$

which is possible by the WNC-condition. We clearly have

$$L(p,x) = \Big(\frac{1}{p} - \frac{1}{2}\Big)\sum L_j(x)\,, \qquad \langle x, x_0\rangle \ge 0\,,$$
$$L(p,x) = \frac{1}{2}\sum L_j(x)\,, \qquad \langle x, x_0\rangle \le 0\,.$$

It follows therefore that up to negligible error

(4.4)
$$\begin{aligned}\int \exp\Big(L(p,x) - \frac{|x|^2}{4t}\Big)\,dx \\ = \int \exp\Big(\Big(\frac{1}{p} - \frac{1}{2}\Big)\sum L_j(x) - \frac{|x|^2}{4t}\Big)\,dx \\ + \text{negligible error.}\end{aligned}$$

A moment reflexion gives on the other hand

$$\Big|\sum L_j\Big|_{V^*} = \rho$$

for the dual norm in V^*. The right hand side of (4.4) can therefore be explicitely computed, this together with our theorem proves our assertion and it gives iv).

To prove v) observe that for every fixed q the function

$$\ell(p,q) = \limsup_{t\to+\infty} \frac{1}{t}\,\log\|\tilde{T}_t\|_{p\to q}$$

is, as a function of p, non decreasing (*cf.* (4.3)) and convex in $1/p$ (by Riesz-Thorin). Under the condition (0.15) of v) it follows therefore that

$$\ell(q) = \ell(1,q) = \left(\frac{1}{2} - \frac{1}{q}\right)^2 \rho^2$$

for some $1 < q < 2$. To finish the proof of v) we shall insert in (0.14) the above value of $\ell(q)$ and prove that this implies the WNC-condition on the L_i's. Towards that observe that we can assume that the L_i's span the space V^* for otherwise we can quotient out $\cap Ker L_i$ and this reduces the integral in (0.14) to a lower dimensional Gaussian. Then (*cf.* the remark that follows the definition of Section 0.4) since

$$L(p,x) = \left(\frac{1}{p} - \frac{1}{2}\right) \sum L_j(x) + \left(1 - \frac{1}{p}\right) \sum L_j^-(x)\,,$$

if G is not WNC, we must have

$$L(p,x) \leq \left(\frac{1}{p} - \frac{1}{2}\right) \sum L_j(x) - \varepsilon\,, \qquad |x| = 1\,,$$

for some $\varepsilon > 0$. This will immediatly give us a contradiction and will complete the proof of v).

The corollaries 1 and 2 are easy to prove. Indeed

$$\left\| \int_t^{t+1} \tilde{T}_s \psi_1 \, ds \right\|_p \geq C \, \|\psi_t\|_p \,, \qquad t > 1\,,$$

by the local Harnack principle. Corollary 1 follows.

If G is as in Corollary 2 we have (*cf.* [8], [10])

$$\|\tilde{T}_t\|_{1\to+\infty}\,,\ \|\tilde{T}_t\|_{1\to 2} = O(e^{-ct^{1/3}})\,, \qquad \|\tilde{T}_t\|_{2\to 2} = O(1)\,,$$

as $t \to +\infty$ for some $c > 0$. By interpolation it follows that

$$\|\tilde{T}_t\|_{p\to q} = O(e^{-ct^{1/3}})\,, \qquad 1 \leq p \leq 2 \leq q \leq +\infty\,,\ p < q\,.$$

If we interpolate this estimate with (4.1) we see that for all $1 \leq p < 2$ there exists $L(p) < 2$ such that

$$\|\tilde{T}_t\|_{p\to q} = O(e^{-ct^{1/3}})\,, \qquad \text{as } t \to +\infty\,,\ L(p) < q \leq +\infty\,.$$

This estimate takes care of the convergence of

$$\int_1^{+\infty} t^{\alpha/2-1}\,\tilde{T}_t\,dt$$

the convergence of $\int_0^1$ is taken care off by standard methods (*cf.* [11]). Corollary 2 follows.

5. Non simply connected groups.

The key to the analysis of non simply connected groups is the following observation. Let G be some connected Lie group and let $K \subset G$ be some compact normal subgroup. We can then use the projected sublaplacian $d\pi(\Delta) = \dot{\Delta}$ by the projection $\pi : G \longrightarrow G/K$ to define T_t, $\tilde{T}_t$, ϕ_t, ψ_t both on G and on G/K. Let $\dot{\psi}_t$ be the corresponding kernel on G/K. It is then clear (*cf.* [8]) that

$$\dot{\psi}_t(x) = \int_K \psi_t(xk)\,dk = \int_K \psi_t(kx)\,dk\,.$$

By an easy application of the local Harnack estimates it follows therefore that

$$\|\dot{\psi}_t\|_{L^p(G/K)} \approx \|\psi_t\|_{L^p(G)}\,, \qquad t \geq 1\,,\ 1 \leq p \leq +\infty\,.$$

This means that, at least the Hardy-Littlewood estimates and the results of Section 0.6 pass through a quotient by some compact subgroup.

To go further we shall have to introduce some notations. We shall say that some soluble connected Lie group Q is *admissible* soluble if there exists some simply connected group $\tilde{Q}$ and some covering map $\theta : \tilde{Q} \longrightarrow Q$ such that $\operatorname{Ker}\theta \cap \tilde{N} = \{e\}$ where $\tilde{N} \subset \tilde{Q}$ is the nilradical of $\tilde{Q}$. The fact that $\theta(\tilde{N}) = N \subset Q$ is the nilradical of Q (*i.e.*, a closed subgroup) implies that

$$\operatorname{dist}(\tilde{N}, \operatorname{Ker}\theta \setminus \{e\}) > 0\,,$$

$$Q/N \cong V \times T = \tilde{Q}/N \cdot \operatorname{Ker}\theta\,,$$

$$V \approx \mathbb{R}^a\,, \qquad T \approx \mathbb{T}^b = (\mathbb{R} \bmod 2\pi)^b\,.$$

We shall say that G some connected Lie group is *admissible* if $G \approx Q\lambda M$ where Q is admissible soluble and M is compact.

Let G now be some amenable connected Lie group. By standard global structure theorems (*cf.* [1], [7], [14]) we see that we can "cover" G by $\theta : Q\lambda M \longrightarrow G$ with $\operatorname{Ker}\theta$ finite (*i.e.*, an "isogeny"). Inside the center of the nilradical of Q on the other hand we can find some compact subgroup $K(\approx \mathbb{T}^b)$ such that G/K is admissible.

The consequence of the above structure theorems and of our previous considerations is that if we can extend our Hardy-Littlewood theory of Section 0.6 from simply connected groups to admissible groups then we automatically have it for all amenable groups. For these theorems it should be observed that the non zero real parts of the roots of G and of the corresponding admissible group (as constructed above) are up to obvious identification the same.

When G is an admissible group then its nilradical $N \subset G$ is simply connected and $G/N \cong V \times K$ where $K = T \times M$ so that the only difference with the simply connected case is the presence of the torus T as a cofactor of M.

By going through the proofs of both the upper and the lower estimates of Section 0 we see therefore that strictly nothing changes neither in the proofs nor in the statements of the results.

The conclusion is in particular that theorems 1 and 2 and their corollaries are valid as stated, for general, not necessarily simply connected, groups.

APPENDIX: The brownian bridge.

Let us recall here some facts on the Brownian bridge (*cf.* [15], [16]) *i.e.*, brownian motion $b(t) \in \mathbb{R}^a = V$ conditioned by

$$b(0) = x\,, \qquad b(t) = y\,.$$

We shall denote this process by $b^t_{x,y}(s)$, $0 \leq s \leq t$, and recall the following well known facts ($\simeq$ denotes equidistributed processes)

i) $b^t_{x,y}(s) \simeq x + s(y-x)/t + b^t_{0,0}(s)$,

ii) $b^t_{0,0}(s) \simeq b(s) - s\,b(t)/t$,

iii) $b^{\lambda t}_{\sqrt{\lambda}x,\sqrt{\lambda}y}(\lambda s) \simeq \sqrt{\lambda}\, b^t_{x,y}(s)$,

iv) $b^t_{0,0}(s) \simeq (1-s/t)\, b(st/(t-s))$,

v) $b^t_{x,y}(s) \simeq b^t_{y,x}(t-s)$, (*cf.* [15], [16] together with the scalling properties of standard brownian motion).

vi) The standard Markov property implies that with respect to the conditional probability $\mathbf{P}[\cdot \mid b^t_{x,y}(t/2)]$ the two "halfs" of the brownian bridge:

$$\{b^t_{x,y}(s):\ 0 < s < t/2], \qquad \{b^t_{x,y}(s):\ t/2 < s < t\},$$

are independents.

We shall now denote by

$$P(D) = \mathbf{P}(L_j(b^t_{0,0}(s)) \leq D:\ 0 < s < t,\ j = 1, 2, \dots, n),$$

where $L_1, \dots, L_n \in V^*$ are linear functionals of V. We shall say that the L_j's satisfy the NC-condition if there exists $x \in V$ such that $L_j(x) > 0$, $j = 1, 2, \dots, n$ (*cf.* Section 0.4). In fact in what follows the L_js will be the real parts of the roots as defined in Section 0. We shall prove the following

Lemma 1. *There exist constants $C, c > 0$ independent of t and D such that:*

$$P(D) \geq C\, e^{-ct/D^2}, \qquad t,\ D > 0. \tag{A.1}$$

If we suppose in addition that the $L_j \in V^*$, $j = 1, 2, \dots,$ satisfy the NC-condition then

$$P(D) \geq C \left(\frac{t}{D^2}\right)^{-c}. \tag{A.2}$$

The estimate (A.1) is a consequence of the following

$$\mathbf{P}(|b^t_{0,0}(s)| \leq D;\ 0 < s < t) \geq C\, e^{-ct/D^2}, \tag{A.3}$$

but (A.3) immediately reduces because of ii) or iv) to the corresponding estimate for brownian motion which is well known (see *e.g.* [7]).

The estimate (A.2) is more subtle to prove. We use v) and vi) to see that

$$P(D) = \int \left(\mathbf{P}(L_j(b^t_{0,0}(s)) \leq D;\ 1 \leq j \leq n,\ 0 < s < t/2 \mid b^t_{0,0}(t/2) = z)\right)^2 \cdot \mathbf{P}(b^t_{0,0}(t/2) \in dz)$$

$$\geq \Big(\int (\cdots)\, \mathbf{P}(b^t_{0,0}(t/2) \in dz) \Big)^2$$
$$= \big(\mathbf{P}(L_j(b^t_{0,0}(s)) \leq D\,;\ 1 \leq j \leq n\,;\ 0 \leq s \leq t/2) \big)^2 .$$

But because of iv) the right hand sideof the above can be replaced by the coresponding expression with $b^t_{0,0}(\cdot)$ replaced by $b(\cdot)$. The corresponding lower estimate is easy to obtain (*cf.* [7]).

References.

[1] Varadarajan, V. S., *Lie groups, Lie algebras and their representations.* Prentice-Hall, 1984.

[2] Hörmander, L., Hypoelliptic second order equations. *Acta Math.* **119** (1967), 147-171.

[3] Chevalley, C., *Théorie des groupes de Lie*, tomes II & III. Hermann 1955.

[4] Varopoulos, N. Th., Diffusion on Lie groups (II). *Can. J. Math.* **46**, (1994), 1073-1093.

[5] Jacobson, *Lie algebras.* Interscience, 1962.

[6] Bourbaki, N., *Eléments de mathématiques, Livre* VI, *Intégration.* Hermann, 1963.

[7] Varopoulos, N. Th., Analysis on Lie group, preprint.

[8] Varopoulos, N. Th., Diffusion on Lie groups (I). *Can. J. Math.* **46** (1994), 438-448.

[9] Alexopoulos, G., An application of homogenization theory to harmonic analysis: Harnack inequalities and Riesz transforms on Lie groups of polynomial growth. *Can. J. Math.* **44** (1992), 691-727.

[10] Varopoulos, N. Th., Thèorie de Hardy-Littlewood sur les groupes de Lie. *C. R. Acad. Sci. Paris, Serie I* **316** (1993), 999-1003.

[11] Varopoulos, N. Th., Saloff-Coste, L., Coulhon, Th., *Analysis and geometry on groups.* Cambridge Univ. Press, 1992.

[12] Lohoué, N., Estimées de type Hardy pour l' opérateur $\Lambda + \lambda$ d' un espace symétrique non compact.*C. R. Acad. Sci. Paris.* **308** (1989), 11-14.

[13] Lohoué, N., Inégalités de Sobolev pour les sous Laplaciens de certains groups unimodulaires. *Geom. Funct. Anal.* **4** (1992), 394-420.

[14] Varopoulos, N. Th., Hardy-Littlewood theory on unimodular groups, to appear in *Ann. Inst. H. Poincaré.*

[15] Knight, F. B., Essentials of Brownian motion and diffusion. *Math. Survey* **18**. Amer. Math. Soc., 1981.

[16] Ikeda, N., Watanabe, S., *Stochastic differential equations and diffusion processes.* North-Holland **24**, 1989.
[17] Varopoulos, N. Th., Mustapha, S., Lecture notes Université Paris VI.

Recibido: 15 de febrero de 1.995

N. Th. Varopoulos
I.U.F. and
Departement de Mathématiques
Université de Paris VI
75005 Paris, FRANCE

Revista Matemática Iberoamericana
Vol. 12, N.° 1, 1996

Good metric spaces without good parameterizations

Stephen Semmes

Abstract. A classical problem in geometric topology is to recognize when a topological space is a topological manifold. This paper addresses the question of when a metric space admits a quasisymmetric parameterization by providing counterexamples to the obvious optimistic conjectures, or, in other words, by providing examples of spaces with many Euclidean-like properties which are nonetheless substantially different from Euclidean geometry. These examples are geometrically self-similar versions of classical topologically self-similar examples from geometric topology, and they can be realized as codimension 1 subsets of Euclidean spaces. Unlike earlier examples going back to Rickman, these sets enjoy good bounds on their geodesic distance functions and good mass bounds (Ahlfors regularity). They are also smooth except for reasonably tame degenerations near small sets, they are uniformly rectifiable, and they have good properties in terms of analysis (like Sobolev and Poincaré inequalities). The construction also produces uniform domains which have many nice properties but which are not quasiconformally equivalent to balls.

1. Introduction.

How can one recognize when a metric space admits a quasisymmetric, bilipschitz, or homeomorphic parameterization by a Euclidean space?

For the purposes of this paper it will be sufficient to consider only subsets of Euclidean spaces instead of abstract metric spaces, and so we restrict the generality of our definitions accordingly. (Note however the results of Assouad [A1], [A2], [A3] on embedding metric spaces nicely into Euclidean spaces, and quasisymmetrically in particular.) A mapping $f : \mathbb{R}^d \longrightarrow \mathbb{R}^n$ is said to be *quasisymmetric* (or a *quasisymmetric embedding*) if it is not constant and if there exists a homeomorphism $\eta : [0,\infty) \longrightarrow [0,\infty)$ such that

$$(1.1) \quad |x-a| \le t\,|x-b| \quad \text{implies} \quad |f(x)-f(a)| \le \eta(t)\,|f(x)-f(b)|$$

for all $t > 0$ and $x, a, b \in \mathbb{R}^d$. This condition means that f distorts relative distances in a bounded way, *e.g.*, if a is much closer to x than b is, then the analogous statement is true for $f(a), f(x)$, and $f(b)$. However, it does not prevent $|f(x)-f(a)|$ from being wildly different from $|x-a|$. The bilipschitz condition requires the stronger property that f distort absolute distances by only a bounded amount, *i.e.*,

$$(1.2) \qquad C^{-1}\,|x-y| \le |f(x)-f(y)| \le C\,|x-y|$$

for some C and all $x, y \in \mathbb{R}^d$.

SIMPLE EXAMPLES. The map $x \mapsto x\,|x|^{a-1}$ on $\mathbb{R}^d$ is quasisymmetric whenever $a > 0$, but it is bilipschitz only when $a = 1$. The mapping $x \mapsto x\,\exp(|x|)$ is a homeomorphism on $\mathbb{R}^d$ but is not quasisymmetric, because quasisymmetric maps cannot grow faster than polynomially at infinity. The mapping on $\mathbb{R}$ given by $x \mapsto |x|^a$ when $x \ge 0$ and $x \mapsto -|x|^b$ when $x \le 0$ is a homeomorphism for all $a, b > 0$, but it is quasisymmetric only when $a = b$.

One can make more amusing examples by introducing some spiralling. Let $\{\theta_t\}$, $t \in \mathbb{R}$, be a one-parameter family of rotations on $\mathbb{R}^d$ which is Lipschitz continuous in t. Then $x \mapsto \theta_{\log|x|}(x\,|x|^{a-1})$ is a quasisymmetric map on $\mathbb{R}^d$ when $a > 0$, and it is bilipschitz when $a = 1$. For embeddings of $\mathbb{R}^d$ into $\mathbb{R}^n$ one can also introduce plenty of corners.

To put the above question about the existence of parameterizations into perspective let us recall a wonderful theorem of Edwards ([E], see also [C1], [C2], [C3], [D]) to the effect that there exist finite polyhedra of dimension 5 (say) which are homeomorphic to the standard 5-sphere but not bilipschitz equivalent to it. (This formulation of the "double suspension theorem" uses also the observation made in [SS, p. 504, Remark (b)].) As this result indicates, there is some serious technology in topology for showing that a space admits a homeomorphic parameterization by a Euclidean space even when one might not expect that to be true. (See Section 2 for a little more information.)

Because of the existence of these strange polyhedral spheres and other examples in [Se5] I have come to the conclusion that bilipschitz parameterizations are too limited for understanding the structure of sets with little smoothness but reasonable behavior. In other words, the strange polyhedral spheres are just finite polyhedra, and the examples constructed in [Se5] are also very reasonable in their behavior, and so they should not receive all the blame if the bilipschitz condition is too stingy to include them in its parameterizations. This does not mean that there is no meaningful characterization of sets which admit bilipschitz parameterizations, nor that there are not plenty of interesting criteria for the existence of bilipschitz parameterizations, but simply that these criteria cannot include some otherwise very reasonable sets. So far the only general criteria known seem to be the ones in [SS] (for polyhedra) and [To1], [To2]. (See [Se5] for some open problems.)

What about quasisymmetric parameterizations? All of the examples in [Se5] were quasisymmetrically equivalent to a Euclidean space, and it is not known whether the strange polyhedral spheres of Edwards and Cannon are quasisymmetrically equivalent to standard spheres. The bottom line of this paper (Theorem 1.12) is that there exist spaces with many good properties but which do not admit quasisymmetric parameterizations. The most compelling of these is a variant of an example in [FS] of a discrete group of homeomorphisms which is not topologically conjugate to a uniformly quasiconformal group. Before getting to that let us start from scratch and consider the question of quasisymmetric parameterizations more thoroughly.

Suppose first that $d = 1$. A set $E \subseteq \mathbb{R}^n$ admits a *quasisymmetric parameterization* by $\mathbb{R}$ if and only if it is a Jordan curve and it satisfies the Ahlfors "3-point condition", *i.e.*, there is a constant C so that if $x, y \in E$ and A is the arc which connects them then $\operatorname{diam} A \leq C\,|x - y|$. (See [TV, Section 4].)

For the $d = 2$ case there is a result [Tu2, Lemma 4] to the effect that the product of a nonrectifiable arc with a line segment cannot be embedded quasisymmetrically into $\mathbb{R}^2$. Assertions of this nature were established first for snowflake curves by Rickman, and then general results were obtained by Väisälä and Tukia. (See [Tu2], [V1].) Keep in mind that there are plenty of nonrectifiable Jordan curves which satisfy the Ahlfors 3-point condition, like snowflake curves. This result allows one to build counterexamples to many reasonable conjectures about the existence of quasisymmetric (and quasiconformal) mappings (see [Tu2]), and thereby makes it difficult if not impossible to find reasonable criteria for the existence of a quasisymmetric parameterization when $d = 2$ without imposing some conditions on the mass. However, it turns out that if we do impose such a condition, then there is a nice positive result (Theorem 1.6 below). Before stating it we need a couple of definitions.

Definition 1.3. *A subset E of $\mathbb{R}^n$ is said to satisfy Condition $(*)$ (with dimension d) if it is closed and if there there is a constant C such that for each $x \in E$ and $r > 0$ there is a (relatively) open set $U \subseteq E$ such that $E \cap B(x,r) \subseteq U \subseteq E \cap B(x, Cr)$ and U is homeomorphic to a d-ball.*

Condition $(*)$ is necessary for E to be quasisymmetrically equivalent to $\mathbb{R}^d$. (A quasisymmetric map takes a ball to a set with approximately the same shape as a ball.) It is not sufficient, however, even when $d = 2$, because of the examples described above. (Note that [Tu2] also covers products of unrectifiable curves with higher-dimensional (standard) cells. See [AV] for related results for products of topological cells when both are permitted to have dimension larger than 1.)

When $d = 1$ Condition $(*)$ implies that E is a Jordan curve which satisfies the Ahlfors 3-point condition, and so it does actually imply quasisymmetric equivalence with $\mathbb{R}$. In general Condition $(*)$ tries to capture some of the geometry implied by the existence of a quasisymmetric parameterization. For instance, Condition $(*)$ forbids cusps and long thin tubes, not to mention crossings.

The next condition requires that the given set be well behaved in terms of Hausdorff measure, and it will be used to avoid the above counterexamples.

Definition 1.4. *A subset E of $\mathbb{R}^n$ is said to be (Ahlfors) regular of dimension d if it is closed and if there is a constant $C > 0$ so that*

$$C^{-1}\, r^d \leq H^d(E \cap B(x,r)) \leq C\, r^d \tag{1.5}$$

for all $x \in E$ and $r > 0$. Here H^d denotes d-dimensional Hausdorff measure (and not cohomology).

A d-plane in $\mathbb{R}^n$ is regular with dimension d, and the same is true of any set which is bilipschitz equivalent to a d-plane. In general (Ahlfors) regularity means that E behaves measure-theoretically like a d-plane, but it can still be very different from a d-plane geometrically. For instance, for each $0 < d < n$ there are Cantor sets in $\mathbb{R}^n$ which are regular with dimension d. There are also snowflake curves and tree-like sets which are regular (with dimension larger than 1). Note that regularity is not necessary for quasisymmetric equivalence with $\mathbb{R}^d$. However, when $d = 2$ regularity and Condition $(*)$ together imply the existence of a quasisymmetric parameterization, modulo some a priori smoothness assumptions.

Theorem 1.6. *Suppose that $E \subseteq \mathbb{R}^n$ is regular and satisfies Condition $(*)$, both with dimension 2. Suppose also that E is smooth and well behaved at infinity. (We need to assume enough to ensure that E is conformally equivalent to the plane.) Then E is quasisymmetrically equivalent to $\mathbb{R}^2$, with a choice of η as in (1.1) which depends on the constants from Definitions 1.3 and 1.4 but which does not depend on our a priori smoothness assumptions in a quantitative way.*

Theorem 1.6 was proved in [Se2, Section 5]. (See [DS3, Section 6] for a related result.) The argument in [Se2] went in two steps, as follows: the a priori assumptions on E together with the uniformization theorem were used to obtain the existence of a conformal parameterization of E, and then classical methods were used to show that the geometric assumptions imply that the conformal parameterization is quasisymmetric with uniform bounds. It turns out that the second step can be made to work in great generality; it is proved in [HK] that a quasiconformal parameterization of a metric space which satisfies certain simple geometric properties is actually quasisymmetric. (See [HK] for the precise statement.) The first step -the uniformization theorem- is of course very special to dimension 2, and in fact Theorem 1.6 fails in dimension 3, as we shall see.

Note that if a set E admits a quasisymmetric parameterization f by $\mathbb{R}^d$ and is regular with dimension d then f satisfies the same kind of estimates as in [Ge]. (That is, the pull-back of $H^d|_E$ to $\mathbb{R}^d$ via f is an A_∞ weight, by the same argument as in [Ge]. See [DS1] and [Se3].

This result implies strong restrictions on the way that f can distort distances, and it implies local Sobolev space conditions on f and its inverse.) Such a set E has many of the same properties as it would if it admitted a bilipschitz parameterization. (This would not be true if E were regular but with a different dimension than d.) When $d = 2$ it is not known whether E might actually be bilipschitz equivalent to $\mathbb{R}^2$ under these conditions, but there are counterexamples when $d = 3$, by [Se5].

Let us consider a couple of other geometric conditions on sets of roughly the same spirit as $(*)$. The first is a more uniform version of $(*)$, while the second is weaker and is given in terms of contractability properties.

Definition 1.7. *A subset E of $\mathbb{R}^n$ is said to satisfy Condition $(**)$ (with dimension d) if it is closed and if there there is a constant C and a locally bounded function $\omega : [0,\infty) \longrightarrow [0,\infty)$ with $\lim_{t\to 0} \omega(t) = 0$ such that for each $x \in E$ and $r > 0$ there is a (relatively) open set $U \subseteq E$, $E \cap B(x,r) \subseteq U \subseteq E \cap B(x, Cr)$, and a homeomorphism g from the unit ball B_d in $\mathbb{R}^d$ onto U such that*

$$|g(y) - g(z)| \le r\,\omega(|y - z|)\,, \qquad \text{for all } y, z \in B_d\,, \tag{1.8}$$

and

$$|g^{-1}(v) - g^{-1}(w)| \le \omega(r^{-1}|v - w|)\,, \qquad \text{for all } v, w \in U\,. \tag{1.9}$$

Roughly speaking, Condition $(**)$ does for sets what quasisymmetry does for mappings.

The difference between this and Condition $(*)$ is that we require here a uniform bound on the moduli of continuity of the homeomorphic parameterizations of the topological d-balls U (and also on the moduli of continuity of their inverses). We do not require that this modulus of continuity be anything in particular -*e.g.*, we do not require Hölder continuity- and we have been careful in (1.8) and (1.9) to make the estimates scale-invariant.

Finite polyhedra which are also topological manifolds (without boundary) provide an amusing class of sets which satisfy $(**)$, or rather the obvious counterpart of $(**)$ for compact sets, in which we consider only small radii. (See Section 11.) The results of Edwards and Cannon imply that there are many strange examples of such polyhedra.

Notice that Condition $(**)$ is satisfied if E admits a quasisymmetric parameterization by $\mathbb{R}^d$. The difference between the two properties is basically that the homeomorphisms g in Definition 1.7 are allowed to depend on x, r in a completely arbitrary fashion, while the existence of a quasisymmetric parameterization means that all the g's can be obtained from a global parameterization of E in a certain way.

One can also look at the uniformity required in Condition $(**)$ in terms of compactness. To understand this point it is helpful to recall the following compactness property of quasisymmetric mappings.

Lemma 1.10. *Suppose that* $f_j : \mathbb{R}^3 \longrightarrow \mathbb{R}^4$ *is a sequence of quasisymmetric embeddings which satisfy* (1.1) *with a fixed choice of* η. *Suppose also that there are two points* $a_0, a_1 \in \mathbb{R}^3$ *and a positive constant* C *such that* $|f_j(a_i)| \leq C$ *and* $|f_j(a_0) - f_j(a_1)| \geq C^{-1}$ *for all* i, j. *Then there is a subsequence of* $\{f_j\}$ *which converges uniformly on compact subsets of* $\mathbb{R}^3$ *to an* η*-quasisymmetric embedding* $f : \mathbb{R}^3 \longrightarrow \mathbb{R}^4$.

This is a well-known and simple consequence of the Arzela-Ascoli theorem. It implies compactness properties for the set of subsets of $\mathbb{R}^4$ which are quasisymmetrically equivalent to $\mathbb{R}^3$ with respect to the Hausdorff topology. Condition $(**)$ has a similar compactness property built in to the definition, but without the benefit of a single parameterization which incorporates all the estimates. By contrast, we shall see that Condition $(*)$ fails to enjoy such compactness.

We shall also consider a condition weaker than $(*)$.

Definition 1.11. *A subset* E *of* $\mathbb{R}^n$ *satisfies Condition* $(\dagger)$ *if it is a topological manifold and if there is a* $C > 0$ *such that* $x \in E$ *and* $r > 0$ *imply that* $E \cap B(x, r)$ *can be contracted to a point inside* $E \cap B(x, Cr)$.

This kind of uniform contractability condition has gained in prominence in recent years. (See [F1], [F2], [GP], [GPW], [P].) Condition $(*)$ implies $(\dagger)$, and the two have similar features, *e.g.*, they both prevent cusps and long thin tubes. It turns out that $(\dagger)$ implies $(*)$ stably, in the sense that $(\dagger)$ for E implies $(*)$ for $E \times \mathbb{R}$, by a theorem of Ferry [F2], at least when E has topological dimension greater or equal than 4. (Basically this comes from [F2, Theorem 4.1], but I am cheating slightly here, because Ferry works only with compact sets and the corresponding versions of $(\dagger)$ and $(*)$.)

The following is the main result of this paper.

Theorem 1.12. a) *There exists a set in* $\mathbb{R}^4$ *which satisfies* (†) (*with dimension* 3) *but not* (∗).

b) *There exists a set in* $\mathbb{R}^4$ *which satisfies* (∗) (*with dimension* 3) *but not* (∗∗).

c) *There exists a set in* $\mathbb{R}^4$ *which satisfies* (∗∗) (*with dimension* 3) *but which does not admit a quasisymmetric parameterization.*

All of these sets can be chosen to be Ahlfors regular with dimension 3 and to have the property that there is a constant L_0 *so that every pair of points* p, q *in the set is contained in a closed subset* W *of the set such that* W *is* L_0*-bilipschitz equivalent to a closed Euclidean 3-ball. (In particular,* p *and* q *can be connected by a curve inside the set with length less or equal than* $L_0^2\,|p-q|$*. Thus the Euclidean distance on these sets are comparable in size to the internal geodesic distances on them.) These sets can also be taken to agree with a 3-plane outside a large ball, and to be homeomorphic to* $\mathbb{R}^3$.

For the record, the statement that two sets are C-bilipschitz equivalent means that there is a bijection between them which satisfies (1.2) for all pairs of points in the domain of the mapping.

The sets in Theorem 1.12 are smooth away from small singular sets, and the degeneracies near the singularities are well controlled (and have a natural self similarity to them). The statement about bilipschitz balls implies also that these sets are uniformly rectifiable in the sense of [DS4]. Thus there is nothing fractal happening, not even asymptotically. These sets are also well behaved in terms of analysis, and in particular there are Sobolev and Poincaré inequalities for them. (See Sections 9 and 10.)

Note that examples as in c) existed before, because of [Tu2], but without Ahlfors regularity (of the correct dimension), or the property about pairs of points being contained in bilipschitz balls, or the good bounds on the geodesic distances.

Part a) of Theorem 1.12 answers a question of Ferry and shows that the stabilization in his theorem is necessary. (He gave a slightly different version of this necessity in [F2, Theorem 4].)

What does Theorem 1.12 mean? The glib answer is that it means that one should assume more than Ahlfors regularity and (∗∗) if one wants to have geometric criteria for the existence of a quasisymmetric parameterization. I am inclined more to the view that the examples of Theorem 1.12 mean that quasisymmetric mappings are too rigid to accommodate some reasonable geometric phenomena, and that there is an interesting middle zone where geometry is approximately Euclidean and good enough for a lot of analysis but still substantially different from Euclidean geometry.

Note that quasiconformal parameterizations of the sets promised in Theorem 1.12 must be quasisymmetric and hence cannot exist. See the comments after Corollary 3.104.

The examples promised in Theorem 1.12 are basically geometric reformulations of classical examples from geometric topology and were inspired by the pictures in [D, Section 9]. Geometric topology is unusual in mathematics for its wealth of concrete examples, including some very interesting (topological) quotients of Euclidean spaces with a lot of topological self-similarity. The main point of Theorem 1.12 is that one can construct sets which contain the same topological information as in these examples but for which the topological self-similarity is converted into actual geometric self-similarity. Specifically, the examples for a), b), and c) of Theorem 1.12 correspond to the construction of the Whitehead continuum, Bing's dogbone space, and Bing doubling, respectively. It is easy to construct similar sets corresponding to other examples of the same type, but the main points of this general procedure are illustrated well by these three examples. The topological features of these classical examples imply interesting properties of the sets constructed here, from which the requirements of a) and b) will follow immediately. In the case of c) there is an additional point which was covered in [FS], and in fact Theorem 1.12.c) turns out to be almost just a reformulation of an example in [FS].

Juha Heinonen pointed out to me that the complementary components of the sets used to prove Theorem 1.12 are "uniform domains" (see (7.13) and (7.14)), and hence cannot be quasiconformally equivalent to a ball. (If they were, then there would have to be a quasisymmetric extension to the boundary, which is impossible.) In particular, if these domains are equipped with their quasihyperbolic metrics, then they cannot be bilipschitzly equivalent to the standard hyperbolic space. This is amusing, because these complementary domains are otherwise so nice. They are topological balls, with topologically tame boundaries, and we can even build them so that there are bilipschitz reflections across their boundaries (*i.e.*, across the sets promised in Theorem 1.12). The complementary components of the set in c) are particularly nice, because they satisfy a version of (**) adapted to domains. (See Theorem 8.1

below.) One could also build examples with the same properties using [Tu2, Lemma 4] (as discussed on [Tu2, p. 518]), but not with the control on the mass, uniform rectifiability, etc., which is available here.

Note that there are general results in [V2] which permit one to conclude that the complementary components of a set are uniform domains under natural uniform conditions on the topology of the set itself. One can also derive higher-order versions of the uniform domain condition in this way. In our case we shall be able to check easily and directly that the domains are uniform, but one should keep the general results in mind.

To understand the details of this paper it could be very helpful to have a copy of [D] handy. In particular [D] has excellent pictures. I shall provide references which permit one to do without [D], but the wonderfully clear book [D] provides the advantages of one-stop shopping. Good general references for aspects of geometric topology related to the topics of this paper include [C1], [C2], [D], [E], [K], and many papers of R. H. Bing.

The paper [Se5] addresses similar questions about bilipschitz instead of quasisymmetric mappings and relies on different examples from geometric topology (like Antoine's necklaces).

Some background information about geometric topology will be given in the next section, and a general construction will begin after that. Specific examples corresponding to Theorem 1.12 are described in Sections 4, 5, and 6, and the complements of these sets are discussed in Sections 7 and 8. Section 9 deals with analysis on these sets, with suitable versions of Sobolev and Poincaré inequalities established in Section 10. The simple fact that finite polyhedra which are topological manifolds satisfy the analogue of (**) for compact sets is given in Section 11, and the last section is devoted to miscellaneous remarks.

2. Some geometric topology.

The examples for Theorem 1.12 come from classical geometric topology, but before getting to that let us review briefly some of the more modern activity concerning the problem of finding homeomorphic parameterizations of a space. A good reference is the expository paper [C1], which begins with the following:

Recognition Problem 2.1. *Find a short list of topological properties,* *reasonably easy to check, that characterize topological manifolds among topological spaces.*

In dimensions 5 and higher there is now a reasonable characterization of topological manifolds coming from work of Edwards and Quinn [E], [Q1], [Q2], modulo a locally defined integer obstruction whose nontriviality was only recently established by Bryant, Ferry, Mio, and Weinberger [B^∞]. That is, they construct spaces which are "almost" manifolds but which have the wrong value of the aforementioned obstruction. Their construction is quite complicated, and the spaces they produced are not well understood. These spaces should be very interesting, because they are so well behaved but still distinct from Euclidean topology, and one can hope that they have interesting geometric realizations on which one can do some analysis.

To put the recognition problem into perspective it is helpful to consider the special case of finite polyhedra. It turns out that a finite polyhedron K is a topological manifold (without boundary) if and only if the link of every simplex in K is a homology sphere of the correct dimension, and if the links of vertices are simply connected when $\dim K > 2$. This result follows from the theorem of Edwards and Cannon [C2], [C3], [E] to the effect that double suspensions of homology spheres are topological spheres. When the dimension is 4 one must also use the Freedman theory [Fr]. As indicated in the introduction, it is known that the local homeomorphic parameterizations promised in this theorem cannot be bilipschitz in certain cases (see [SS] for details), and the existence of quasisymmetric homeomorphisms is an open problem.

The bottom line is that topologists have some serious technology for establishing the existence of homeomorphic parameterizations. They tend not to provide any estimates on the extent to which their parameterizations distort distances, but there are also examples which show that the homeomorphisms which they produce have to be complicated.

Of course the preceding discussion about the existence of homeomorphic parameterizations ignores local-to-global issues of the type addressed by the Schönflies theorem, the annulus and Poincaré conjectures, the H and S cobordism theorems, and surgery theory. The local and global questions are not truly separate -*e.g.*, the properties of a space at infinity can be reformulated in terms of local properties of the one-point compactification near the new point- and the distinction is particularly dubious in the context of parameterizations with scale-invariant bounds, as in bilipschitz and quasisymmetric conditions, where an obstruction to doing something globally can give rise to obstruction to doing it locally with a uniform bound.

Now let us consider a much more restricted version of the recognition problem which will be more directly connected to the proof of Theorem 1.12. There is an old result of R. L. Moore to the effect that a Hausdorff topological space X is homeomorphic to $\mathbf{S}^2$ if it can be realized as the image of a continuous mapping $f : \mathbf{S}^2 \to X$ with the property that $f^{-1}(x)$ and $\mathbf{S}^2 \setminus f^{-1}(x)$ are nonempty and connected for each $x \in X$. This is not to say that f is itself a homeomorphism; f could collapse an arc or a disk to a point, for instance. Observe that if we collapse a circle in $\mathbf{S}^2$ to a point we get an X which consists of two 2-spheres touching at a point. In this case both the hypothesis and conclusion of Moore's theorem fail to hold.

What happens in 3 dimensions? It turns out that there are some subtle negative results. To explain this it is helpful to introduce some auxiliary notions. (See [D] for details.) A *decomposition* G of $\mathbb{R}^3$ is simply a partition of $\mathbb{R}^3$, *i.e.*, a collection of disjoint subsets of $\mathbb{R}^3$ whose union is all of $\mathbb{R}^3$. (See [D, bottom of p. 7], or [K, p. 86].) Given a decomposition G of $\mathbb{R}^3$ we can form the usual quotient space $\mathbb{R}^3/G$ (as a topological space). Under reasonable hypotheses (which include the requirement that each element of G be a closed subset of $\mathbb{R}^3$) one knows that $\mathbb{R}^3/G$ is a Hausdorff space, and every decomposition that we shall consider will have a Hausdorff quotient. One would like to know when $\mathbb{R}^3/G$ is a topological manifold, which might even be homeomorphic to $\mathbb{R}^3$ itself. Moore's theorem provides a nontrivial criterion for this in 2 dimensions, but this is a much harder problem in 3 dimensions.

If F is a closed subset of $\mathbb{R}^3$, then we can define a decomposition associated to it by taking G to consist of F together with $\{x\}$ for all $x \in \mathbb{R}^3 \backslash F$. Thus $\mathbb{R}^3/G$ is simply the space that one gets by shrinking F to a point and leaving the rest of $\mathbb{R}^3$ alone. For instance, if one takes F to be a line segment, then $\mathbb{R}^3/G$ is homeomorphic to $\mathbb{R}^3$ again. This is also true if one takes F to be a (standard) closed 2-disk, or a (standard) closed 3-ball. If F is taken to be the unit sphere, then $\mathbb{R}^3/G$ is homeomorphic to $\mathbb{R}^3$ with a 3-sphere attached to it at one point.

What happens if F is a (standard) circle? Set $X = \mathbb{R}^3/G$, and let p denote the point in X which corresponds to F. Then $X\backslash\{p\}$ is not simply connected, because it is homeomorphic to $\mathbb{R}^3\backslash F$, and we can take a circle which links F to get a homotopically nontrivial loop in $X\backslash\{p\}$. This implies that X is not homeomorphic to $\mathbb{R}^3$, and a local version of the same argument shows that X is not a topological manifold at p.

This example shows that the most naive transcription of Moore's theorem does not work in 3-dimensions, since the complement of F in $\mathbb{R}^3$ is connected. There is a more interesting example based on the Whitehead continuum W in $\mathbb{R}^3$. The main properties of this continuum are the following: (i) it is cell-like, which means that it can be contracted to a point inside of any open set which contains it; (ii) when viewed as a subset of $\mathbf{S}^3$, its complement is contractable (unlike a circle, as in the previous example); (iii) there is an open set $U \subseteq \mathbb{R}^3$ which contains W such that there are loops in arbitrarily small neighborhoods of W which do not intersect W but which cannot be contracted to a point in $U \setminus W$. Equivalently, although $\mathbf{S}^3 \setminus W$ is contractable, it is not simply connected at infinity. Notice that a standard line segment satisfies (i) and (ii) above, but not (iii). If we let G be the decomposition associated to $F = W$ as before, then $X = \mathbb{R}^3/G$ is again not homeomorphic to $\mathbb{R}^3$. Indeed, if $p \in X$ corresponds to W, then $X \setminus \{p\}$ is simply connected in this case, but there is an open set V (the image in X of U in (iii)) which contains p and which has the property that there are loops in $X \setminus \{p\}$ which are as close as we want to p but which are not homotopically trivial in $V \setminus \{p\}$. Thus X is not a topological manifold at p.

The Whitehead continuum can be constructed through an iterative procedure as follows. One starts with a solid torus T in $\mathbb{R}^3$ and another solid torus T_1 embedded inside T as in [D, Figure 9-7, p. 68]. (See also [K, p. 81ff].) T_1 should be embedded into T in such a way that it is homotopically trivial but clasped to make it isotopically nontrivial. That is, T_1 cannot be deformed to a standard torus inside T without crossing itself. One then iterates this construction by identifying T_1 with T to get a new torus T_2 inside T_1, and then repeating the process indefinitely to get a sequence of nested solid tori T_j. The Whitehead continuum is obtained by taking the intersection of the T_j's. The key property (iii) above can be reduced to the fact that the meridional circle in ∂T pictured in [D, Figure 9-7] cannot be contracted to a point in T without touching the intersection of the T_j's. (See [D, Proposition 9, p. 76] and the remarks which preceed it. See also [K, p. 82].)

We shall use the Whitehead continuum in Section 4 to prove Theorem 1.12.a), and we shall employ similar iterative constructions for b) and c). Before we consider specific examples in detail we should formulate this iterative procedure in more general terms, starting with a basic definition from [D, bottom of p. 61].

Definition 2.2. *A defining sequence in $\mathbb{R}^3$ is a sequence $\{C_i\}$ of closed subsets of $\mathbb{R}^3$ such that each C_i is the closure of a bounded open set with smooth boundary and such that C_{i+1} is contained in the interior of C_i for each i. The C_i's need not be connected.*

Given a defining sequence $\{C_i\}$ in $\mathbb{R}^3$, we can define a decomposition G of $\mathbb{R}^3$ by taking the elements of G to be the components of $\cap C_i$ together with the singletons from $\mathbb{R}^3 \setminus \cap C_i$. In the earlier discussion of the Whitehead continuum each $C_i = T_i$ had only one component, as did $\cap C_i$. In the other examples considered here the number of components of C_i grows exponentially.

The iterative procedures that we shall use will be represented by the following. For the record, a "domain" is a connected open set.

Definition 2.3. *An initial package $\mathcal{P}$ consists of a bounded smooth domain D in $\mathbb{R}^3$, a finite collection $D_1, \ldots, D_n$ of smooth subdomains with disjoint closures contained in D, and mappings ϕ_j, $j = 1, \ldots, n$, such that each ϕ_j is a diffeomorphism from a neighborhood of $\overline{D}$ onto a neighborhood of $\overline{D}_j$ which maps D onto D_j.*

For example, in the construction of the Whitehead continuum we had an initial package (with $n = 1$) consisting of $D = T$, $D_1 = T_1$, and any reasonable choice of ϕ_1.

To an initial package $\mathcal{P}$ we can associate a defining sequence $\{C_i\}$ by taking C_0 to be $\overline{D}$, C_1 to be $\cup \overline{D}_j$, C_2 to be the union of the images of the $\overline{D}_k$'s under ϕ_j inside each $\overline{D}_j$, and so forth. Thus C_i is the union of n^i images of $\overline{D}$ under various compositions of the ϕ_j's.

As explained in [D, Section 9], there are some very interesting decompositions of $\mathbb{R}^3$ which can be obtained from an initial package in this way. One such decomposition G, due to Bing [B3], has a nonmanifold quotient $\mathbb{R}^3/G$ (called "the dogbone space") even though G has the property of being *cellular*. A compact set K in $\mathbb{R}^3$ is said to be cellular if for each open set $U \subseteq \mathbb{R}^3$ with $U \supseteq K$ there is a topological 3-ball contained in U which contains K, and a decomposition G of $\mathbb{R}^3$ is said to be cellular if each of its elements is a cellular set. (See p. 35 and Corollary 2A on p. 36 of [D].) The Whitehead continuum is definitely not cellular [D, p. 76, Proposition 9], and at one point it was apparently hoped that a cellular quotient of $\mathbf{S}^3$ would be $\mathbf{S}^3$ again. Bing's example shows that this is not true, and we shall use it to prove Theorem 1.12.b).

There is an older example which Bing considered in [B1] for which the quotient is homeomorphic to $\mathbb{R}^3$ but in a nontrivial way. For instance, this decomposition has a symmetry about a 2-plane which gives rise to an involution on $\mathbf{S}^3$ whose fixed point set is a wild sphere. This example was also used in [FS], and it will be used here to prove Theorem 1.12.c).

Our next task will be to take the construction of a decomposition and an associated quotient of $\mathbb{R}^3$ from an initial package and adapt it in such a way as to have better geometric properties. For a bare bones version of Theorem 1.12 all we really need to do is deform the Euclidean metric smoothly on D in such a way that the ϕ_j's become similarities near ∂D. A straightforward iterative construction would then allow us to build a metric with respect to which the ϕ_j's are similarities on all of $\overline{D}$. This metric would deteriorate near the elements of the decomposition, as it should. In this way we can build a metric space which is topologically equivalent to the decomposition space and which has nice geometric self-similarity properties. For the sake of concreteness and other benefits it is better to construct theses spaces as subsets of Euclidean spaces, and the examples used in this paper even fit into $\mathbb{R}^4$. In fact their embeddings into $\mathbb{R}^4$ and their complementary components have some especially nice properties for which we shall take extra care to make manifest. If one simply wants to build such sets without worrying about extra properties then some of the efforts and assumptions in the next section are unnecessary. (See the remarks after Definition 3.2.)

3. The general construction.

The next definition describes excellent packages, which consist of an initial package $\mathcal{P}$ together with objects which will allow us to convert $\mathcal{P}$ into a topologically equivalent object in $\mathbb{R}^4$ for which the analogue of the ϕ_j's are similarities. Recall that a (Euclidean) similarity on $\mathbb{R}^4$ is an affine transformation which is a combination of a translation, dilation by a positive number (called the dilation factor of the similarity), and an orthogonal transformation.

Convention 3.1. *P denotes the $x_4 = 0$ hyperplane in $\mathbb{R}^4$, and from now on we identify it with $\mathbb{R}^3$, so that any objects living on $\mathbb{R}^3$ (like an initial package) will be viewed as living on P.*

Definition 3.2. *An excellent package $\mathcal{E}$ consists of an initial package $\mathcal{P} = \{D, D_1, \ldots, D_n, \phi_1, \ldots, \phi_n\}$ as in Definition 2.3 together with a bounded smooth domain Ω in $\mathbb{R}^4$, smooth subdomains $\omega_1, \ldots, \omega_n$ of Ω with disjoint closures contained in Ω, another collection of smooth subdomains $\Omega_1, \ldots, \Omega_n$ with disjoint closures contained in Ω, orientation-preserving similarities ψ_j, $1 \le j \le n$, on $\mathbb{R}^4$ which all preserve P and have the same dilation factor $\rho \in (0, 1/10)$, and a diffeomorphism θ on $\mathbb{R}^4$ which satisfy the following properties:*

$$\overline{\Omega} \cap P = \overline{D}\,,\ \overline{\omega}_j \cap P = \overline{D_j}\,,\ \textit{and}\ \partial\Omega,\ \partial\omega_j\ \textit{all intersect } P \textit{ transversely;} \tag{3.3}$$

$$\psi_j(\Omega) = \Omega_j \textit{ for all } j\,; \tag{3.4}$$

$$\theta(\Omega) = \Omega \textit{ and } \theta = \textit{the identity on } \mathbb{R}^4 \backslash \Omega \textit{ and on a neighborhood of } \partial\Omega; \tag{3.5}$$

$$\theta(\omega_j) = \Omega_j \textit{ and } \theta = \psi_j \circ \phi_j^{-1} \textit{ on a neighborhood of } \overline{D}_j \textit{ in } P \textit{ for each } j\,. \tag{3.6}$$

Thus Ω and the ω_j's are fattened-up versions of D and the D_j's in $\mathbb{R}^4$, the Ω_j's are "straightened" versions of the ω_j's, and θ converts the slightly twisted D_j's into the straighter $\Omega_j \cap P$'s at the cost of deforming P inside Ω.

These excellent packages have more structure than we actually need for Theorem 1.12. Instead of the requirement that θ exist as a diffeomorphism on all of $\mathbb{R}^4$ it would be enough to have θ as an embedding of P into $\mathbb{R}^4$, say. If were also willing to work in a larger space than $\mathbb{R}^4$ then this weaker version of an excellent package would exist for any initial package. ($\mathbb{R}^7$ is plenty large enough.) As it is, the above definition of an excellent package imposes some topological restrictions on an initial package. These restrictions will be satisfied in the examples that we shall consider, and the extra structure that we obtain will be pleasant to have. Note that the kind of bare-bones deformation of the metric for an initial package described at the end of Section 2 involves no topological obstructions whatsoever.

For the rest of this section we shall assume that we are given an excellent package $\mathcal{E}$ as above and build some surfaces from it. This construction is similar to the one used in [Se5, Section 5] (with Lemma 5.6 there providing the excellent package).

Define $M^1 \subseteq \mathbb{R}^4$ by $M^1 = \theta(P)$. Thus M^1 is the same as P outside Ω, it agrees with $\psi_j(D) = \Omega_j \cap P$ inside Ω_j, and it is some smooth manifold in $\Omega \setminus \cup\Omega_j$. We want to define a sequence of submanifolds M^j with many twists like the ones in M^1, but to do this we should first sort out the relevant codings.

Let $\mathcal{S}_l$ denote the finite sequences $\alpha = \{a_i\}_{i=1}^l$ with l terms such that $a_i \in \{1, \ldots, n\}$ for all i. Thus we can identify $\mathcal{S}_1$ with $\{1, \ldots, n\}$, and it will be useful to consider $\mathcal{S}_0$ as a set with just one element, the empty sequence, sometimes denoted $\varnothing$. Define ψ_α and Ω_α for $\alpha \in \mathcal{S}_l$ recursively in the following manner. If $l = 1$, so that we can view α as an element of $\{1, \ldots, n\}$, then we simply take ψ_α and Ω_α to be as in Definition 3.2 above. If $l > 1$ and $\alpha = \{a_i\}_{i=1}^l$, then let $\alpha' = \{a_i\}_{i=1}^{l-1}$ be the "parent" of α in $\mathcal{S}_{l-1}$ and set $\Omega_\alpha = \psi_{\alpha'}(\Omega_{a_l})$ and $\psi_\alpha = \psi_{\alpha'} \circ \psi_{a_l}$. We view the empty sequence in $\mathcal{S}_0$ as being the parent of the elements of $\mathcal{S}_1$, so that the preceding equations hold with $\Omega_\varnothing = \Omega$ and $\psi_\varnothing$ taken to be the identity. We shall call two elements of $\mathcal{S}_l$ "siblings" if they have the same parent, and extend this terminology to the Ω_α's as well. With these conventions we have the following properties for ψ_α and Ω_α for each $\alpha \in \mathcal{S}_l$, $l \geq 1$, with α' the parent of α:

(3.7) ψ_α is a similarity with dilation factor ρ^l, and $\psi_\alpha(P) = P$,

(3.8) $$\Omega_\alpha = \psi_\alpha(\Omega),$$

(3.9) $\overline{\Omega}_\alpha \subset \Omega_{\alpha'}$ and $\overline{\Omega}_\alpha$ is disjoint from its siblings in $\Omega_{\alpha'}$.

The Ω_α's have additional nesting properties, which we state as a lemma.

Lemma 3.10. *Suppose that $\alpha \in \mathcal{S}_l$ and $\beta \in \mathcal{S}_k$. Then either Ω_α and Ω_β are disjoint (and have disjoint closures), or $\Omega_\alpha \subseteq \Omega_\beta$, in which case $l \geq k$ and α is a descendant of β, or $\Omega_\beta \subseteq \Omega_\alpha$, in which case $k \geq l$ and β is a descendant of α. In particular Ω_α and Ω_β are disjoint (and have disjoint closures) when $k = l$ and $\alpha \neq \beta$.*

To see this choose $\gamma \in \mathcal{S}_m$ to be the common ancestor of α and β with m as large as possible (but perhaps $= 0$). If γ is equal to either α or β, then one is an ancestor of the other, and we are in business. Otherwise, α and β are descended from distinct children of γ, and disjointness follows from (3.9).

We can define ϕ_α and D_α in the same way as ψ_α and Ω_α, so that ϕ_α and D_α are given to us as part of our initial package when $\alpha \in \mathcal{S}_1$, $\phi_\alpha = \phi_{\alpha'} \circ \phi_{a_l}$ when $l > 1$, $\alpha = \{a_i\}_{i=1}^l$, and α' is the parent of α, and D_α is defined to be $\phi_\alpha(D)$. The D_α's satisfy the same nesting properties as the Ω_α's, in the sense that the analogue of Lemma 3.10 for the D_α's is true, with the same proof. If we set $C_l = \cup_{\alpha \in \mathcal{S}_l} \overline{D}_\alpha$, then this is equivalent to the defining sequence mentioned after Definition 2.3.

Set

$$F = \bigcap_{l=1}^{\infty} \bigcup_{\alpha \in \mathcal{S}_l} \overline{\Omega}_\alpha \,, \tag{3.11}$$

and let $\mathcal{S}$ denote the collection of all infinite sequences $s = \{s_i\}$ of elements of $\{1, \ldots, n\}$, so that $\mathcal{S}$ is the natural limit of the $\mathcal{S}_l$'s. Thus F is a Cantor set in $\mathbb{R}^4$ which actually lies in P, because of (3.7) and (3.8), and there is a natural bijection $f : \mathcal{S} \to F$ which is defined in the obvious manner. (Each element s of $\mathcal{S}$ determines a nested sequence of Ω_α's which converges to a point (by (3.7)-(3.9)), and $f(s)$ is defined to be this point. Conversely, every element of F must arise from such a nested sequence of Ω_α's, by Lemma 3.10.) If we perform the same construction for the $D'_\alpha s$ instead of the Ω_α's, then we might not get a Cantor set but a more complicated set with nontrivial components. These components are the nontrivial elements of the decomposition associated to the defining sequence C_l above (as described just after Definition 2.2). (See Sublemma 3.40.)

Set $Y = \overline{\Omega} \setminus \cup_{j=1}^n \Omega_j$ and $Y_\alpha = \psi_\alpha(Y)$. These compact sets in $\mathbb{R}^4$ are the closures of smooth domains. As before let us set $Y_\varnothing = Y$, where $\varnothing$ is the empty sequence in $\mathcal{S}_0$. If $\alpha \in \mathcal{S}_l$ and $\beta \in \mathcal{S}_k$, and if Y_α intersects Y_β, then one of α, β is the parent of the other, and their intersection equals $\partial\Omega_\gamma$, where γ is whichever of α and β is the child. This is an easy consequence of Lemma 3.10, (3.9), and the definitions. Notice also that

$$\mathbb{R}^4 = (\mathbb{R}^4 \backslash \Omega) \cup \Big(\bigcup_{l=0}^{m} \bigcup_{\alpha \in \mathcal{S}_l} Y_\alpha \Big) \cup \Big(\bigcup_{\alpha \in \mathcal{S}_{m+1}} \Omega_\alpha \Big) \tag{3.12}$$

for each $m \geq 0$, and

$$\mathbb{R}^4 = (\mathbb{R}^4 \backslash \Omega) \cup \Big(\bigcup_{l=0}^{\infty} \bigcup_{\alpha \in \mathcal{S}_l} Y_\alpha \Big) \cup F \,. \tag{3.13}$$

The basic building block for the construction of the M^j's is

$$\Sigma = \theta\Big(\overline{D} \setminus \bigcup_{j=1}^{n} D_j\Big). \tag{3.14}$$

This is a compact embedded 3-dimensional submanifold of $\mathbb{R}^4$ with boundary which satisfies $\Sigma \subset Y$ and $\Sigma = P$ on a neighborhood of ∂Y inside Y (by (3.5) and (3.6)). Setting $\Sigma_\alpha = \psi_\alpha(\Sigma)$ for $\alpha \in \mathcal{S}_l$, $l \geq 0$ (so that $\Sigma_\varnothing = \Sigma$ when $l = 0$, as usual), we have that

$$\Sigma_\alpha \subset Y_\alpha \tag{3.15}$$

and

$$\Sigma_\alpha = P \text{ on a neighborhood of } \partial Y_\alpha \text{ inside } Y_\alpha\,. \tag{3.16}$$

Define M^j for $j \geq 1$ by

$$M^j = (P \setminus D) \cup \Big(\bigcup_{l=0}^{j-1} \bigcup_{\alpha \in \mathcal{S}_l} \Sigma_\alpha\Big) \cup \Big(\bigcup_{\alpha \in \mathcal{S}_j} \Omega_\alpha \cap P\Big). \tag{3.17}$$

In accordance with our usual conventions for $l = 0$ we set $M^0 = P$. It is easy to see that this agrees with $M^1 = \theta(P)$ from before, because of the definitions and the properties of θ. All the M^j's are embedded smooth submanifolds of $\mathbb{R}^4$, because of (3.16) and the disjointness properties of the Y_α's. Also,

$$M^j = M^k \quad \text{outside} \quad \bigcup_{\alpha \in \mathcal{S}_j} \Omega_\alpha\,, \qquad \text{if } j \leq k\,. \tag{3.18}$$

This implies in particular that the M^j's are converging in the Hausdorff topology to

$$M = (P \setminus D) \cup \Big(\bigcup_{l=0}^{\infty} \bigcup_{\alpha \in \mathcal{S}_l} \Sigma_\alpha\Big) \cup F\,. \tag{3.19}$$

This is a smooth embedded submanifold away from the Cantor set $F \subseteq P$. It will sometimes be convenient to denote M by M^∞ to make the notation more uniform.

These submanifolds have the self-similarity property that

$$(3.20) \qquad \Omega_\alpha \cap M^j = \psi_\alpha(\Omega \cap M^{j-l}) \quad \text{when } \alpha \in \mathcal{S}_l \text{ and } 0 \le l \le j\,.$$

Here we allow $j = \infty$, which gives the relevant property for M. This equality is not hard to derive from the definitions (which were chosen precisely so that this would be true).

These are basically the sets that we are interested in, but for technical reasons we shall need to define another set $\widetilde{M}$ which is not quite as singular as M but which contains a (small) copy of each M^j. To do this let $\{B_k\}_{k=1}^{\infty}$ be a sequence of balls in $\mathbb{R}^4$ with disjoint doubles whose radii tend to 0 and whose centers lie on P and converge to the origin, and let $A_k : \mathbb{R}^4 \longrightarrow \mathbb{R}^4$ be affine mappings composed of translations and (nonzero) dilations such that $A_k(\overline{\Omega}_k) \subseteq B_k$ and $A_k(P) = P$. Set $\widetilde{M}^k = A_k(M^k)$, so that $\widetilde{M}^k = P$ outside B_k (and the interesting part of $\widetilde{M}^k$ is contained in B_k). Let $\widetilde{M}$ be the subset of $\mathbb{R}^4$ such that $\widetilde{M} = \widetilde{M}^k$ inside each B_k and $\widetilde{M} = P$ outside $\bigcup_k B_k$. (Compare with [Tu1, Example on p. 69] and [HY, Example 6.6].) For convenience we also require that there exist balls B'_k with radius $\le 100\,\text{radius}\, B_k$ such that $B_l \subseteq B'_k$ when $l \ge k$ but $B_l \cap B'_k = \varnothing$ when $l < k$. This is not hard to arrange, by taking the B_k's to be $B(z_k, 2^{-2k-3})$ with $|z_k| = 2^{-2k}$ and setting $B'_k = B(0, 2^{-2k+1})$, for instance. For the clarity of future arguments it is best for us to simply require that the B_k's be chosen in this manner, and with all the z_k's lying on the same line through the origin, rather than worry about the level of generality in which we can do this construction.

It is clear that the M's have a lot of self-similarity, but we also need to keep track of their topological properties. Recall the definition of the D_α's and C_l's (just before (3.11)), and let G be the decomposition associated to the defining sequence $\{C_l\}$, as discussed just after Definition 2.2. We want to show that M is homeomorphic to $\mathbb{R}^3/G$, where G is the decomposition associated to our initial package in the manner described after Definition 2.3. We also want to build some parameterizations of the M^j's which approximate the aforementioned homeomorphism in a nice way.

Lemma 3.21. *There exist diffeomorphisms h_j from $\mathbb{R}^4$ onto itself, $0 \le j < \infty$, and a continuous mapping h from $\mathbb{R}^4$ onto itself with the following properties: $h_j(\mathbb{R}^3) = M^j$ and $h(\mathbb{R}^3) = M$; $h = h_j = \theta$ on $\mathbb{R}^4 \setminus (\cup_{i=1}^n \theta^{-1}(\Omega_i))$; $h_j \to h$ uniformly on $\mathbb{R}^4$; $h(D_\alpha) = M \cap \Omega_\alpha$ for α in any $\mathcal{S}_l$ and $h_j(D_\alpha) = M^j \cap \Omega_\alpha$ for α in any $\mathcal{S}_l$ with $l \le j$; h*

is constant on each element of the decomposition G, and it induces a homeomorphism from $\mathbb{R}^3/G$ onto M. There is also a homeomorphism $\widetilde{h}$ from $\mathbb{R}^4$ onto itself which maps $\mathbb{R}^3$ onto $\widetilde{M}$ and which equals the identity off of each B_k (and off each $A_k(\overline{\Omega})$), which maps each $B_k \cap P$ onto $B_k \cap \widetilde{M}$, and which is a diffeomorphism away from the origin.

Let us define first little copies of θ on the various Ω_α's. Set

$$\theta_\alpha = \psi_\alpha \circ \theta \circ \psi_\alpha^{-1}, \tag{3.22}$$

so that

$$\theta_\alpha(\Omega_\alpha) = \Omega_\alpha \text{ and } \theta_\alpha = \text{ the identity on } \mathbb{R}^4 \backslash \Omega_\alpha \text{ and on a neighborhood of } \partial\Omega_\alpha\,. \tag{3.23}$$

If α is the empty sequence in $\mathcal{S}_0$, then $\theta_\alpha = \theta$. If $\beta = \{b_i\}_{i=1}^{l+1} \in \mathcal{S}_{l+1}$ is the child of α with $b_{l+1} = p$, then

$$\theta_\alpha = \psi_\beta \circ \phi_p^{-1} \circ \psi_\alpha^{-1} \text{ on a neighborhood of } \psi_\alpha(\overline{D}_p) \text{ in } P \tag{3.24}$$

by (3.6).

Let g_l denote the composition of all θ_α for $\alpha \in \mathcal{S}_l$. Because the Ω_α's for $\alpha \in \mathcal{S}_l$ are pairwise disjoint (by Lemma 3.10), these θ_α's commute, and so we need not to worry about how we do the composition. Note that $g_0 = \theta$,

$$g_l(\Omega_\beta) = \Omega_\beta \quad \text{whenever } \beta \in \mathcal{S}_k\,,\ k \leq l\,, \tag{3.25}$$

and

$$g_l = \text{ the identity on } \mathbb{R}^4 \backslash \Big(\bigcup_{\gamma \in \mathcal{S}_k} \Omega_\gamma \Big) \quad \text{when } k \leq l\,. \tag{3.26}$$

These observations follow easily from Lemma 3.10.

Define h_j for $j \geq 1$ by

$$h_j = g_{j-1} \circ g_{j-2} \circ \cdots \circ g_0\,. \tag{3.27}$$

These are obviously diffeomorphisms, and $h_1 = g_0 = \theta$. We can take h_0 to be the identity, for completeness. Note that

$$h_k \circ h_j^{-1} = \text{ the identity on } \mathbb{R}^4 \backslash \Big(\bigcup_{\alpha \in \mathcal{S}_j} \Omega_\alpha \Big) \tag{3.28}$$

and

$$h_k \circ h_j^{-1}(\Omega_\alpha) = \Omega_\alpha , \qquad \text{for all } \alpha \in \mathcal{S}_j \tag{3.29}$$

when $j \leq k$. In particular

$$h_j = h_1 \text{ on } \mathbb{R}^4 \setminus \Big(\bigcup_{i=1}^{n} \theta^{-1}(\Omega_i) \Big) \text{ when } j > 1 . \tag{3.30}$$

Let us check that

$$\begin{aligned} &h_j = \psi_\alpha \circ \phi_\alpha^{-1} \text{ on } \overline{D}_\alpha \quad (\text{and hence } h_j(D_\alpha) = \Omega_\alpha \cap P) \\ &\text{when } \alpha \in \mathcal{S}_j . \end{aligned} \tag{3.31}$$

We do this by induction. When $j = 1$ this reduces to (3.6) and the definitions of D_α and Ω_α. Suppose now that we know (3.31) for some value of j and that we want to verify it for $j + 1$. Let $\beta \in \mathcal{S}_{j+1}$ be given, and let $\alpha \in \mathcal{S}_j$ be its parent. Set $p = b_{j+1} \in \{1, \ldots, n\}$, where $\beta = \{b_i\}_{i=1}^{j+1}$. From our induction hypothesis we get that $h_j(\overline{D}_\beta) \subseteq h_j(D_\alpha) \subseteq \Omega_\alpha$, and so $h_{j+1} = g_j \circ h_j = \theta_\alpha \circ h_j$ on $\overline{D}_\beta$. Our induction hypothesis also gives $h_j(D_\beta) = \psi_\alpha \circ \phi_\alpha^{-1}(D_\beta)$. By definitions we have that $D_\beta = \phi_\beta(D)$ and $\phi_\beta = \phi_\alpha \circ \phi_p$, and so $\phi_\alpha^{-1}(D_\beta) = \phi_\alpha^{-1}(\phi_\beta(D)) = \phi_p(D) = D_p$. Thus $h_j(\overline{D}_\beta) = \psi_\alpha(\overline{D}_p)$, and this permits us to use (3.24) to obtain $h_{j+1} = \psi_\beta \circ \phi_\beta^{-1}$ on $\overline{D}_\beta$ from our induction hypothesis that (3.31) holds for j and α. This in turn implies that $h_{j+1}(D_\beta) = \psi_\beta(D)$, and this last is the same as $\Omega_\beta \cap P$ because of (3.3), (3.7), and (3.8). This proves (3.31).

Observe that

$$h_k(D_\alpha) \subseteq \Omega_\alpha , \qquad \text{when } \alpha \in \mathcal{S}_j , \; j \leq k . \tag{3.32}$$

This follows from (3.31) and (3.29).

Set $E = \overline{D} \setminus \cup_{j=1}^{n} D_j$ and $E_\alpha = \phi_\alpha(E)$ for any α in any $\mathcal{S}_l$. Thus E_α is the same as $\overline{D}_\alpha$ with the children of D_α removed (*i.e.*, the D_β's with $\beta \in \mathcal{S}_{l+1}$ a child of α). As usual we have $E_\varnothing = E$ for the empty sequence in $\mathcal{S}_0$. Let us check that

$$h_k(E_\alpha) = \Sigma_\alpha , \qquad \text{when } \alpha \in \mathcal{S}_j , \; j < k . \tag{3.33}$$

It suffices to check this when $k = j+1$, because $h_k \circ h_{j+1}^{-1} =$ the identity on Σ_α when $\alpha \in \mathcal{S}_j$ and $k \geq j + 1$, because of (3.29) and (3.15).

Because $h_j(E_\alpha) \subseteq h_j(\overline{D}_\alpha) \subseteq \overline{\Omega}_\alpha$ we get that $g_j = \theta_\alpha$ on $h_j(E_\alpha)$, while (3.31) implies that $h_j(E_\alpha) = \psi_\alpha(E)$. Hence $h_{j+1}(E_\alpha) = g_j(h_j(E_\alpha)) = \theta_\alpha(\psi_\alpha(E)) = \psi_\alpha(\theta(E))$ by (3.22). Using (3.14) we get $\psi_\alpha(\theta(E)) = \psi_\alpha(\Sigma) = \Sigma_\alpha$, which proves (3.33).

It is now easy to check that $h_j(D_\alpha) = M^j \cap \Omega_\alpha$ for α in any $\mathcal{S}_l$ with $l \leq j$, using (3.33), (3.31), the definition (3.17) of M^j, and the nesting properties of the Ω_β's (as in Lemma 3.10). We also have that

$$h_k(\mathbb{R}^3 \backslash \Big(\bigcup_{\alpha \in \mathcal{S}_k} D_\alpha)\Big) \subseteq \mathbb{R}^4 \backslash \Big(\bigcup_{\alpha \in \mathcal{S}_k} \Omega_\alpha \Big) \tag{3.34}$$

because of (3.33). This implies that

$$h_m = h_k \text{ on } \mathbb{R}^3 \backslash \Big(\bigcup_{\alpha \in \mathcal{S}_k} D_\alpha \Big), \qquad \text{when } m \geq k, \tag{3.35}$$

because of (3.28).

From (3.28) and (3.29) we have that

$$\|h_j - h_k\|_\infty \leq \sup_{\alpha \in \mathcal{S}_j} \operatorname{diam} \Omega_\alpha = \rho^j \operatorname{diam} \Omega \qquad \text{when } j \leq k, \tag{3.36}$$

where ρ is the dilation factor of the ψ_j's (as in Definition 3.2). This implies that the h_j's converge uniformly to a continuous mapping $h : \mathbb{R}^4 \longrightarrow \mathbb{R}^4$. Notice that

$$h = h_j \text{ on } \mathbb{R}^3 \backslash \Big(\bigcup_{\alpha \in \mathcal{S}_j} D_\alpha \Big), \tag{3.37}$$

$$h = h_1 = \theta \text{ on } \mathbb{R}^4 \backslash \Big(\bigcup_{i=1}^{n} \theta^{-1}(\Omega_i) \Big), \tag{3.38}$$

and

$$h(\overline{D}_\alpha) \subseteq \overline{\Omega}_\alpha, \qquad \text{for any } \alpha \in \mathcal{S}_j \text{ and any } j, \tag{3.39}$$

because of (3.35), (3.30), and (3.32).

Let $\{C_l\}$ be the defining sequence associated to the D_α's, so that $C_l = \cup_{\alpha \in \mathcal{S}_l} \overline{D}_\alpha$, and set $C = \cap C_l$. By definition our decomposition G of $\mathbb{R}^3$ consists of the components of C and the singletons in $\mathbb{R}^3 \backslash C$. We understand h well on $\mathbb{R}^3 \backslash C$, because of (3.37), and we want to understand it on C. We must first analyze the components of C.

Let $\mathcal{S}$ be as before (defined just after (3.11)). Given a sequence $s = \{s_i\}_{i=1}^{\infty} \in \mathcal{S}$, let A_s denote the intersection of the $\overline{D}_\alpha$'s which come from ancestors α of s. That is, $\alpha \in \mathcal{S}_j$ is an ancestor of s if $\alpha = \{s_i\}_{i=1}^{j}$. It is easy to check that the A_s's are pairwise disjoint, using the nesting properties of the D_α's.

Sublemma 3.40. *$C = \bigcup_{s \in \mathcal{S}} A_s$, and the A_s's are the connected components of C.*

C contains all the A_s's by definition. Conversely, if $p \in C$, then for each j there is an $\alpha_j \in \mathcal{S}_j$ such that $p \in \overline{D}_{\alpha_j}$. The analogue of Lemma 3.10 for the D's instead of the Ω's implies that α_{j+1} must be a child of α_j for each j. This means that the α_j's combine to form a sequence $s \in \mathcal{S}$, and it is clear that $p \in A_s$.

Each A_s is connected. For this we use the fact that each D_α is connected, since our original domain D is (by the definition of domain). Thus A_s is the decreasing intersection of compact connected sets, and it is an elementary general fact that A_s must itself be connected under these conditions. (If A_s were disconnected, it would be contained in the union of two disjoint open sets, each of which intersects it, and the same would then have to be true for some $\overline{D}_\alpha$.)

Each A_s is a component of C. If not, there would be an A_s and an A_t which lie in the same component of C, with $s \neq t$. Let $\alpha \in \mathcal{S}_j$ be the common ancestor of s and t with j as large as possible (but perhaps equal to 0). Then C is contained in the union of the complement of D_α and the closures of the children of D_α, and A_s and A_t lie in different children of D_α. Since the closures of these children are disjoint and lie in D_α we see that A_s and A_t cannot both touch the same component of C. This proves Sublemma 3.40.

Recall the (bijective) mapping $f : \mathcal{S} \longrightarrow F$ defined just after the definition of the Cantor set F in (3.11). We have that

$$h(p) = f(s), \qquad \text{for all } p \in A_s \text{ and } s \in \mathcal{S}, \tag{3.41}$$

because of (3.39) and the definitions of A_s and $f(s)$. It is easy to check that

$$h(D_\alpha) = M \cap \Omega_\alpha, \qquad \text{for any } \alpha \text{ in any } \mathcal{S}_l, \tag{3.42}$$

because of (3.19), (3.37), (3.18), (3.41), and the corresponding statement for the h_j's (just before (3.34)). In other words, h looks like the

h_j's away from C, M looks like the M^j's away from F, and C and F correspond under h the way that they should because of (3.41).

It is not hard to check that h induces a homeomorphism from $\mathbb{R}^3/G$ onto M. Sublemma 3.40 provides us with a complete understanding of the decomposition G, we know how h behaves on and near the nontrivial elements of G because of (3.41) and (3.42), and we understand h away from C because of (3.37).

The last part of Lemma 3.21, about $\widetilde{h}$, is an easy consequence of the earlier part. That is, we set $\widetilde{h}$ = the identity off the $A_k(\Omega)$'s and $\widetilde{h} = A_k \circ h_k \circ A_k^{-1}$ on each $A_k(\Omega)$. Because each h_k equals the identity off Ω and maps Ω onto itself it is easy to see that this is a homeomorphism, and it maps $\mathbb{R}^3$ onto $\widetilde{M}$ by definition. It is also clear that it is a diffeomorphism away from the origin.

This completes the proof of Lemma 3.21.

RECAPITULATION 3.43. We started with an excellent package, which contained an initial package. From the initial package we can generate a decomposition of $\mathbb{R}^3$ as in Section 2. The excellent package is, roughly speaking, a topologically equivalent version of the initial package which lives in $\mathbb{R}^4$ and has better geometric properties (like self-similarity). We have now used the excellent package to construct sets M, M^j, and $\widetilde{M}$, and Lemma 3.21 tells us that they have the correct compatibility with the decomposition associated to the initial package. To prove Theorem 1.12 we shall choose specific excellent packages (in Sections 4, 5, and 6) and use the resulting sets M, $\widetilde{M}$.

Before we consider specific examples we shall establish some general properties of the construction which will be relevant for all of the examples. For this we shall need a case-by-case analysis of the positions of balls centered on these sets whose basic structure will be used repeatedly in this paper, and so we establish it first. For the moment we restrict ourselves to M and M^j and forget about $\widetilde{M}$.

Lemma 3.44. *Let $x \in M^j$ and $r > 0$ be given, where we allow $j = \infty$ in which case $M^j = M$. Let $a > 0$ be a small number that we get to choose but which should be small enough so that $2a \operatorname{diam} \Omega < \operatorname{dist}(\Omega_p, \Omega_q)$ for all $p, q = 1, \ldots, n$, $p \neq q$. Then one of the following alternatives holds:*

i) $B(x, r) \cap \Omega = \varnothing$,

ii) $B(x, r) \cap \Omega \neq \varnothing$ *and* $r \geq a \operatorname{diam} \Omega$,

iii) $B(x,r) \cap \Omega \neq \varnothing$, $r < a \operatorname{diam} \Omega$, *and* $B(x,r) \not\subseteq \Omega$,

iv) *there is an* α *in some* $\mathcal{S}_l$, $0 < l \leq j$, *such that* $B(x,r) \subseteq \Omega_\alpha$ *and* $r \geq a \operatorname{diam} \Omega_\alpha$,

v) *there is an* α *in some* $\mathcal{S}_l$, $0 \leq l \leq j$, *such that* $B(x,r) \subseteq \Omega_\alpha$, $r \leq a \operatorname{diam} \Omega_\alpha$, *and either* $l = j$ *or* $B(x,r) \cap Y_\alpha \neq \varnothing$.

Recall that Y_α was defined just before (3.12), and note that $l = j$ is not an option when $j = \infty$.

Lemma 3.44 is quite straightforward. If none of the first three cases hold then we have $r < a \operatorname{diam} \Omega$ and $B(x,r) \subset \Omega$. Choose $l \leq j$ as large as possible so that $B(x,r) \subset \Omega_\alpha$ for some $\alpha \in \mathcal{S}_l$, where $l = 0$ is allowed. If also $r \geq a \operatorname{diam}\Omega_\alpha$, then iv) obtains (unless $l = 0$, in which case ii) was already satisfied). If $r \leq a \operatorname{diam}\Omega_\alpha$, then v) has to hold because we took l to be as large as possible. That is, if $l < j$ and $B(x,r) \cap Y_\alpha = \varnothing$, then we could replace Ω_α by one of its children. For this last step we need to know that a is sufficiently small so that $B(x,r)$ cannot touch two different children of Ω_α, and the condition on a in the lemma ensures precisely this.

Let us now establish the Ahlfors regularity of these sets.

Lemma 3.45. *If* $\rho^3 n \leq 1$, *where* ρ, n *are as in Definition* 3.2, *then the sets* M, M^j, *and* $\widetilde{M}$ *are all regular with dimension* 3, *and with a constant that is bounded independently of* j.

Notice first that M, M^j, and $\widetilde{M}$ are all closed.

We should begin with some preliminary facts. The first is that

$$H^3(F) = 0\,. \tag{3.46}$$

This follows from the definition (3.11) of F, which implies that $F \subseteq \bigcup_{\alpha \in \mathcal{S}_l} \overline{\Omega}_\alpha$ for each l, so that

$$H^3(F) \leq \limsup_{l \to \infty} \sum_{\alpha \in \mathcal{S}_l} (\operatorname{diam} \overline{\Omega}_\alpha)^3 \leq \limsup_{l \to \infty} n^l \rho^{3l} (\operatorname{diam} \Omega)^3 \tag{3.47}$$

by definition of Hausdorff measure. (Do not forget that $\mathcal{S}_l$ has n^l elements.) This implies (3.46), since we are assuming that $\rho^3 n < 1$. Notice that a similar argument implies that

$$\text{Hausdorff dimension}\,(F) \leq d\,, \qquad \text{if } \rho^d n < 1\,. \tag{3.48}$$

In our examples we shall have the freedom to choose ρ to be as small as we wish, and so we can make F have Hausdorff dimension as small as we wish.

Next we check that

$$H^3(\Omega \cap M^j) \leq C < \infty\,, \tag{3.49}$$

for some constant C which does not depend on j. In view of (3.46) and the definitions (3.17) and (3.19) of M^j and M we are reduced to estimating

$$\sum_{l=0}^{j-1} \sum_{\alpha \in \mathcal{S}_l} H^3(\Sigma_\alpha) + \sum_{\alpha \in \mathcal{S}_j} H^3(\Omega_\alpha \cap P)\,. \tag{3.50}$$

Since Hausdorff measure behaves properly under similarities we have that $H^3(\Sigma_\alpha) = \rho^{3l}\, H^3(\Sigma)$ when $\alpha \in \mathcal{S}_l$ and $H^3(\Omega_\alpha \cap P) = \rho^{3j}\, H^3(\Omega \cap P)$ when $\alpha \in \mathcal{S}_j$, and so (3.50) reduces to

$$\sum_{l=0}^{j-1} n^l\, \rho^{3l}\, H^3(\Sigma) + n^j\, \rho^{3j}\, H^3(\Omega \cap P)\,. \tag{3.51}$$

The desired bound follows from our assumption that $\rho^3\, n < 1$, since $H^3(\Sigma)$ and $H^3(\Omega \cap P)$ are finite.

Notice also that each Σ_α satisfies the "compact" version of regularity, namely that

$$C_0^{-1}\, s^3 \leq H^3(\Sigma_\alpha \cap B(y,s)) \leq C_0\, s^3\,, \tag{3.52}$$

for some constant C_0 (which does not depend on α) and all $y \in \Sigma_\alpha$ and $0 < s < \operatorname{diam} \Sigma_\alpha$. For Σ itself (3.52) is a consequence of its smoothness, while for the general Σ_α (3.52) reduces to the case of Σ because everything behaves properly under similarities. The same reasoning implies that

$$C_0^{-1}\, s^3 \leq H^3(\Omega_\alpha \cap P \cap B(y,s)) \leq C_0\, s^3\,, \tag{3.53}$$

for some constant C_0 (which does not depend on α) and all $y \in \Omega_\alpha$ and $0 < s < \operatorname{diam} \Omega_\alpha$.

To prove Lemma 3.45 let us deal first with M and M^j, and let us take $M = M^\infty$ as before. Let $x \in M^j$ and $r > 0$ be given, so that we want to show that

$$C^{-1}\, r^3 \leq H^3(M^j \cap B(x,r)) \leq C\, r^3\,, \tag{3.54}$$

for some constant C which does not depend on x, r, or j. Fix an $a > 0$ which is small enough so that Lemma 3.44 can be applied, and also so that

$$(3.55) \qquad 2a \operatorname{diam} \Omega < \rho \operatorname{dist}(\partial\Omega, \Omega_p) < \operatorname{dist}(\partial\Omega, \Omega_p),$$

for $p = 1, \ldots, n$. Lemma 3.44 provides us with five alternatives to consider separately. In case i) we have that $M^j \cap B(x,r) = P \cap B(x,r)$, and in particular that $x \in P$, and (3.54) follows.

Now suppose that ii) holds. In this case the upper bound in (3.54) is automatic, because of the corresponding bound for P and (3.49). The lower bound is slightly a nuisance, but it is not deep. Let us first check that

$$(3.56) \qquad H^3((P \backslash \Omega) \cap B(x,r)) \geq C^{-1} r^3,$$

for some constant C when $x \in P \backslash \Omega$. When $r \leq 10 \operatorname{diam} \Omega$ this follows from the smoothness of Ω, while for $r \geq 10 \operatorname{diam} \Omega$ it holds also for $x \in \Omega$, because $(P \backslash \Omega) \cap B(x,r)$ must then contain the intersection of P with a ball of radius $r/3$. These observations imply that in order to establish the lower bound in (3.54) we may as well assume that $x \in \Omega$ and $r \leq 10 \operatorname{diam} \Omega$. Since we are already assuming in ii) that $r \geq a \operatorname{diam} \Omega$, we have that r is bounded and bounded from below. Let m be the smallest positive integer such that $\rho^{m+1} < a/2$, and notice that m does not depend on x, r, or j. The main point in the rest of the argument is that the parts of M^j which correspond to levels above m do not really matter. If $j \leq m$, then either $x \in \Sigma_\alpha$ for an $\alpha \in \mathcal{S}_k$ with $k < j$ or $x \in \Omega_\alpha \cap P$ for some $\alpha \in \mathcal{S}_j$ (see (3.17)). In this case we can derive the desired lower bound in (3.54) from (3.52) or (3.53) (applied with $s =$ a small (but not too small) multiple of r). Thus we may assume that $j > m$, so that $\Sigma_\beta \subseteq M^j$ whenever $\beta \in \mathcal{S}_l$, $l \leq m$. If $x \in \Sigma_\alpha$ for an $\alpha \in \mathcal{S}_k$ with $k \leq m$, then we can again derive the lower bound in (3.54) from (3.52). We are left with the case where $x \in \Omega_\alpha$ for some $\alpha \in \mathcal{S}_{m+1}$, because of the nesting properties of the Ω_γ's (as in Lemma 3.10). Our choice of m implies that $\operatorname{diam} \Omega_\alpha < r/2$, and hence there is a point y in $B(x, r/2)$ which lies on (the boundary of) Σ_β, where $\beta \in \mathcal{S}_m$ is the parent of α. This permits us to reduce to (3.52) again since $H^3(M^j \cap B(x,r)) \geq H^3(\Sigma_\beta \cap B(y, r/2))$. This establishes (3.54) when ii) holds.

When iii) holds, our assumption (3.55) on a implies that $B(x,r)$ cannot touch the Ω_p's, $1 \leq p \leq n$, so that $M^j \cap B(x,r) = ((P \backslash \Omega) \cup$

$\Sigma) \cap B(x,r)$. In this case (3.54) is an immediate consequence of the smoothness of Σ and D.

If iv) holds, then we can reduce to $l = 0$ (and $\Omega_\alpha = \Omega$) using the self-similarity property (3.20), and this is just a special case of ii).

We are left with v). As before we can use (3.20) to reduce to $l = 0$ and $\Omega_\alpha = \Omega$. If $j = 0$ then $M^j = P$ and (3.54) is immediate. If $j = 1$ then $M^j = M^1$ is smooth and (3.54) is again clear. Thus we may assume that $j > 1$. The key observation now is that

$$B(x,r) \cap \Omega_\gamma = \varnothing, \qquad \text{for all } \gamma \in \mathcal{S}_2 . \tag{3.57}$$

To see this first notice that (3.55) implies that

$$2a \operatorname{diam} \Omega < \operatorname{dist}(\partial\Omega_\beta, \Omega_\gamma) \tag{3.58}$$

whenever $\beta \in \mathcal{S}_1$ and $\gamma \in \mathcal{S}_2$ is a child of β. (Do not forget that ρ is the dilation factor of the similarity ψ_β, and do not forget (3.8) either.) Our assumption v) implies that $B(x,r)$ intersects Y (defined just before (3.12)), and so if $B(x,r)$ intersected some Ω_γ, $\gamma \in \mathcal{S}_2$, then (3.58) would not be true, since we have also that $r < a \operatorname{diam} \Omega$ from v). Thus (3.57) is true, which implies that

$$B(x,r) \cap M^j = B(x,r) \cap \Big(\Sigma \cup \Big(\bigcup_{\beta \in \mathcal{S}_1} \Sigma_\beta\Big)\Big) \tag{3.59}$$

since we have also that $B(x,r) \subseteq \Omega$ by v). The regularity estimate (3.54) follows easily (from (3.52), for instance).

This proves that the M^j's are regular with a uniformly bounded constant. It remains to deal with $\widetilde{M}$. Let $x \in \widetilde{M}$ and $r > 0$ be given, and recall the definition of $\widetilde{M}$ and the related notation (just after (3.20)). If $B(x,r)$ is disjoint from all the B_k's, then $B(x,r) \cap \widetilde{M} = B(x,r) \cap P$, and we are in business. If $B(x,r) \subseteq 2\,B_k$ for some k, then $x \in A_k(M^k)$, $B(x,r) \cap \widetilde{M} = B(x,r) \cap A_k(M^k)$, and the necessary estimates on $H^3(B(x,r) \cap \widetilde{M})$ follow from the regularity of M^k. From this case we get that

$$H^3(\widetilde{M} \cap B_k) \le C\,(\operatorname{radius} B_k)^3 , \tag{3.60}$$

for all k. Now suppose that $B(x,r)$ intersects some B_k's but that it is not contained in any $2\,B_k$. Let K denote the set of k's such that $B(x,r)$ intersects B_k. Since $B(x,r) \not\subseteq 2\,B_k$ for any k we have that

$2r \geq \operatorname{radius} B_k$ when $k \in K$, and so $\cup_{k \in K} B_k \subseteq B(x, 5r)$. Because the B_k's are all disjoint and centered on P we obtain from this that

$$\sum_{k \in K} (\operatorname{radius} B_k)^3 \leq (5r)^3 \tag{3.61}$$

since $H^3((\cup_{k \in K} B_k) \cap P) \leq H^3(B(x, 5r) \cap P)$. Because (3.60) implies that

$$\begin{aligned} H^3(B(x,r) \cap \widetilde{M}) \leq H^3&(B(x,r) \cap P) \\ &+ \sum_{k \in K} H^3(B(x,r) \cap \widetilde{M} \cap B_k) \leq C r^3 , \end{aligned} \tag{3.62}$$

for some constant C, we get the upper bound that we need from (3.61). For the lower bound we observe that

$$H^3(B(x,r) \cap P \cap (2 B_k \backslash B_k)) \geq C^{-1} (\operatorname{radius} B_k)^3 \tag{3.63}$$

when $k \in K$, because $B(x, r)$ intersects such a B_k but is not contained in $2 B_k$. Hence

$$\begin{aligned} H^3(B(x,r) \cap \widetilde{M}) &\geq H^3\Big((B(x,r) \cap P) \backslash \Big(\bigcup_{k \in K} B_k \Big)\Big) \\ &\geq C^{-1} H^3(B(x,r) \cap P) , \end{aligned} \tag{3.64}$$

for some constant C. This uses also the disjointness of the $2 B_k$'s. Thus we get the lower bound on $H^3(B(x,r) \cap \widetilde{M})$ that we needed, so that $\widetilde{M}$ is also regular.

This completes the proof of Lemma 3.45. The same basic structure of the argument will be used repeatedly in this paper. That is, we shall need to prove various properties about balls on M or on an M^j, and we shall use Lemma 3.44 to distinguish some cases. Case i) will always by trivial, and iii) and v) will typically be easy because of the smoothness of M and the M^j's away from the singular set F. Cases ii) and iv) are generally about the same as each other and often require more specific information about the excellent package. To simplify these future arguments we collect first some information in the following lemma that will be common to many of them.

Lemma 3.65. *Consider M^j, $j \leq \infty$, with $j = \infty$ corresponding to M. There exist a small constant a and a large constant C_0 (depending on* *the excellent package but not on j) so that if $x \in M^j$ and $r > 0$ satisfy* i), iii), *or* v) *in Lemma* 3.44, *or if they satisfy* ii) *and also $j < \infty$, then there is an open set W in $\mathbb{R}^4$ with $B(x,r) \subseteq W \subseteq B(x, C_0 r)$ and a diffeomorphism λ from $B(0,r)$ onto W such that $\lambda(P \cap B(0,r)) = M^j \cap W$. In the cases* i), iii), *and* v) (*but not* ii)) *we can also take λ to be C_0-bilipschitz, i.e.,*

$$C_0^{-1}\,|y-z| \le |\lambda(y) - \lambda(z)| \le C_0\,|y-z|, \tag{3.66}$$

for all $y, z \in B(0,r)$.

In other words, M^j can be flattened out in a nice way near $B(x,r)$. Note the scale-invariance of the bilipschitz condition (3.66). This is very important when we are working with small balls near the singular set, because we shall not have uniform bounds on higher derivatives of λ.

Let us prove Lemma 3.65. Let a be small, to be chosen soon, and let $x \in M^j$ and $r > 0$ be given. If (x,r) satisfies i) in Lemma 3.44, then $B(x,r) \cap M^j = B(x,r) \cap P$, and we can simply take $W = B(x,r)$ and λ to be a translation. If ii) holds, so that $B(x,r) \cap \Omega \ne \varnothing$ and $r \ge a \operatorname{diam} \Omega$, then $B(x,Cr) \supseteq \Omega$ for $C = 1 + a^{-1}$. In this case $B(x,Cr) \cap M^j = h_j(B(x,Cr) \cap P)$, where h_j is the diffeomorphism (for $j < \infty$) promised in Lemma 3.21, and so we can simply take $W = h_j(B(x,Cr))$ and λ to be a translation of h_j. If iii) holds, then $B(x,r)$ must intersect $\partial\Omega$. If a is small enough, then $B(x,r) \cap M^j = B(x,r) \cap ((P \backslash \Omega) \cup \Sigma)$. The right hand side is a smooth submanifold with boundary, and $B(x,r)$ stays away from the boundary when a is sufficiently small. In this case M^j is a small smooth perturbation of a 3-plane inside $B(x,r)$, and it is easy to get the desired W and λ (with $C_0 = 2$, for instance). Before we deal with v) let us introduce some auxiliary notation and definitions. We may as well assume that $j \ge 2$, because the $j = 0$ case is trivial and we can simply use the diffeomorphism h_1 provided by Lemma 3.21 to get Lemma 3.65 when $j = 1$.

Given $\beta \in \mathcal{S}_k$, $k \le j$, set $\Sigma'_\beta = \Sigma_\beta$ when $k < j$ and $\Sigma'_\beta = \overline{\Omega}_\beta \cap P$ when $k = j$. This is just a convenient way to allow for the slightly exceptional case where $k = j < \infty$ without having make additional statements. Define N_α for α in some $\mathcal{S}_l$, $l \le j$, in the following manner. If $0 < l \le j-1$ set

$$N_\alpha = \Sigma'_\delta \cup \Sigma'_\alpha \cup \left(\bigcup \Sigma'_\beta\right), \tag{3.67}$$

where δ is the parent of α in $\mathcal{S}_{l-1}$ and the union is taken over the β's in $\mathcal{S}_{l+1}$ which are children of α. If $l = j$ define N_α in the same way except that we drop the Σ'_β's (since the children of α do not matter). If $l = 0$ then replace Σ'_δ in (3.67) with $P\backslash D$ but keep the rest. In each case N_α is a smooth embedded submanifold of $\mathbb{R}^4$ which contains Σ'_α, with some room to spare. Specifically, there is a constant $C > 0$ so that

$$\{z \in M^j : \operatorname{dist}(z, \Sigma'_\alpha) < C^{-1} \operatorname{diam} \Omega_\alpha\} \subseteq N_\alpha , \tag{3.68}$$

and in fact the left hand side does not get too close to the boundary of N_α. This constant C does not depend on α, j, or l; this is easy to check, using the usual self-similarity argument based on (3.20) to reduce to the (finitely many) $l = 0, 1$ cases.

Now suppose that v) in Lemma 3.44 is true, and let α, l be as in v). Remember from v) that either $l = j$ or $B(x,r)$ intersects Y_α. If a is small enough then we have that $B(x,r) \cap M^j = B(x,r) \cap N_\alpha$, because of (3.68), and $B(x,r)$ stays away from the boundary of N_α. In this case we can get the desired W and λ as soon as a is small enough, for the same reasons of smoothness as in case iii). We can even get uniform estimates (which do not depend on α or l) because the self-similarity provided by (3.20) permits us to reduce the problem to a finite number of models for the N_α's. This proves Lemma 3.65.

Next we deal with the property about bilipschitz balls in the conclusion of Theorem 1.12. This will occupy us for the remainder of the section, and the reader may wish to skip the long argument for the time being. Let us assume that

$$P\backslash D \text{ and } \overline{D}\backslash \bigcup D_i \text{ are connected.} \tag{3.69}$$

This assumption will hold in all of our examples.

Proposition 3.70. *Assuming that our initial package satisfies* (3.69), *there is a constant L so that if $E = M, \widetilde{M}$, or M^j for some $j < \infty$, then every pair of distinct points $p, q \in E$ is contained in a closed subset W of E such that W is L-bilipschitz equivalent to a closed Euclidean 3-ball. In particular there is a curve in E which joins them and which has length less or equal than $L^2 |p - q|$.*

This proposition (which is a variant of [Se5, Proposition 4.25]) is basically trivial, but it takes some space to do it with a moderate amount of care. Suppose for simplicity that we are working with M rather than M^j or $\widetilde{M}$. The first step is to connect p and q by a nice curve. "Nice" means in particular that the curve should avoid the singularities of M as much as possible. It may be necessary for the curve to do a fair amount of looping near p and q, because of the twisting of M, but we can understand this in a clear and simple way because we understand the singularities of M so well (by construction). The second step is to "fatten up" this curve to get a subset of M which is bilipschitz equivalent to a 3-ball. The amount of fattening is allowed to degenerate linearly as we move toward the endpoints of the curve, because we only want a bilipschitz ball (as opposed to a smooth ball). By choosing the curve to move away from the singularities of M as fast as possible we shall have that M is very flat near points on the curve (at the appropriate scale), and this will allow us to fatten the curve sufficiently.

The proof will show that we can choose W so that $W\backslash\{p,q\}$ is contained in the smooth part of E (*i.e.*, it is disjoint from F when $E = M$ and it does not contain the origin when $E = \widetilde{M}$), and that the bilipschitz equivalence between W and a ball can be taken to be smooth away from p and q. This observation will be useful in Section 10.

Much of the structure of an excellent package is unnecessary for Proposition 3.70, in the same way as discussed just after Definition 3.2. In particular the existence of these bilipschitz balls involves only "internal" properties of the M's, and not their relationship with the ambient space.

Let us now begin the proof of Proposition 3.70. Suppose first that $E = M^j$, where $j = \infty$ is allowed, and let $p, q \in M^j$ be given, $p \neq q$. We may as well assume that $j \geq 2$, since $M^0 = P$ and M^1 is bilipschitz equivalent to P (via the diffeomorphism h_1 from Lemma 3.21).

Given u, v in P and $\varepsilon > 0$ let $S(u,v)$ denote the segment which connects u to v and let $S_\varepsilon(u,v)$ be the set of points x in P such that $\operatorname{dist}(x, S(u,v)) \leq \varepsilon \operatorname{dist}(x, \{u,v\})$. Thus $S_\varepsilon(u,v)$ is the union of two truncated cones, one with vertex u, the other with vertex v. It is also bilipschitz equivalent to a Euclidean 3-ball, with a bilipschitz constant which depends only on ε. In order to produce a set W as in the proposition it is better to think of W as being bilipschitz equivalent to some $S_\varepsilon(u,v)$ rather than a round ball. Typically W will look like a twisted version of $S_\varepsilon(u,v)$ which spirals around at the ends. (To get the smoothness we want one should smooth out the spherical "corner" in the middle of $S_\varepsilon(u,v)$, but that is easy to do.)

It will be more convenient in the proof to use a slightly different analysis of cases than the one in Lemma 3.44. We begin with an easy one.

Lemma 3.71. *If $p, q \in M^j$ satisfy $p, q \in N_\alpha$ for some $\alpha \in \mathcal{S}_l$, $l \leq j$, then the conclusions of Proposition* 3.70 *hold for M^j with this choice of p and q. Here N_α is as in* (3.67).

To see this, forget about all this specific notation for a moment, and let N be a compact connected smooth 3-dimensional embedded submanifold of $\mathbb{R}^4$ (with boundary). Then any pair of points in N are contained in a closed subset of N which is bilipschitz equivalent to a closed Euclidean 3-ball. This is not hard to prove, and we leave it as an exercise. Lemma 3.71 is a special case of this statement, at least when $l > 0$, modulo the issue of getting uniform bounds on the bilipschitz constants. These uniform bounds come from the self-similarity property (3.20), which ensures that the N_α's are all similar to a finite collection of models. The argument for $l = 0$ is similar but modifications are needed because N_α is now unbounded (but equal to P outside Ω). In bounded regions this case behaves in the same way as the previous one, but in unbounded regions it behaves like the corresponding question for $P \backslash B$, where B is some ball. The point is simply that one must sometimes be careful to choose W so that it avoids the hole. Again the details are left to the reader.

The next lemma covers the most interesting case for Proposition 3.70.

Lemma 3.72. *The conclusions of Proposition* 3.70 *hold for M^j when $j = \infty$ and $p, q \in F$.*

Choose $\delta \in \mathcal{S}_m$ such that $p, q \in \Omega_\delta$ and m is as large as possible. Let α_l and β_l be the unique elements of $\mathcal{S}_l$ such that $p \in \Omega_{\alpha_l}$ and $q \in \Omega_{\beta_l}$. Thus $\alpha_m = \beta_m = \delta$, and α_{m+1} and β_{m+1} are both children of δ, but they are distinct children, since m is maximal. Choose points $p_l \in \partial\Sigma_{\alpha_l}$ and $q_l \in \partial\Sigma_{\beta_l}$ in an arbitrary manner. There is a constant C which depends only on our excellent package so that

$$C^{-1} \operatorname{diam} \Omega_{\alpha_l} \leq |p_l - p_{l+1}| \leq C \operatorname{diam} \Omega_{\alpha_l} , \tag{3.73}$$

$$C^{-1} \operatorname{diam} \Omega_{\beta_l} \leq |q_l - q_{l+1}| \leq C \operatorname{diam} \Omega_{\beta_l} , \tag{3.74}$$

$$(3.75) \qquad C^{-1}\operatorname{diam}\Omega_\delta \le |p_{m+1} - q_{m+1}| \le C\operatorname{diam}\Omega_\delta\ ,$$

$$(3.76) \qquad C^{-1}\operatorname{diam}\Omega_\delta \le |p - q| \le C\operatorname{diam}\Omega_\delta\ .$$

Note that $p_l \to p$ and $q_l \to q$ as $l \to \infty$. We shall build our bilipschitz ball W by combining a family of smooth tubes which connect the successive p_l's and q_l's.

Remember that $\operatorname{diam}\Omega_\alpha = \rho^l \operatorname{diam}\Omega$ when $\alpha \in S_l$. Thus $\{|p_l - p_{l+1}|\}_l$ and $\{|q_l - q_{l+1}|\}_l$ are approximately geometric sequences.

In order to prove Lemma 3.72 it suffices to find $\varepsilon > 0$ and a bilipschitz mapping $f : S_\varepsilon(z,w) \to M$ (with uniform choices of ε and the bilipschitz constant) such that $f(z) = p$ and $f(w) = q$. We shall define f in stages. To understand how f is constructed it is helpful to visualize the region $f(S_\varepsilon(z,w))$ that we shall have to construct. It will be a union of 3-dimensional tubes in M, where the tubes connect the successive p_l's and q_l's. These tubes will be diffeomorphic to rectangles and they will be neither too thin nor too close to F. To build these tubes we shall first choose some smooth Jordan arcs which connect the successive p_l's and q_l's, and the tubes will be little tubular neighborhoods of these arcs. The next sublemma deals with the existence of these Jordan arcs.

Sublemma 3.77. *Given any α in any S_l and any pair of points a, b in different components of the boundary of Σ_α, we can find an arc $\gamma \subseteq \Sigma_\alpha$ which connects a to b and has the following properties: if u and v are two points on γ, then the length of the arc in γ which connects u to v is bounded by $C|u - v|$; inside $B(a, C^{-1}\operatorname{diam}\Omega_\alpha)$ the curve γ agrees with the line segment in P which emanates from a, is orthogonal to $\partial\Sigma_\alpha$ at a, and goes inside Σ_α, and similarly for b (remember that the part of Σ_α near its boundary lies in P); if $u \in \gamma$, then $\operatorname{dist}(u, \partial\Sigma_\alpha) \ge C^{-1}\operatorname{dist}(u, \{a,b\})$ (so that γ does not get close to the boundary except near the endpoints); for each positive integer i the Euclidean norm of the i^{th} derivative of the arclength parameterization of γ is bounded by $C(i)(\operatorname{diam}\Omega_\alpha)^{1-i}$. (Notice that this estimate is scale-invariant.) These constants C and $C(i)$ depend only on our excellent package.*

This is an easy exercise. The main points are that we can reduce to the case where $l = 0$ and $\Sigma_\alpha = \Sigma$ using the self-similarity principle (3.20), and that Σ is a smooth connected (by (3.69)) compact manifold with boundary which agrees with P near its boundary. Thus we can certainly connect any pair of points in Σ with a curve, but by being a little bit careful we can choose the curve so that it avoids self-intersections and the boundary of Σ, we can make it smooth, etc.

Next we connect a sequence of curves as provided by Sublemma 3.77.

Sublemma 3.78. *There exist points z and w in P and a bilipschitz map $f : S(z,w) \longrightarrow M$ with uniformly bounded bilipschitz constant which satisfy the following properties: $f(z) = p$ and $f(w) = q$;*

$$C^{-1} \operatorname{dist}(t, \{z,w\}) \le \operatorname{dist}(f(t), F) \le C \operatorname{dist}(t, \{z,w\}), \tag{3.79}$$

for all $t \in S(z,w)$ (so that the image of f avoids the singular set F as much as possible); f is is smooth away from the endpoints z and w, and if $f^{(i)}$ denotes the i^{th} order derivative of f on $S(z,w)$, $f' = f^{(1)}$, then $|f'(t)| = 1$ and

$$|f^{(i)}(t)| \le C(i) \operatorname{dist}(t, \{z,w\})^{-i+1}, \tag{3.80}$$

for all $t \in S(z,w) \backslash \{z,w\}$ and $i \ge 1$. These constants depend only on our excellent package (and i in the case of $C(i)$).

Given $l > m$ let γ_l be the curve provided by Sublemma 3.77 for $\alpha = \alpha_l$, $a = p_l$, and $b = p_{l+1}$. Thus the length of γ_l is comparable to $|p_l - p_{l+1}|$, which is controlled by (3.73). Similarly, let $\widetilde{\gamma}_l$ be the curve which corresponds to β_l, q_l, and q_{l+1}, and let γ be the curve which corresponds to δ, p_{m+1}, and q_{m+1}. We shall choose f so that its image is precisely the union of all these curves together with p and q.

Notice first that the sum of the lengths of all the γ's is controlled by a convergent geometric series and hence is finite. In fact this total length is comparable to $|p - q|$, because of (3.76). Choose $z, w \in P$ so that $|z - w|$ equals the sum of the lengths of the γ's. (Except for this z and w can be arbitrary.) Set $f(z) = p$ and $f(w) = q$, and define f on $S(z,w)$ in such a way that it is really just the concatenation of the arclength parameterizations of the γ's, ordered in the obvious way. This mapping is smooth on $S(z,w) \backslash \{z,w\}$, because of the properties of the γ's in Sublemma 3.77.

Each arc γ_l corresponds to a segment I_l in $S(z,w)$, and the length of I_l and its distance to z are both comparable to $|p_l - p_{l+1}|$. Similar statements apply to the other γ's. Using these observations it is easy to derive (3.80) from the corresponding property in Sublemma 3.77, while (3.79) uses also the fact that

$$C^{-1} \operatorname{diam} \Omega_\alpha \leq \operatorname{dist}(x, F) \leq C \operatorname{diam} \Omega_\alpha \tag{3.81}$$

when $x \in \Sigma_\alpha$.

The Lipschitzness of f and the fact that $|f'| = 1$ on $S(z,w) \backslash \{z,w\}$ follow from the fact that we are using the arclength parameterizations of the γ's. It remains to show that f is bilipschitz, *i.e.*, that $|f(u) - f(v)| \geq C^{-1}|u - v|$ whenever $u, v \in S(z,w)$. If $f(u)$ and $f(v)$ lie in the same arc among the γ's then this follows immediately from the chord-arc property in Sublemma 3.77. If $f(u)$ and $f(v)$ lie in adjacent arcs among the γ's then this estimate can also be derived from the properties of these arcs given in Sublemma 3.77. If they do not lie in adjacent γ's -in particular if one of u or v equals z or w- then we use the fact that we know where the arcs lie in terms of the Σ_α's together with the various nesting and separation properties of the Y_α's and Ω_α's. For instance, we know that $\Sigma_{\alpha_l} \subseteq \Omega_{\alpha_{m+1}}$ and $\Sigma_{\beta_l} \subseteq \Omega_{\beta_{m+1}}$ when $l > m$, and we know that $\operatorname{dist}(\Omega_{\alpha_{m+1}}, \Omega_{\beta_{m+1}}) \geq C^{-1} \operatorname{diam} \Omega_\delta$ since $\alpha_{m+1} \neq \beta_{m+1}$ by the maximality of m. This implies that $|f(u) - f(v)| \geq C^{-1}|p - q| \geq C^{-1}|u - v|$ when $f(u)$ lies in one of the γ_l's and $f(v)$ lies in one of the $\widetilde{\gamma}_k$'s. Similarly, $\operatorname{dist}(\Sigma_{\alpha_l}, \Omega_{\alpha_k}) \geq C^{-1} \operatorname{diam} \Omega_{\alpha_l}$ when $k > l + 1$, as one can check by reducing to the case where $k = l + 2$ and $l = 0$, using the nesting properties in Lemma 3.10 and self-similarity (3.20). This implies that $|f(u) - f(v)| \geq C^{-1} \operatorname{diam} \Omega_{\alpha_l}$ if $f(u)$ lies in γ_l and $f(v)$ lies in γ_k with $k > l+1$. In this case we have that $|u - v| \leq C \operatorname{diam} \Omega_{\alpha_l}$ and hence $|f(u) - f(v)| \geq C^{-1}|u - v|$. The other cases can be handled in the same way, and we obtain that f is indeed bilipschitz. This proves Sublemma 3.78.

From now on we assume that f is defined on $S(z,w)$ as in Sublemma 3.78. We want to extend f to $S_\varepsilon(z,w)$ for suitable $\varepsilon > 0$. To do this we analyze the unit normals to M along the image of f, then we shall determine a first-order (linear) approximation to this extension of f, and then we build a true extension of f.

Sublemma 3.82. *There is a smooth function $\nu : S(z,w) \backslash \{z,w\} \longrightarrow \mathbb{R}^4$ such that $\nu(t)$ is normal to M at $f(t)$, $|\nu(t)| = 1$, and the derivatives of ν satisfy*

$$|\nu^{(i)}(t)| \leq C(i) \operatorname{dist}(t, \{z,w\})^{-i}, \tag{3.83}$$

for all $t \in S(z,w) \backslash \{z,w\}$ and $i \geq 1$. These constants depend only on i and our excellent package.

This is easy to see. The point is that there is a smooth unit normal on Σ simply because Σ is smooth, and there are smooth unit normals to the Σ_α's by self-similarity. We can get ν by piecing together these unit normals. The estimates (3.83) come from the estimates on f in Sublemma 3.78 and the self-similarity of the Σ_α's. This proves Sublemma 3.82.

From now on (in this proof of Lemma 3.72) we let ν be as in the Sublemma. Set $v_0 = (w-z)/|w-z|$ and $v_1 = (0,0,0,1)$ (= the standard unit normal to P). When we write $f'(t)$ we shall mean the derivative of f at t in the direction v_0.

Sublemma 3.84. *There is a smooth map ϕ from $S(z,w)\backslash\{z,w\}$ into rotations on $\mathbb{R}^4$ such that ϕ satisfies $\phi(t)v_0 = f'(t)$, $\phi(t)v_1 = \nu(t)$, and*

$$|\phi^{(i)}(t)| \leq C(i)\,\mathrm{dist}(t,\{z,w\})^{-i}, \tag{3.85}$$

for all $t \in S(z,w)\backslash\{z,w\}$ and $i \geq 0$, where $C(i)$ depends only on i and our excellent package.

Let us first resolve the "differentiated" version of this problem, in which we construct a family of antisymmetric linear mappings which will turn out to be $\phi'(t)\phi(t)^{-1}$. Given $t \in S(z,w)\backslash\{z,w\}$ define linear mappings $\psi_0(t), \psi_1(t) : \mathbb{R}^4 \longrightarrow \mathbb{R}^4$ by

$$\psi_0(t)\,\zeta = f''(t)\,\langle f'(t),\zeta\rangle + \nu'(t)\,\langle \nu(t),\zeta\rangle\,, \tag{3.86}$$

$$\begin{aligned}\psi_1(t)\,\zeta = {}& \nu(t)\,\langle \nu'(t), f'(t)\rangle\,\langle f'(t),\zeta\rangle \\ &+ f'(t)\,\langle f''(t),\nu(t)\rangle\,\langle \nu(t),\zeta\rangle\,.\end{aligned} \tag{3.87}$$

Note that $f'(t)$ and $\nu(t)$ are orthogonal to each other for all t, so that $\psi_0(t)f'(t) = f''(t)$ and $\psi_0(t)\nu(t) = \nu'(t)$. We also have that

$$\begin{aligned}\langle f''(t), f'(t)\rangle = 0\,, \qquad \langle \nu'(t),\nu(t)\rangle = 0\,,\\ \langle \nu'(t), f'(t)\rangle + \langle \nu(t), f''(t)\rangle = 0\,,\end{aligned} \tag{3.88}$$

because $\langle f'(t), f'(t)\rangle$, $\langle \nu(t),\nu(t)\rangle$, and $\langle \nu(t), f'(t)\rangle$ are all constant and hence have vanishing derivative. (Remember from Sublemma 3.78 that $|f'(t)| \equiv 1$.) This implies that $\psi_1(t)$ is antisymmetric and that $\psi_0^t(t)f'(t) = \psi_1(t)f'(t)$ and $\psi_0^t(t)\nu(t) = \psi_1(t)\nu(t)$ for all t, where ψ_0^t denotes the transpose of ψ_0. Set $\psi = \psi_0 - \psi_0^t + \psi_1$. Thus $\psi(t)$ is antisymmetric and $\psi(t)f'(t) = f''(t)$ and $\psi(t)\nu(t) = \nu'(t)$ for all t.

We want to define ϕ now by solving the differential equation $\phi' = \psi\phi$. Let u be the midpoint of $S(z,w)$, and take $\phi(u)$ to be any rotation which satisfies $\phi(u)v_0 = f'(u)$, $\phi(u)v_1 = \nu(u)$. With this choice made we can extend ϕ to all of $S(z,w)\backslash\{z,w\}$ by solving the aforementioned differential equation. Because ψ is always antisymmetric we get that every $\phi(t)$ is a rotation. Our choice of ψ also ensures that $\phi(t)v_0 = f'(t)$ and $\phi(t)v_1 = \nu(t)$ for all t. To get the bounds (3.85) we observe that

$$|\psi^{(i)}(t)| \leq C(i)\,\mathrm{dist}\,(t,\{z,w\})^{-i-1}\,, \tag{3.89}$$

for all $t \in S(z,w)\backslash\{z,w\}$ and $i \geq 0$, where $C(i)$ depends only on i and our excellent package. This follows from the definition of ψ and straightforward computation. The bounds for ϕ follow easily from this (and the fact that every $\phi(t)$ is a rotation, and hence has norm one). This proves Sublemma 3.84.

Sublemma 3.90. *There is a small number $\eta > 0$ so that f admits an extension to a smooth mapping (also called f) from $S_\eta(z,w)$ into M such that the differential of this extension at $t \in S(z,w)$ equals the restriction of $\phi(t)$ to P, where ϕ is as in Sublemma* 3.84, *and such that*

$$|\nabla^i f(x)| \leq C(i)\,\mathrm{dist}(x,\{z,w\})^{-i+1}\,, \tag{3.91}$$

$$C^{-1}\,\mathrm{dist}\,(x,\{z,w\}) \leq \mathrm{dist}\,(f(x),F) \leq C\,\mathrm{dist}\,(x,\{z,w\})\,, \tag{3.92}$$

for all $x \in S_\eta(z,w)\backslash\{z,w\}$ and $i \geq 1$. These constants η, C, and $C(i)$ depend only on our excellent package.

There are several ways to prove this, all of them boring. Here's one. Define an auxiliary extension g of f as a map from $S_\eta(z,w)$ into $\mathbb{R}^4$ by taking g to be affine in the directions perpendicular to $S(z,w)$, with the affine mapping chosen in the obvious way using ϕ from Sublemma 3.84. It is not hard to check that g satisfies the analogues of (3.91) and (3.92), using (3.80) and (3.85), at least if η is small enough (for the first inequality in (3.92)). In particular g is Lipschitz, with a uniform bound. This Lipschitz bound implies that the image of g stays very close to M compared to its distance to F when η is small. To make this precise let π denote the orthogonal projection of P onto the line through $S(z,w)$. Then

$$\begin{aligned} |g(x) - f(\pi(x))| &= |g(x) - g(\pi(x))| \\ &\leq C\,|x - \pi(x)| \\ &\leq C\,\eta\,\mathrm{dist}\,(x,\{z,w\}) \end{aligned} \tag{3.93}$$

when $x \in S_\eta(z,w)$. We are using here the fact that $\pi(x) \in S(z,w)$ when $x \in S_\eta(z,w)$ (and $\eta < 1$) and also the Lipschitzness of g.

We want to define $f(x)$ for $x \in S_\eta(z,w)$ by taking the point on M which is closest to $g(x)$. A priori this is dangerous, but here we need only deal with pieces of M which are far from the singular set compared to the length scale at which we are working, and there is no problem with the nearest-point-projections on such small pieces of M. More precisely, there is a small constant $a > 0$ depending only on our excellent package so that if $\xi \in M \backslash F$ and $B = B(\xi, a\,\mathrm{dist}\,(\xi, F))$, then the mapping Π on B which takes a point and sends it to the (unique) nearest point in M is well defined, smooth, and satisfies

$$\sup_B |\nabla^i \Pi| \leq C(i)\,\mathrm{dist}\,(\xi, F)^{-i+1}\,, \tag{3.94}$$

for $i \geq 1$. This is not hard to prove, using the smoothness of M away from F and the self-similarity property (3.20) to get the uniform estimates. (To do this from scratch one must compute a little to reduce to the inverse function theorem.)

Once we have these nearest-point-projections on small balls like B we can get f as in Sublemma 3.90 by projecting g onto M. This will only work when η is small enough, which ensures that the image points of g lie in balls like the ones just described, because of (3.93) and (3.79). It is not hard to check that this definition of f satisfies (3.91) and (3.92). Also, we defined g so that it had the correct differential along $S(z,w)$, and the nearest-point-projections onto M don't change that (because the image of the differential lies in the tangent space to M). This proves Sublemma 3.90.

Sublemma 3.95. *Let $f : S_\eta(z,w) \longrightarrow M$ be as in Sublemma* 3.90. *If $\varepsilon > 0$ is small enough, then the restriction of f to $S_\varepsilon(z,w)$ is bilipschitz, with ε and the bilipschitz constant depending only on our excellent package.*

The point here is that we chose ϕ carefully to make the extension spread out in the right way. We shall show first that f is bilipschitz on certain small balls using our choice of ϕ, and then we shall use the bilipschitzness of f on $S(z,w)$ to control the global behavior of f.

Given $t \in S(z,w)$ set $r = \mathrm{dist}\,(t, \{z,w\})$ and $B = B(t) = B(t, b\,r) \cap P$, where $b > 0$ is small and to be chosen. Let us show that if b is small enough then the restriction of f to B is bilipschitz with a bounded constant. This is an easy consequence of Sublemma 3.90. Let $A(x)$ be the affine function from P into $\mathbb{R}^4$ defined by $A(x) = f(t) + \phi(t)x$, where ϕ is as in Sublemma 3.84. This is the affine Taylor approximation to f at t, and it preserves distances, since $\phi(t)$ is a rotation. We can estimate $f - A$ (or rather its gradient) on B using (3.91), and we get

$$\begin{aligned}\sup_B |\nabla(f - A)| &\le C\, b\, r \sup_B |\nabla^2 f| \\ &\le C\, b\, r \sup_B \operatorname{dist}(x, \{z, w\})^{-1} \le C\, b\,. \end{aligned} \tag{3.96}$$

This uses also the fact that $\operatorname{dist}(x, \{z, w\}) \approx r$ when $x \in B$ (assuming $b \le 1/2$, say). If b is small enough then we conclude that

$$\frac{1}{2}|x - y| \le |f(x) - f(y)| \le 2\,|x - y|\,, \qquad \text{when } x, y \in B\,, \tag{3.97}$$

since A preserves distances. Choose such a b and let it be fixed from now on.

Now let x, y be any pair of points in $S_\varepsilon(z, w)$, and let us check the bilipschitz condition for them. We already know from (3.91) that f is Lipschitz on $S_\varepsilon(z, w)$, and so we need only concern ourselves with getting a lower bound for $|f(x) - f(y)|$. We may assume that x and y do not both belong to any ball $B(t), t \in S(z, w)$ as above, since we have (3.97) already. Thus $y \notin B(\pi(x))$, and this implies that $|x - y| \ge 10^{-2}\, b \operatorname{dist}(x, \{z, w\})$ if ε is small enough. We get the same inequality with the roles of x and y reversed, and so

$$|x - y| \ge 10^{-2}\, b \max\{\operatorname{dist}(x, \{z, w\}), \operatorname{dist}(y, \{z, w\})\}\,. \tag{3.98}$$

Since f is bilipschitz on $S(z, w)$ (Sublemma 3.78) we get that

$$|f(\pi(x)) - f(\pi(y))| \ge C^{-1}\,|\pi(x) - \pi(y)|\,. \tag{3.99}$$

This implies that

$$|f(\pi(x)) - f(\pi(y))| \ge C^{-1}\,|x - y|\,, \tag{3.100}$$

because (3.98) yields $|x - \pi(x)| + |y - \pi(y)| \le C\, b^{-1}\, \varepsilon\, |x - y| \le 10^{-1}\,|x - y|$ when ε is small enough. To get back to $|f(x) - f(y)|$ we observe that

$$|f(\pi(x)) - f(x)| \le C\, \varepsilon \operatorname{dist}(x, \{z, w\}) \le C\, b^{-1}\, \varepsilon\, |x - y| \tag{3.101}$$

because of the Lipschitzness of f, the fact that $x \in S_\varepsilon(z,w)$, and (3.98). We also have the same estimate for y instead of x, and we conclude from (3.100) that

$$|f(x) - f(y)| \geq C^{-1}\,|x - y| \tag{3.102}$$

if ε is small enough.

This completes the proof of Sublemma 3.95, and Lemma 3.72 follows.

Lemma 3.103. *The conclusions of Proposition* 3.70 *hold for* M^j *when* $p, q \in \Omega$, *whether or not* $j = \infty$.

Again choose δ in some $\mathcal{S}_m$ so that $p, q \in \Omega_\delta$ and m is as large as possible. We may as well assume that $m < j - 1$ and that one of p and q lies in Ω_γ for some $\gamma \in \mathcal{S}_{m+2}$, since otherwise we can apply Lemma 3.71. This implies that $|p - q| \geq C^{-1} \operatorname{diam} \Omega_\delta$ for some constant C (which depends only on the excellent package); if $|p - q|$ were small compared to $\operatorname{diam} \delta$, then we could use the fact that one of p and q lies in an Ω_γ, $\gamma \in \mathcal{S}_{m+2}$, to conclude that $p, q \in \Omega_\zeta$ for some child $\zeta \in \mathcal{S}_{m+1}$ of δ, in contradiction to the maximality of m.

Under these conditions we can apply the same basic construction as in the proof of Lemma 3.72. It can happen now that now one or both of p and q does not lie in F, or that $j < \infty$, so that the sequences of α_l's and β_l's might have to stop in a finite number of steps. In fact, we could have that one of p or q lies in Σ_δ, so that there would be no α_l's, or no β_l's. Thus it may be necessary to modify the construction at one or both "ends", but the estimates and underlying principles remain the same. One chooses points like the p_l's and the q_l's, one connects these points with nice curves in M^j (using Sublemma 3.77, extended slightly to include Σ'_α's when $j < \infty$), one combines the curves and parameterizes the union by a bilipschitz map as in Sublemma 3.78, one extends this mapping as in Sublemma 3.90 (using a good family of frames as in Sublemma 3.84), and then one checks bilipschitzness as in Sublemma 3.95. The details are left to the reader.

Lemmas 3.71, 3.72, and 3.103 cover all the possible locations of $p, q \in M^j$ except for $p \in \Omega_\alpha$ for some $\alpha \in \mathcal{S}_1$ and $q \in P \backslash \Omega$ (or the other way around). (See (3.17) and (3.19).) In this case we have that $|p - q| \geq C^{-1}$ for some constant C. Set $m = 0$ and let δ be the empty sequence in $\mathcal{S}_0$, so that $\Omega_\delta = \Omega$. We can choose $\alpha_l \in \mathcal{S}_l$ and $p_l \in \partial\Omega_{\alpha_l}$ as in the beginning of the proof of Lemma 3.72, except that these sequences may stop in a finite number of steps. We can use the same basic construction as in the proof of Lemma 3.72 to connect p to an auxiliary point on the boundary of $\partial\Omega$ through a nice sequence of curves, as in Sublemmas 3.77 and 3.78. We can connect q to this auxiliary point in a nice way from $P\backslash\Omega$, simply because $P \cap \Omega$ is a bounded smooth domain in P and $P\backslash\Omega$ is connected (by (3.69)). (This is similar to part of Lemma 3.71.) If we do these things in a non-stupid manner then we can fatten up this connection between p and q to get a bilipschitz 3-ball in M^j which contains them. There is a minor difference in this situation, however. If S_ϵ denotes the analogue of $S_\epsilon(z, w)$ (from the proof of Lemma 3.72) adapted to this situation, then the proportion of S_ϵ devoted to the connection from p to $\partial\Omega$ will be comparable in size to the diameter of Ω (a positive constant). If $|p - q|$ is very large, then this will be a small proportion of S_ϵ, much less than half, and most of S_ϵ will be devoted to the connection from q to $\partial\Omega$. This does not pose a serious problem, but it does mean that the bulge in the middle of S_ϵ should be placed away from Ω, where everything is flat. The details are again left to the reader.

This proves Proposition 3.70 in the case where $E = M$ or M^j. Suppose now that $E = \widetilde{M}$. Let $p, q \in \widetilde{M}$ be given, $p \neq q$. Let $\{B_k\}$ be the sequence of balls used in the definition of $\widetilde{M}$ (just after (3.20)), and let $\widetilde{M}^k$ be the affine image of M^k with the interestesting part squeezed into B_k, as in the definition of $\widetilde{M}$. If $p, q \in (3/2)\, B_k$ for some k, then we can use the previous result for M^k to obtain that p and q are contained in a set $W' \subseteq \widetilde{M}^k$ which is bilipschitz equivalent to a Euclidean 3-ball. If one is careful about the previous construction one can choose W' so that $W' \subseteq 2B_k$, but one can also simply force this to happen, in the following manner. We can choose W' so that $W' \subseteq CB_k$ for some uniformly bounded constant C; if this is not true to begin with, it simply means that W' is unnecessarily large, and we can replace it with a smaller subset. Let Ψ be a bilipschitz map from CB_k into $2B_k$ which equals the identity on $(3/2)B_k$ and maps $(CB_k\backslash(3/2)B_k)\cap P$ into $2B_k \cap P$. If we set $W = \Psi(W')$, then $W \subseteq \widetilde{M}^k \cap 2B_k$ and hence $W \subseteq \widetilde{M}$. We also have that W is uniformly bilipschitz equivalent to a Euclidean 3-ball and contains p and q. Thus the case where $p, q \in (3/2)\, B_k$ for some k can be reduced to the previous results.

If neither p nor q lie in any B_k, then it is not hard to show directly that they are contained in a subset W of $P \setminus (\cup B_k) \subseteq \widetilde{M}$ which is bilipschitz equivalent to a Euclidean 3-ball with a uniform bound. It is easier to think of W as being bilipschitz equivalent to a set like $S_\epsilon(u, v)$

as described at the beginning of the proof of Proposition 3.70 rather than a standard ball, so that it is easier to visualize the way the ends are placed into $P \setminus (\cup B_k)$. It is also convenient to choose the B_k's in the rather specific manner described in the paragraph after (3.20).

We are left with the case where p lies in some B_k and q lies outside $(3/2)\, B_k$. Suppose first that q does not lie in any other B_l. We can use the same kind of argument as in Lemmas 3.72 and 3.103 to connect p to an auxiliary point in $\partial((3/2)\, B_k)$ in a nice way (as in Sublemma 3.78), using a family of smooth arcs. We can then connect from there to q inside $P \setminus (\cup B_l)$ by a more direct construction, since the structure of $P \setminus (\cup B_l)$ is so simple. These two connections can be combined and then filled out to get a bilipschitz 3-ball in $\widetilde{M}$ which connects p and q. This combination and filling-out will be realized concretely as a bilipschitz map from a set of the form $S_\varepsilon(z,w)$ into $\widetilde{M}$, with the restriction of this map to one end of $S_\varepsilon(z,w)$ providing the connection from p to $\partial((3/2)\, B_k)$, and the rest corresponding to the connection from there to q. The diameter of $S_\varepsilon(z,w)$ should be comparable to $|p-q|$, while the diameter of the piece at the end corresponding to the connection from p to $\partial((3/2)\, B_k)$ should be comparable to the radius of B_k. These sizes are not inconsistent with each other, because $|p-q|$ is at least one-half the radius of B_k. It may well be that $|p-q|$ is much larger than the radius of B_k, in which case we should be slightly careful to put the middle of $S_\varepsilon(z,w)$ far away from the B_l's. This type of detail is awkward but not at all deep.

If instead q does lie in some B_l, then we use the method of Lemmas 3.72 and 3.103 to connect p to an auxiliary point in $\partial((3/2)\, B_k)$ and q to an auxiliary point in $\partial((3/2)\, B_l)$ by well-behaved curves (as in Sublemma 3.78). We can connect these auxiliary points in a nice way inside $P \setminus (\cup B_m)$ by a direct construction, and these three connections can be combined and filled out in such a way as to get a bilipschitz 3-ball in $\widetilde{M}$ which connects p and q. If we think of the combination of these connections as being represented by a bilipschitz map from a set of the form $S_\varepsilon(z,w)$, then the connections from p to $\partial((3/2)\, B_k)$ and from q to $\partial((3/2)\, B_l)$ will correspond to pieces of $S_\varepsilon(z,w)$ at the two ends of $S_\varepsilon(z,w)$. These two pieces will have sizes comparable to the radii of B_k and B_l, respectively, while the diameter of $S_\varepsilon(z,w)$ should be about the same as $|p-q|$. These sizes are consistent with each other, because $|p-q|$ is at least one-half the radius of each of B_k and B_l. As usual, if $|p-q|$ is much larger than the radii of B_k and B_l, then we have to be careful to map the large middle of $S_\varepsilon(z,w)$ away from the various B_i's.

This completes the proof of Proposition 3.70, modulo various packages of details left to the reader. In all cases the argument can be understood in terms of connecting sequences of smooth curves together and then parameterizing these curves and filling out these parameterizations to get a bilipschitz mapping defined on a set of the form $S_\varepsilon(z,w)$. There are some minor variations among the various cases -whether the sequence of constituent curves is finite or infinite, whether we work only in bounded regions like the Ω_α's or we have to go wandering outside to the large flat regions of the M's, or whether we have to worry about where the "bulge" in the middle of $S_\varepsilon(z,w)$ should be sent- but the actual constructions are simpler than their gory detailed descriptions.

Note that if we were only interested in the last part of Proposition 3.70 (about connecting p and q by a curve in E whose length is bounded by a constant times $|p-q|$) then the preceding proof would simplify substantially. For instance, under the assumptions of Lemma 3.72 we would need much less than Sublemma 3.78.

Corollary 3.104. *If our initial package satisfies* (3.69) *and* $E = M, \widetilde{M}$, *or* M^j, $j < \infty$, *then* E *is linearly locally connected, with constants that do not depend on* j *in the latter case. This means that there is a constant* C *so that for each* $x \in E$ *and* $t > 0$ *we have that any two points in* $E \cap B(x,t)$ *lie in the same component of* $E \cap B(x, Ct)$, *and any two points in* $E \backslash B(x,t)$ *lie in the same component of* $E \backslash B(x, C^{-1}t)$.

This is an easy consequence of Proposition 3.70. One could also prove it more directly, in the same way that it is much easier to connect pairs of points in the M's by curves which are not too long than it is to get the bilipschitz balls in Proposition 3.70. If one were to try to give a more direct proof of the corollary, then Lemma 3.44 would be rather convenient for proving the second part of the linear local connectedness condition, but it is better to apply it to $B(x,(1+a^{-1})^{-1}t)$, where a is as in Lemma 3.65, than to $B(x,t)$ itself.

The main result of [HK] provides a good reason to care about linear local connectedness. This result states that if a 3-dimensional regular set E is linearly locally connected, and if E admits a homeomorphic parameterization by $\mathbb{R}^3$ by a mapping which satisfies the pointwise definition of quasiconformality, then in fact this mapping must be quasisymmetric. Thus the sets promised in Theorem 1.12 cannot even be quasiconformally equivalent to $\mathbb{R}^3$. In this particular situation it is a little ridiculous to cite the theorem in [HK], because the same result could be obtained more directly using standard results about quasiconformal mappings between domains in $\mathbb{R}^n$ and the fact that the Σ_α's are all smooth and similar to each other. (One could also use Proposition 3.70.)

In the next three sections we consider specific examples of excellent packages.

4. The Whitehead example.

We shall continue to use the definitions and notation of the preceding section here.

Let D be a smooth solid torus in $\mathbb{R}^3$, and let D_1 be another smooth solid torus whose closure is contained in D and which is clasped inside of D in the usual manner for generating the Whitehead continuum, *i.e.*, in the manner shown in [D, p. 68, Figure 9-7] (see also [K, p. 81ff]). Note that this clasping prevents D_1 from being isotopic to a standard (small) solid torus in D, but D_1 is homotopically trivial in D. In other words, we can deform D_1 inside D into something small, but D_1 has to cross itself along the way. Let ϕ_1 be a smooth diffeomorphism which maps a neighborhood of $\overline{D}$ onto $\overline{D_1}$ and sends D onto D_1. (One can think of simply bending D around in $\mathbb{R}^3$ to get D_1.) This gives us an initial package, and note that it satisfies (3.69).

To get an excellent package we take Ω and ω_1 to be (4-dimensional) solid tori in $\mathbb{R}^4$ which satisfy (3.3). The property of being a "4-dimensional solid torus" here means in particular that Ω is diffeomorphic to $D \times (-1,1)$ and similarly for ω. It is helpful to take ω_1 to be much flatter than Ω in the x_4 direction, to provide plenty of room to move around. We take ρ to be small and positive but otherwise at our disposal, and we choose ψ_1 to be a combination of a translation (by an element of P) and dilation by ρ such that $\Omega_1 = \psi_1(\Omega)$ and its closure lie in a ball in Ω. To build θ we use the fact that (unlike D_1 and D) ω_1 is not really clasped inside Ω, because of the freedom of movement provided by the extra dimension. We can lift up one end of the clasping part of ω_1 (in the x_4 direction) while leaving the other clasping end alone, and then we can bring it around in Ω and shrink it until ω_1 is deformed into Ω_1. One can do this process in such a way that $\omega_1 \cap P = D_1$ is deformed to $\Omega_1 \cap P$, and one can even extend this motion to all of $\mathbb{R}^4$ without ever moving points outside of Ω. These observations are more easily made rigorous using the following well-known "isotopy extension" result.

Lemma 4.1. *Let O be a bounded open set in $\mathbb{R}^n$ and let K be a compact subset of O. Suppose that $g(x,t)$ is a smooth $\mathbb{R}^n$-valued mapping defined for x in a neighborhood of K and t in $[0,1]$ such that $g(x,0) = x$ for all x and $g(\cdot,t)$ is a diffeomorphism onto its image for each t. Assume also that $g(K \times [0,1]) \subseteq O$. Then there exists a smooth mapping $G : \mathbb{R}^n \times [0,1] \longrightarrow \mathbb{R}^n$ such that $g(x,t) = G(x,t)$ when $x \in K$, $G(\cdot,t)$ is a diffeomorphism on $\mathbb{R}^n$ for each t, $G(x,t) = x$ when $x \in \mathbb{R}^n \setminus O$, and $G(x,0) = x$ for all x.*

To see this we begin by observing that there is a smooth time-dependent vector field $V(y,t)$ on $\mathbb{R}^n$ such that $V(g(x,t),t) = \partial g(x,t)/\partial t$ for all $x \in K$ and $t \in [0,1]$ and such that $V(y,t)$ vanishes whenever x lies outside any prescribed neighborhood of $g(K \times [0,1])$. Indeed, we start by defining V on pairs of the form $(g(x,t),t)$ for x near K and t in $[0,1]$, and we can extend V to all $(x,t) \in \mathbb{R}^n \times [0,1]$ using a cut-off function. Let $G(x,t)$ be the solution of the ordinary differential equation $\partial G(x,t)/\partial t = V(G(x,t),t)$ for all $(x,t) \in \mathbb{R}^n \times [0,1]$, with the initial condition $G(x,0) = x$ for all x. The uniqueness theorem for ordinary differential equations and standard facts imply that G has the required properties. This proves Lemma 4.1.

Using Lemma 4.1 and the deformation process described above one can build a diffeomorphism θ on $\mathbb{R}^4$ which satisfies (3.5), $\theta(\omega_1) = \Omega_1$, and also $\theta(D_1) = \Omega_1 \cap P$. If one is careful one can get that $\theta = \psi_1 \circ \phi_1^{-1}$ on a neighborhood of $\overline{D}_1$ in P, but it is simpler to just redefine ϕ_1 so that this is true.

The conclusion of all this is that we get an excellent package associated to our initial package, and we can take ρ to be as small as we want. The construction of Section 3 produces sets M^j and $\widetilde{M}$.

Theorem 4.2. *$\widetilde{M}$ satisfies (†) but not (∗). The M^j's for $j < \infty$ satisfy (†) with constants which remain bounded, but they do not satisfy (∗) with bounded constants.*

This result together with Lemma 3.45 and Proposition 3.70 imply Theorem 1.12.a).

Of course the M^j's for $j < \infty$ and $\widetilde{M}$ are homeomorphic to $\mathbb{R}^3$ (and hence topological manifolds) by Lemma 3.21.

Let us first prove the bounded contractability conditions for the M^j's. Let a be chosen as in Lemma 3.65, and let $0 \leq j < \infty$, $x \in M^j$, and $r > 0$ be given. We want to check that we can contract $B(x,r) \cap M^j$ to a point inside $B(x, Cr) \cap M^j$ for some constant C which does not depend on x, r, or j. Lemma 3.44 permits us to consider separately the five cases listed there. All but case iv) are covered by Lemma 3.65, and so we suppose that iv) holds.

Let Ω_δ be the parent of Ω_α, and let us show that

(4.3) $\quad \Omega_\alpha \cap M^j$ can be contracted to a point inside $\Omega_\delta \cap M^j$.

To prove this we may as well assume that $l = 1$, so that $\Omega_\delta = \Omega$, since otherwise we can use ψ_δ to pull everything back to Ω. (This will change M^j to M^{j-l}, as in (3.20), but that is okay.) Using the homeomorphism h_j from Lemma 3.21 we can reduce further to the problem of contracting D_1 to a point in D. This we can do, because of the specific choice of the Whitehead initial package. (It is not true for arbitrary initial packages.)

This proves that the M^j's satisfy (†) uniformly, and so we consider now $\widetilde{M}$. Recall its definition and related notation from the paragraph after (3.20). Part of this argument also works in general, and so we split it off as a separate lemma.

Lemma 4.4. *Suppose that we have started with any excellent package and constructed $\widetilde{M}$ as in Section 3. Let $b > 0$, $x \in \widetilde{M}$ and $r > 0$ be given, and assume that for each k we have either $B(x,r) \cap B_k = \varnothing$ or $r > b \operatorname{radius} B_k$, where $\{B_k\}$ is the sequence of balls in the definition of $\widetilde{M}$. Then there is a relatively open set $U \subseteq \widetilde{M}$ which is homeomorphic to a 3-ball and satisfies $B(x,r) \cap \widetilde{M} \subseteq U \subseteq B(x, Cr) \cap \widetilde{M}$, where C depends on b but not on x or r.*

Let x, r be given as in the lemma. If $B(x,r) \cap B_k = \varnothing$ for all k then $B(x,r) \cap \widetilde{M} = B(x,r) \cap P$ and there is nothing to do. Suppose that $B(x,r) \cap B_k \neq \varnothing$ for some k and choose such a k which is as small as possible. Let B'_k be as in the paragraph defining $\widetilde{M}$ (just after (3.20)), so that $\operatorname{radius} B'_k \leq 100 \operatorname{radius} B_k$, $B_l \subset B'_k$ when $l \geq k$, and $B_l \cap B'_k = \varnothing$ when $l < k$. Let $\widetilde{h}$ be as in Lemma 3.21, and set $U = \widetilde{h}((B(x,r) \cup B'_k) \cap P)$. This is a topological 3-ball, because the union of two intersecting (standard) 3-balls is a topological 3-ball. We also have that

$$U = \Big(\Big((B(x,r) \cup B'_k) \setminus \Big(\bigcup_{l \geq k} B_l\Big)\Big) \cap P\Big) \bigcup \Big(\Big(\bigcup_{l \geq k} B_l\Big) \cap \widetilde{M}\Big) \tag{4.5}$$
$$\subseteq B(x,r) \cup B'_k\,,$$

because of the properties of $\widetilde{h}$ in Lemma 3.21. The hypothesis of Lemma 4.4 guarantees that $\operatorname{radius} B_k \leq b^{-1}r$, which implies in turn that $\operatorname{diam} U \leq Cr$ for a suitable constant C. It is easy to check also that $B(x,r) \cap \widetilde{M} \subseteq U$ because of the properties of $\widetilde{h}$ in Lemma 3.21. This proves Lemma 4.4.

Let us now show that $\widetilde{M}$ satisfies (†). We already know that it is a topological manifold, and so we need only check the contractability condition. Let $b > 0$ be small, which we get to choose. Let $x \in \widetilde{M}$ and $r > 0$ be given. Because of Lemma 4.4 we may as well assume that there is a k such that $B(x,r)$ intersects B_k and $r \leq b \operatorname{radius} B_k$. These conditions imply in particular that $B(x,r) \subseteq 2\,B_k$. Hence $B(x,r) \cap \widetilde{M} = B(x,r) \cap A_k(M^k)$, and since we already know that (†) holds uniformly for the M^j's, and since the similarity A_k does not affect the (†) property (or its constant) we get that there is a constant C' such that $B(x,r) \cap A_k(M^k)$ can be contracted to a point inside $B(x,C'r) \cap A_k(M^k)$. If b is small enough (depending on C'), then $B(x,C'r) \subseteq 2\,B_k$ too, so that $B(x,C'r) \cap A_k(M^k) = B(x,C'r) \cap \widetilde{M}$. Thus $B(x,r) \cap \widetilde{M}$ can be contracted to a point inside $B(x,C'r) \cap \widetilde{M}$, which is what we wanted. This proves that $\widetilde{M}$ also satisfies (†).

(Incidentally, the choice of b in the preceding argument is a little bit stupid, in the sense that if one looks carefully one sees that a far less small choice of b would work fine. However, this additional complication is not needed for the proof.)

It remains to show that (∗) is bad. This will be derived from a famous property of the Whitehead continuum. Since $n = 1$ we have that each $\mathcal{S}_l$ has only one element, and so if the defining sequence $\{C_l\}$ is constructed from our initial package as described after Definition 2.3, then each C_l has only one component, namely D_α for $\alpha =$ the unique element of $\mathcal{S}_l$. Set $W = \cap_l C_l$. This is the Whitehead continuum, and it is the only nondegenerate element of the decomposition G associated to $\{C_l\}$ as discussed just after Definition 2.2. We shall sometimes find it convenient to view W as a subset of the 3-sphere $\mathbf{S}^3$.

Recall that a compact set $K \subseteq \mathbb{R}^3$ is said to be cellular if for each open set $V \supseteq K$ there is a topological 3-ball U such that $K \subseteq U \subseteq V$.

(See [D, p. 35].) A decomposition G of $\mathbb{R}^3$ is said to be cellular if each element of G is a cellular subset of $\mathbb{R}^3$. (See [D, p. 36, Corollary 2A].) A topological space X is said to be simply connected at ∞ if for each compact set $K \subseteq X$ there is a compact set $L \subseteq X$ with $L \supseteq K$ such that every loop in $X \backslash L$ can be contracted to a point in $X \setminus K$. (See [K, p. 83].) Note that $\mathbf{S}^3 \setminus K$ is simply-connected at ∞ if $K \subseteq \mathbb{R}^3$ is cellular.

Proposition 4.6. *W is not a cellular subset of $\mathbb{R}^3$, and in fact $\mathbf{S}^3 \backslash W$ is not simply connected at ∞.*

The first statement is a reformulation of [D, p. 76, Proposition 9] (see also the top of p. 69 of [D]). The second statement is discussed on [K, p. 82-83].

Lemma 4.7. *If $\widetilde{M}$ satisfies $(*)$, or if the M^j's satisfy $(*)$ with a uniformly bounded constant, then W is cellular.*

Suppose first that the M^j's satisfy $(*)$ with a uniformly bounded constant C_0. Let l be given, and let j be larger than l and at our disposal. Let α and β be the unique elements of $\mathcal{S}_l$ and $\mathcal{S}_j$, respectively, and fix a point x in $\Omega_\beta \cap M^j$. Our assumption on the M^j's implies that there is a topological 3-ball U such that

$$\Omega_\beta \cap M^j \subseteq U \subseteq B(x, C_0 \operatorname{diam} \Omega_\beta) \cap M^j . \tag{4.8}$$

If $j - l$ is sufficiently large, depending on C_0, then $B(x, C_0 \operatorname{diam} \Omega_\beta) \subseteq \Omega_\alpha$, so that $U \subseteq \Omega_\alpha \cap M^j$. We can use the homeomorphism h_j in Lemma 3.21 to bring U back to $\mathbb{R}^3$, and we get a topological 3-ball $V = h_j^{-1}(U)$ such that $D_\beta \subseteq V \subseteq D_\alpha$. Since l was arbitrary and we obtain that W is cellular.

The same argument works if we assume that $\widetilde{M}$ satisfies $(*)$. The point is that $\widetilde{M}$ contains a copy of $\Omega \cap M^j$ for each j, and we never left Ω in the preceding argument. This proves Lemma 4.7.

Proposition 4.6 and Lemma 4.7 imply that $\widetilde{M}$ does not satisfy $(*)$, and that the M^j's do not satisfy $(*)$ with a uniformly bounded constant. This proves Theorem 4.2.

REMARK 4.9. Lemma 4.7 works for any excellent package. That is, if we start with an excellent package and construct M^j and $\widetilde{M}$ as in

Section 3, and if either $\widetilde{M}$ satisfies $(*)$ or the M^j's satisfy $(*)$ with a uniformly bounded constant, then the decomposition G associated to the initial package (as in Section 2) is cellular, *i.e.*, each element of G is a cellular subset of $\mathbb{R}^3$. This can be proved by exactly the same argument as in Lemma 4.7, with minor additional complications when the number n of D_i's is greater than 1.

REMARK 4.10. The set M corresponding to this excellent package is homeomorphic to $\mathbb{R}^3$ with the Whitehead continuum contracted to a point. This famous non-manifold has the property that its product with $\mathbb{R}$ is homeomorphic to $\mathbb{R}^4$. (See [K, p. 87, Theorem 1] and [D, p. 83, Theorem 3].) This space $M \times \mathbb{R}$ arises in one description of the simplest Casson handles, as in p. 86 and the bottom half of p. 83 of [K]. These Casson handles have been shown in [Bi] to be exotic in a certain sense (with respect to their smooth structures). One can imagine that the quasiconformal version of this exoticness is also true, because of [DoS].

5. Bing's dogbone space.

We shall continue to use freely the notations and definitions of Section 3.

To prove Theorem 1.12.b) we shall use the construction of Bing's dogbone space, which is given in [B3] and as [D, p. 64, Example 4]. For this we define an intial package by taking D to be a smooth solid two-handled torus which is embedded in $\mathbb{R}^3$ in the standard way (no funny business with the two handles). We also take $n = 4$ and the D_i's to be solid two-handled tori in D arranged in the manner pictured in [B3, p. 486, Figure 1] and in [D, p. 65, Figure 9-4] (and with diffeomorphisms ϕ_i as in Definition 2.3 chosen in a non-stupid manner). This arrangement satisfies the requirements of Definition 2.3 and (3.69), and it has the additional feature that each D_i is (individually) embedded in D in a topologically standard (unlinked) manner. In particular,

(5.1) for each $i = 1, 2, 3, 4$ there is a topological 3-ball U_i such that $\overline{D}_i \subseteq U_i \subseteq D$.

This is obvious from the pictures; the point is that each D_i is, as an individual domain, not linking with itself inside D in any way. (As a group the D_i's are definitely linked, not homologically, but in the sense that they cannot be disentangled by an isotopy on D. This is related to the fact that the U_i that works for one i will have to intersect the other D_j's.) Condition (5.1) implies that the decomposition G associated to this initial package (as discussed in Section 2) is cellular (in the sense described in the preceding section). However, the quotient is not a manifold.

Theorem 5.2. (Bing, [B3, p. 498, Theorem 13]) *$\mathbb{R}^3/G$ is not a manifold.*

For the proof of Theorem 1.12.b) (5.1) and Theorem 5.2 are the only properties that we shall need. Note that there are other examples in Section 9 (beginning on p. 61) of [D] of cellular decompositions of $\mathbb{R}^3$ which are obtained from initial packages and whose quotients are not topological manifolds. Actually, [D] does not address directly the issue of manifold quotients, only the stronger property of "shrinkability" of the decomposition, but a theorem of Armentrout [Ar] implies that the two properties are equivalent for (cellular) decompositions (as remarked near the bottom of p. 22 of [D]). The bottom line is that there are other examples that we could use to get Theorem 1.12.b).

We can also build an excellent package for this initial package. Let Ω and ω_i, $1 \leq i \leq 4$, be solid versions of D and the D_i's in $\mathbb{R}^4$ which satisfy (3.3). The phrase "solid version" means that Ω should be diffeomorphic to $D \times (-1, 1)$, and similarly for the ω_i's. It is also a good idea to require that the ω_i's lie in a thin slab $\{x \in \mathbb{R}^4 : |x_4| < \varepsilon\}$, while Ω should contain a much fatter slab around most of D (and near the ω_i's in particular). This allows us to translate an ω_i "up" (in the positive x_4 direction) away from the other ω's and to move it around up there without getting too close to the others. This ensures that the ω_i's are not linked in $\mathbb{R}^4$ in any manner. Let ψ_i, $1 \leq i \leq 4$, be similarities on $\mathbb{R}^4$ with a common dilation factor ρ which map P to itself and which send Ω to domains Ω_i with disjoint closures in Ω. It is convenient to require also that the Ω_i's stay away from all the ω_i's. We do not care too much about the specific value of ρ but it should be reasonably small and we may take it to be as small as we want. The main point now is that we can build a θ as in Definition 3.2. To do this we take ω_1, we lift it up in the positive x_4 direction away from the slab $\{x \in \mathbb{R}^4 : |x_4| < \varepsilon\}$, we shrink it and slide it around until it is the same as Ω_1 but translated up a bit, and then we set it down onto Ω_1. In this whole process we take care not to touch the ω_i's or the Ω_i's for $i \neq 1$, and also to remain inside Ω the whole time. We then repeat the process for ω_2, ω_3, ω_4. Lemma 4.1 allows us to extend these deformations to all of $\mathbb{R}^4$ in such a way that points in or near the complement of Ω are not moved, and points in the ω_i's and the Ω_i's move only when they are supposed to. As a result we get a diffeomorphism θ as in Definition 3.2. (To be honest, it is simpler to first build θ and then choose the ϕ_i's so that (3.6) holds, rather than adjusting θ to fit the ϕ_i's.)

Thus we have an excellent package associated to our initial package, where ρ can be made as small as we want, and so we can use the construction of Section 3 to produce sets M, M^j, and $\widetilde{M}$.

Theorem 5.3. *$\widetilde{M}$ satisfies $(*)$ but not $(**)$. The M^j's, $0 \leq j < \infty$, satisfy $(*)$ with a uniformly bounded constant, but they do not satisfy $(**)$ with uniform choices of the constant and modulus of continuity.*

Let us first show that the M^j's satisfy $(*)$ with a uniformly bounded constant. Let a be as in Lemma 3.65, and let j, $x \in M^j$, and $r > 0$ be given. Lemma 3.44 allows us to consider separately the cases i)-v) listed there, but Lemma 3.65 implies that we need only consider iv). Let α be as in iv) in Lemma 3.44, and let $\delta \in \mathcal{S}_{l-1}$ be its parent. It suffices to show that there exists a relatively open set U_α in M^j which is homeomorphic to a 3-ball and satisfies

$$\Omega_\alpha \cap M^j \subseteq U_\alpha \subseteq \Omega_\delta \cap M^j . \tag{5.4}$$

For the usual self-similarity reasons (*i.e.*, (3.20), but with α replaced by δ) we can reduce this to the case where $l = 1$ (and j is replaced by $j - l + 1$). This case reduces to (5.1) because of Lemma 3.21. (Take U_α to be the image under h_{j-l+1} of the appropriate U_i.) This proves (5.4) and the fact that the M^j's satisfy $(*)$ with a bounded constant.

To show that $\widetilde{M}$ satisfies $(*)$, one uses Lemma 4.4 to reduce to the previous fact for the M^j's. The argument is practically identical to the corresponding step in Section 4 (just after (4.5)), and we do not repeat it.

Next we show that $(**)$ is bad for this excellent package.

Lemma 5.5. *If $\widetilde{M}$ satisfies $(**)$, or if the M^j's satisfy $(**)$ with uniform choices of the constant and modulus of continuity, then M is a topological manifold. (This works for any excellent package, and not just the particular ones considered in this section.)*

This is pretty straightforward. We have very precise control on the convergence of the M^j's to M which implies convergence in the Hausdorff topology in particular. A uniform version of (**) would force M to satisfy (**) also, and of course (**) certainly implies that M is a topological manifold. If $\widetilde{M}$ satisfies (**), then we use the fact that $\widetilde{M}$ contains a translation and dilation of the most interesting part of each M^j. Since (**) is preserved by similarities we can use the same argument to conclude that M is a topological manifold.

Theorem 5.3 now follows from Lemma 5.5, Lemma 3.21 (which states that M is homeomorphic to $\mathbb{R}^3/G$, where G is the decomposition associated to our initial package), and Theorem 5.2. Theorem 1.12.b) follows from Theorem 5.3, Lemma 3.45, and Proposition 3.70.

Incidentally, the fact that Bing's dogbone space $\mathbb{R}^3/G$ could be embedded topologically in $\mathbb{R}^4$ was observed long ago [Cu].

REMARK 5.6. Note that we can make the singular set F of M have Hausdorff dimension as small as we want in this example, by taking the parameter ρ to be small. See (3.48).

6. Bing doubling.

We shall use the definitions and notations from Section 3 freely in this section.

To prove Theorem 1.12.c) we use another example studied by Bing [B1], [D, p. 62, Example 1]. We start with a smooth solid torus D in $\mathbb{R}^3$, and we take $n = 2$ and D_1, D_2 to be two disjoint smooth solid tori in D which are folded over and linked as in [B1, p. 357, Figure 3] and [D, p. 63, Figure 9-1]. Each of these two tori are (separately) embedded in a topologically trivial manner in D, and we have that

(6.1) there are open sets $U_1, U_2 \subseteq D$ with $\overline{D}_i \subseteq U_i$ such that $\overline{U}_1, \overline{U}_2$ are each diffeomorphic to the closed unit 3-ball.

However, as a pair, D_1 and D_2 are linked, in the sense that they cannot be pulled apart by an isotopy of D onto itself. Note that D, D_1, and D_2 satisfy (3.69).

Let Ω and ω_1, ω_2 be solid versions of the D's in $\mathbb{R}^4$ which satisfy (3.3). As before, "solid version" means that Ω should be diffeomorphic to $D \times (-1, 1)$, with $\Omega \cap P$ corresponding to $D \times \{0\}$, and similarly for the ω_i's. One should not go out of one's way to choose them stupidly, and we shall see in the next section that it is a good idea to choose them to be symmetric about P. As in the preceding section, it is better to choose ω_1, ω_2 to lie much closer to P than Ω, so that we can disentangle ω_1 from ω_2 in Ω with ease. For instance, it is convenient to require that there exist $\varepsilon > 0$ so that we have the following solid version of (6.1):

$$\omega_i \subseteq U_i \times (-\varepsilon, \varepsilon) \subseteq \overline{U}_i \times [-\varepsilon, \varepsilon] \subseteq \Omega\,. \tag{6.1$'$}$$

Let us ask also that $\omega_i \supseteq D_i \times (-b, b)$ for some $b > 0$ (for a minor technical convenience in Section 8).

Let ψ_1 and ψ_2 be similarities on $\mathbb{R}^4$ with the same dilation factor ρ which map P onto itself and which send Ω onto domains Ω_1, Ω_2 in Ω with disjoint closures and which stay away from the ω_i's. As usual it is good for ρ to be small, and we can take it to be as small as we want. For the same reason as in Section 5 we can build a mapping θ which satisfies the requirements of Definition 3.2. We lift up ω_1 (in the positive x_4 direction), deform it into a copy of Ω_1 sitting just over Ω_1, and drop it onto Ω_1, and then we repeat the process for Ω_2, taking care that the deformations stay inside Ω and do not disturb the other players (*i.e.*, we do not touch ω_2 when deforming ω_1). We can use Lemma 4.1 to extend these deformations to all of $\mathbb{R}^4$ in such a way that points outside Ω never move and points in $\omega_1, \omega_2, \Omega_1, \Omega_2$ move only when they are supposed to. In the end we get a mapping θ with the right properties. As usual, we can simply define the ϕ_i's as in Definition 2.3 from this construction of θ (or make unnecessary efforts to adjust θ to previous choices of the ϕ_i's). The bottom line is that we have initial and excellent packages in this case, and so we get the associated decomposition G of $\mathbb{R}^3$ (as in Section 2) and the sets M, M^j constructed in Section 3. (We do not need $\widetilde{M}$ for this example.)

It turns out that this decomposition is well behaved topologically, but for nontrivial reasons.

Theorem 6.2 (Bing [B1]) *$\mathbb{R}^3/G$ is homeomorphic to $\mathbb{R}^3$, and in fact there exists a homeomorphism f from $\mathbb{R}^3/G$ onto $\mathbb{R}^3$ which agrees with the "identity" on the complement of $D_1 \cup D_2$.*

This result is given in [B1, Section 3, Paragraph III]. See also [B5].

Theorem 6.3. *M satisfies* $(**)$.

Let a be chosen as in Lemma 3.65. Let $x \in M$ and $r > 0$ be given, so that we want to find a topological ball U and a parameterization of it which satisfy the conditions in Definition 1.7. If x, r satisfy i), iii), or v) in Lemma 3.44, then we are in business, because of Lemma 3.65. Suppose that we are in case iv) in Lemma 3.44. Because of (3.20) (with $j = \infty$, so that $M^{j-l} = M$) we can reduce to the case where $l = 1$. Note that there is nothing fishy going on here with the uniform estimates, because they are all chosen to behave properly under similarities. (Do not forget that the radius r changes also with the similarity.) Let h be as in Lemma 3.21, and choose $i = 1, 2$ so that $\Omega_\alpha \cap M = h(D_i)$, where Ω_α is as in iv) in Lemma 3.44 (and hence $\alpha \in \mathcal{S}_1$). Set $U = h(U_i)$, where U_i is as in (6.1). Lemma 3.21 tells us that h descends to a homeomorphism from $\mathbb{R}^3/G$ onto M, and so $\overline{U}$ is homeomorphic to $\overline{U}_i/G$. Our mapping f above (from Theorem 6.2) provides us with a homeomorphism from $\overline{U}_i/G$ onto $\overline{U}_i$. The bottom line is that U is a topological ball whose closure is homeomorphic to a closed 3-ball. Since there are only two choices here there is no problem with getting the uniform estimates required in Definition 1.7, since continuous maps between compact sets are uniformly continuous. In other words, in case iv) we get our uniformity because the self-similarity (3.20) allows us to reduce to a finite number of mappings. The whole construction in Section 3 was designed to make this happen.

We are left with ii) in Lemma 3.44, which is slightly a nuisance but not deep. Set $r_1 = r_1(r) = 2r + \operatorname{diam}\Omega$, so that

$$r_1 \leq C_0\, r \tag{6.4}$$

(with $C_0 = 2 + a^{-1}$) by ii), and let h be as in Lemma 3.21 again. Notice that $B(x, r_1) \supseteq \Omega$, and that

$$h(B(x,r_1) \cap P) = ((B(x,r_1) \cap P) \setminus D) \cup (\Omega \cap M) \supseteq B(x,r) \cap M\,. \tag{6.5}$$

Setting $U = h(B(x,r_1) \cap P)$, we have that $B(x,r) \cap M \subseteq U = B(x,r_1) \cap M$.

Let F be the homeomorphism from $\mathbb{R}^3$ onto M obtained in the following manner. We know that h descends to a homeomorphism from $\mathbb{R}^3/G$ onto M, and that f above is a homeomorphism from $\mathbb{R}^3/G$ onto $\mathbb{R}^3$, and we take F to be the composition of the former with the inverse of the latter. From the properties of h and f we get that F equals the identity off D and that F maps D onto $\Omega \cap M$. Thus $U = F(B(x, r_1))$, and U is a topological ball in particular. We need to check that we can parameterize U with the right kind of uniform estimates for the moduli of continuity.

Since F is a homeomorphism which equals the identity off the compact set D there is a locally bounded function $\zeta : [0,\infty) \longrightarrow [0,\infty)$ such that $\lim_{t\to 0} \zeta(t) = 0$,

$$|F(y) - F(z)| \leq \zeta(|y - z|) \qquad \text{when } y, z \in \mathbb{R}^3, \tag{6.6}$$

$$|F^{-1}(v) - F^{-1}(w)| \leq \zeta(|v - w|) \qquad \text{when } v, w \in M, \tag{6.7}$$

and

$$\zeta(t) = 2t \qquad \text{when } t > \operatorname{diam} \Omega. \tag{6.8}$$

These estimates are not quite what we want, because they do not scale properly. The estimates that we need come down to the existence of a function $\xi : [0,\infty) \longrightarrow [0,\infty)$ such that $\lim_{t\to 0} \xi(t) = 0$,

$$|F(y) - F(z)| \leq r\, \xi(r_1^{-1} |y - z|) \tag{6.9}$$

when $y, z \in B(x, r_1) \cap \mathbb{R}^3$, and

$$|F^{-1}(v) - F^{-1}(w)| \leq r_1\, \xi(r^{-1} |v - w|) \tag{6.10}$$

when $v, w \in B(x, r) \cap M$.

It is important here that ξ not depend on r (or x); if not for this we could simply compute ξ from ζ. As it is, we are lead to try

$$\xi(t) = \sup\{r^{-1}\, \zeta(r_1 t) + r_1^{-1}\, \zeta(r\, t) : \ r \geq a \operatorname{diam} \Omega\}, \tag{6.11}$$

(where r_1 is still related to r as above). This is actually finite and locally bounded for all t, because of (6.8), but we need to check that $\lim_{t\to 0} \xi(t) = 0$. Let ε in $(0,1)$ be given, and choose $\delta \in (0,1)$ so that $\zeta(t) < \varepsilon$ when $t < \delta$. Suppose that $t < C_0^{-1}\, \varepsilon\, \delta$ (where C_0 is as in (6.4)), and let us show that $\xi(t)$ is small. Let ξ_1, ξ_2, and ξ_3 be defined as in (6.11), but where the supremum is limited to the ranges $r > t^{-1} \operatorname{diam} \Omega$, $\varepsilon^{-1} \leq r \leq t^{-1} \operatorname{diam} \Omega$, and $a \operatorname{diam} \Omega \leq r < \varepsilon^{-1}$. Thus ξ is the same as the maximum of the ξ_i's. For ξ_1 we use (6.8) and (6.4) to get

$$\xi_1(t) \leq r^{-1}\, (2 r_1 t) + r_1^{-1}\, (2 r t) \leq C\, t \leq C \varepsilon. \tag{6.12}$$

For ξ_2 we use the fact that $\sup\{\zeta(s) : s \leq C_0 \operatorname{diam} \Omega\} < \infty$ to obtain

$$\xi_2(t) \leq C\, \varepsilon. \tag{6.13}$$

As for ξ_3 we simply use (6.4) and the fact that r is bounded from below to get

$$\xi_3(t) \leq C \sup\{\zeta(s) : s \leq \delta\} \leq C\,\varepsilon\,. \tag{6.14}$$

Altogether we get that $\xi(t) \leq C\varepsilon$ when $t < C_0^{-1}\,\varepsilon\,\delta$, and so $\lim_{t\to 0}\xi(t) = 0$, as desired. This completes the proof of Theorem 6.3.

Theorem 6.15. *M does not admit a quasisymmetric parameterization by $\mathbb{R}^3$.*

This is basically a small perturbation of [FS, Theorem 2.1]. This result in [FS] says that a certain discrete group of homeomorphisms on $\mathbf{S}^3$ is not homeomorphically conjugate to a group of quasiconformal mappings with uniformly bounded dilatation. The present story amounts to building a different metric space where a topologically equivalent form of this group acts uniformly quasiconformally, and we are concluding that this metric space cannot be quasiconformally equivalent to $\mathbf{S}^3$. However it is easier to prove the theorem directly. The idea of the proof is that the sets $\Omega_\alpha \cap M$ all look alike and are reasonably well-shaped, while any homeomorphism from M to $\mathbb{R}^3$ has to twist at least some of these sets rather severely, and more so than a quasisymmetric map can.

To make this precise, let g be a homeomorphism from M onto $\mathbb{R}^3$, let $\mathcal{T}_l^0$ denote the collection of subsets of M of the form $\overline{\Omega}_\alpha \cap M$ with $\alpha \in \mathcal{S}_l$, and let $\mathcal{T}_l$ be the set of images of elements of $\mathcal{T}_l^0$ under g. We want to show that no matter how g is chosen the geometry of some of the elements of $\mathcal{T}_l$ will have to degenerate as $l \to \infty$. This will come down to a lemma in [FS].

Let F be the homeomorphism described just after (6.5). Using F we obtain that the elements of $\mathcal{T}_l^0$ are homeomorphic to smooth solid tori in $\mathbb{R}^3$ when $l = 0, 1$, and they even have neighborhoods in M which are homeomorphic to solid tori. The same is true for all l because of the self-similarity property (3.20). Thus the elements of $\mathcal{T}_l$ are locally flat topological solid tori in $\mathbb{R}^3$. Let T denote the unique element of $\mathcal{T}_0$. This is the solid torus in which all the action takes place, because it contains all the others.

If $\Gamma \in \mathcal{T}_l$, define length(Γ) to be the infimum of the (parameterized) length of the loops in the interior of Γ which represent a generator in the fundamental group of Γ. The next lemma is a consequence of [FS, p. 81, Lemma 2.2].

Lemma 6.16. *There is a constant $C > 0$ so that*

$$2^{-l} \sum_{\Gamma \in \mathcal{T}_l} \operatorname{length}(\Gamma) \geq C^{-1}, \tag{6.17}$$

for all l. In particular, for each l there is at least one $\Gamma \in \mathcal{T}_l$ such that $\operatorname{length}(\Gamma) \geq C^{-1}$.

The second statement follows from the first, because $\mathcal{T}_l$ has 2^l elements (since $n = 2$ in this example). The idea behind the first part is pretty simple. Fix an l, and choose for each $\Gamma \in \mathcal{T}_l$ a loop γ in the interior of Γ which represents a generator of the fundamental group of Γ. Consider the combination of all 2^l of these γ's. The point is that every time we increase l by 1 we get to double the number of γ's, but we also have to double the number of times that this system of curves "goes around" in T. ("Goes around" should be interpreted geometrically, and not homologically, because these curves do not go around in T homologically at all.) It is easy to believe this after staring at the pictures, and a proof is given in [FS].

Lemma 6.18. *Assume that $g : M \longrightarrow \mathbb{R}^3$ is actually quasisymmetric. Then there is a constant $C > 0$ so that* $\operatorname{length}\Gamma \leq C \operatorname{diam}\Gamma$ *for all $\Gamma \in \mathcal{T}_l$ and any l.*

Let $\Gamma \in \mathcal{T}_l$ be given, and let $\alpha \in \mathcal{S}_l$ be chosen so that $\Gamma = g(\overline{\Omega}_\alpha \cap M)$. Let us first choose a loop γ_0 in $\Omega_\alpha \cap M$ which represents a generator of its fundamental group and which comes from some fixed smooth curve in $\Omega \cap M$. That is, we first make a nice choice of such a curve (call it τ) in $\Omega \cap M$, and then we use the self-similarity property (3.20) and take $\gamma_0 = \psi_\alpha(\tau)$. Thus γ_0 is smooth at the scale of $\varepsilon \operatorname{diam}\Omega_\alpha$ and $\operatorname{dist}(\gamma_0, M \backslash \Omega_\alpha) \geq \delta \operatorname{diam}\Omega_\alpha$ for fixed $\varepsilon, \delta > 0$ which do not depend on α or l.

Set $\gamma = g(\gamma_0)$. This represents a generator in the fundamental group of Γ, but it may have infinite length. However, there is a fixed $\delta' > 0$, which does not depend on Γ, such that $\operatorname{dist}(\gamma, \mathbb{R}^3 \backslash \Gamma) \geq \delta' \operatorname{diam}\Gamma$. This follows from the corresponding property of γ_0 and the quasisymmetry condition. This gives us enough room to deform γ inside Γ to a loop with length less or equal than $C \operatorname{diam}\Gamma$, as desired. (To be honest, to check this carefully one should notice that γ cannot oscillate too many times at the scale of $(\delta'/10)\operatorname{diam}\Gamma$, say, because of the smoothness property of γ and the quasisymmetry of g. This implies that we only need to make a bounded number of modifications to γ in a bounded number of little balls. Note that the condition $\operatorname{dist}(\gamma, \mathbb{R}^3 \backslash \Gamma) \geq \delta' \operatorname{diam} \Gamma$ does not prevent γ from looping around many, many times in a little neighborhood of itself, even though we know that this cannot happen here.)

Theorem 6.15 follows from Lemmas 6.16 and 6.18, because $\operatorname{diam} \Gamma$ tends to 0 as $l \to \infty$ when $\Gamma \in \mathcal{T}_l$ (since the corresponding statement is true for $\mathcal{T}_l^0$ and g is continuous). Theorem 1.12 .c) now follows from Theorems 6.3 and 6.15 together with Lemma 3.45 and Proposition 3.70.

REMARK 6.19. I have a philosophical explanation for Theorem 6.15 which I cannot back up with a proof but which I would like to share with the reader. The self-similarity properties of M stem from the fact that we constructed Σ as in (3.14) so that its big boundary component is similar to each of its small boundary components. The corresponding part of our initial package is the set $D \setminus \cup_j \overline{D}_j$, and one reason that we cannot build a quasisymmetric parameterization of M is that $D \setminus \cup_j \overline{D}_j$ does not possess a version of this property. Specifically, if ∂D were conformally equivalent to each ∂D_j, then we could try to build a quasiconformal parameterization of M by building a suitable quasiconformal map from $D \setminus \cup_j \overline{D}_j$ onto Σ and putting copies of it on top of itself. In order to have the dilatation not build up and become unbounded in the limit we need this quasiconformal building block to be conformal at the ends, which is why we would need to have ∂D be conformally equivalent to each ∂D_j. However, it is not true that 2-dimensional tori are all conformally equivalent, and this is the source of the problem. I do not know how to turn this explanation into an alternative proof of Theorem 6.15, but I think that it would be interesting to do so, especially for the purpose of understanding other examples of this type. (See also Section 12.)

7. The complementary components, part 1.

Throughout this section we assume that we are given an excellent package as in Definition 3.2, and we shall consider the behavior of the complementary components of M, the M^j's, and $\widetilde{M}$. We shall use freely the definitions and notation of Section 3.

Let U^+ and U^- denote the two components of $\mathbb{R}^4 \backslash P$, *i.e.*, U^+ is the set of points $x \in \mathbb{R}^4$ such that $x_4 > 0$, and similarly for U^-. Define X^+ and X^- by

$$X^{\pm} = \Big(\overline{\Omega} \backslash \bigcup_{j=1}^{n} \omega_j\Big) \cap U^{\pm} . \tag{7.1}$$

In this section we require that

$$X^{\pm}, \Omega \cap U^{\pm}, \text{ and } U^{\pm} \backslash \Omega \text{ are connected.} \tag{7.2}$$

This condition is not necessarily minimal, but it is valid in the examples in Sections 4, 5, and 6 and others like them, and it is nicely clear.

Lemma 7.3. *Each of M, the M^j's, and $\widetilde{M}$ has exactly two complementary components in $\mathbb{R}^4$.*

For $\widetilde{M}$ and the M^j's this is an immediate consequence of Lemma 3.21. For M we have to be slightly more careful.

Let $g_l : \mathbb{R}^4 \longrightarrow \mathbb{R}^4$ be as in the proof of Lemma 3.21, $l = 0, 1, 2, \ldots$ and define a mapping $e : \mathbb{R}^4 \backslash M \longrightarrow \mathbb{R}^4$ by

$$e(x) = \lim_{l \to \infty} g_0^{-1} \circ g_1^{-1} \circ \cdots \circ g_l^{-1} . \tag{7.4}$$

This limit exists because $g_k(x) = x$ for any $x \in \mathbb{R}^4 \backslash M$ and all sufficiently large k, by (3.26). Also, if h_l is as in (3.27),

$$h_l(e(x)) = x , \tag{7.5}$$

for any $x \in \mathbb{R}^4 \backslash M$ and all sufficiently large l. In particular we have that e maps $\mathbb{R}^4 \backslash M$ into $\mathbb{R}^4 \backslash P$. Since e equals the identity outside Ω we conclude that $\mathbb{R}^4 \backslash M$ has at least two components.

Let us analyze $\mathbb{R}^4 \backslash M$ a little more in order to show that it has at exactly two components. Recall the definition of Y (shortly before (3.12)) and set $Y^+ = \theta(X^+), Y^- = \theta(X^-)$. Set $Y_\alpha^+ = \psi_\alpha(Y^+)$ and $Y_\alpha^- = \psi_\alpha(Y^-)$ for any α in any $\mathcal{S}_l$, and define V^+ and V^- by

$$V^{\pm} = (U^{\pm} \backslash \Omega) \cup \bigcup_{l=0}^{\infty} \bigcup_{\alpha \in \mathcal{S}_l} Y_\alpha^{\pm} . \tag{7.6}$$

It is easy to see that

$$\mathbb{R}^4 \backslash M = V^+ \cup V^- , \tag{7.7}$$

because of (3.13) and (3.19). Let us check that

$$V^+ \text{ and } V^- \text{ are connected.} \tag{7.8}$$

To see this note that if $\alpha \in S_l$, $l \geq 1$, and δ is the parent of α, then

$$Y_\alpha^+ \text{ touches } Y_\delta^+ \tag{7.9}$$

(at their respective boundaries), and similarly for Y_α^-. When $l = 1$ (7.9) reduces easily to the fact that the ψ_j's from Definition 3.2 are required to be orientation-preserving. The general case can be reduced to this one by using the fact that $\psi_\alpha = \psi_\delta \circ \psi_j$ for some j (as in the definition of ψ_α just before (3.7)). We also have that

$$Y^+ \text{ touches } U^+ \backslash \Omega \tag{7.10}$$

(at their boundaries), and similarly for Y^-, and (7.8) follows easily from these observations and our assumption (7.2). This proves Lemma 7.3.

Note that the complementary domains of M^j also admit an expression like (7.7). That is, if we define V_j^+ and V_j^- by

$$V_j^\pm = (U^\pm \backslash \Omega) \cup \Big(\bigcup_{l=0}^{j-1} \bigcup_{\alpha \in S_l} Y_\alpha^\pm \Big) \cup \Big(\bigcup_{\alpha \in S_j} \Omega_\alpha \cap U^\pm \Big), \tag{7.11}$$

then

$$\mathbb{R}^4 \backslash M^j = V_j^+ \cup V_j^-, \tag{7.12}$$

because of (3.12) and (3.17), and $V_j^\pm$ are connected for the same reasons as for (7.8). Thus $V_j^\pm$ are the two complementary components of M^j.

Next we want to show that these various complementary domains are *uniform domains.* Recall that a domain O in $\mathbb{R}^4$ is a uniform domain if there exists a constant C so that for each pair of points $x, y \in O$ we can find a path Γ in O which connects x and y and satisfies

$$\operatorname{diam} \Gamma \leq C\,|x - y| \tag{7.13}$$

and

$$\operatorname{dist}(z, \partial O) \geq C^{-1} \operatorname{dist}(z, \{x, y\}) \qquad \text{when } z \in \Gamma. \tag{7.14}$$

(There are many equivalent characterizations of uniform domains, and this condition is a little weaker in appearance than some of the others.) Bounded smooth domains are uniform domains, but a domain with an outward-pointing cusp is not. Note that this condition is scale-invariant.

Lemma 7.15 *The complementary components of M, the M^j's, and $\widetilde{M}$ in $\mathbb{R}^4$ are uniform domains, with constants bounded independently of j (in the case of the M^j's).*

This is a fairly straightforward exercise. We could derive this lemma from the general results in [V2], but let us give instead a direct argument. Consider M and the M^j's and their complementary components V^+ and V_j^+. Let us call sets of the form $U^+ \backslash \Omega$, Y_α^+ ($\alpha \in \mathcal{S}_l, l < j$), and $U^\pm \cap \Omega_\alpha$ ($\alpha \in \mathcal{S}_j$) *building blocks* for V^+ or V_j^+, as appropriate. Notice that these building blocks are all uniform domains, with uniformly bounded constant. In the case of $U^+ \backslash \Omega$ this is true because it has smooth boundary and looks like a half-space at infinity. For the other building blocks this uniform estimate follows from the fact that they are all similar to one of a finite number of bounded smooth domains. The union of two building blocks which touch is also a uniform domain with bounded constant, for the same reasons. (Remember also that a Y_α and a Y_β can touch only when one of α and β is the parent of the other, as observed just before (3.12).)

Suppose now that we are given two points x, y in V^+ or V_j^+. We want to connect them with a curve which satisfies (7.13) and (7.14) (with $O = V^+$ or V_j^+, as appropriate). If x and y lie in the same building block then we are in business, by the preceding remarks, and also when they lie in different building blocks which touch. Thus we may assume that x and y lie in disjoint building blocks. Consider first the case where $x \in Y_\alpha^+$ and $y \in Y_\beta^+$ for some $\alpha \neq \beta$, $\alpha \in \mathcal{S}_l$, $\beta \in \mathcal{S}_k$, $k, l < j$. Our assumption of disjointness implies that neither of α or β is the parent of the other. Let δ be the "last" common ancestor of α and β, so that either $\alpha = \delta$, $\beta = \delta$, or α and β are descended from different children of δ. Notice that

$$\operatorname{dist}(Y_\alpha^+, Y_\beta^+) \le |x - y|. \tag{7.16}$$

$$\operatorname{diam} Y_\alpha^+ + \operatorname{diam} Y_\beta^+ \le 2 \operatorname{diam} \Omega_\delta \le C \operatorname{dist}(Y_\alpha^+, Y_\beta^+). \tag{7.17}$$

(The last inequality reduces via the similarity ψ_δ to the fact that the distance between the children of Ω is positive (when δ is different from both α and β), and to the fact that the distance between Y and any of the Ω_η's with $\eta \in \mathcal{S}_2$ is positive (when δ equals one of α and β). We get a uniform estimate because the similarity allows us to reduce to a finite number of cases.) We can build a curve which goes from x to ∂Y_α^+, then ascends to ∂Y_γ^+ for the various successive ancestors γ of α until we get to δ, and then goes down to ∂Y_γ^+ for the successive descendants γ of δ until we get to ∂Y_β^+, from which we can connect easily to β. This curve will satisfy (7.13) because it will be contained in Ω_δ and because $\operatorname{diam} \Omega_\delta \leq C\,|x-y|$, by (7.16) and (7.17). If we are a little bit careful we can choose the curve so that (7.14) also holds. (The point is to stay as far away from $\partial Y_\gamma^+ \cap M^j$ as possible -*i.e.*, at distance greater or equal than $C^{-1} \operatorname{diam} Y_\gamma^+$- for all the intermediate γ's. In Y_α^+ and Y_β^+ it may be necessary to let the curve get close to the boundary of V^+ or V_j^+ because of the positions of x and y.) The remaining possible situations where x and y lie in disjoint building blocks (*e.g.*, $x \in U^+ \backslash \Omega$, or $y \in U^+ \cap \Omega_\alpha$ for an $\alpha \in \mathcal{S}_j$) are handled in essentially the same way.

Thus one can show that V^+ and the V_j^+'s are uniform domains, with bounded constants. The same argument works for the V^-'s. It is easy to show that the complementary components of $\widetilde{M}$ are uniform domains using the corresponding statement for the M^j's and the construction of $\widetilde{M}$ (just after (3.20)). For this one should go through the usual analysis of cases, *i.e.*, the cases where the given pair of points both lie in the same $2B_k$, or they both lie outside all the B_k's, or they lie in different $2B_k$'s, or one lies in some $2B_k$ and the other lies outside all the other B_l's. The details are left to the reader. This completes the proof of Lemma 7.15.

In our examples we also have that there exist bilipschitz reflections across M, the M^j's, and $\widetilde{M}$. To see this we first define a suitable symmetry condition for an excellent package.

Definition 7.18. *An excellent package as in Definition* 3.2 *is said to be symmetric if Ω, the ω_j's and the Ω_j's are symmetric about P, and if the restriction of θ to a neighborhood of each $\overline{\omega}_j$ commutes with the obvious reflection about P.*

It is important here that we do not require that θ commute with the reflection about P everywhere, because that will not be true in the interesting examples.

Lemma 7.19. *We can choose the excellent packages in Sections 4, 5, and 6 to be symmetric.*

This is straightforward but slightly tedious. There is absolutely no problem about making the various Ω's and ω_j's be symmetric, but the symmetry condition on θ is slightly more complicated. The main point is that we always obtained θ by first building suitable isotopies on the ω's and then extending them to all of Ω using Lemma 4.1. The first step is the only one that really matters here and is the one which is most easily controlled. It is easy to check that it can be carried out in such a manner as to get the desired symmetry condition. In the examples in Sections 5 and 6, for instance, these isotopies on the ω's could be taken to be translations in the positive x_4 direction, followed by a certain "unwinding" operation centered on the relevant vertical translate $P'$of P, followed by a translation back down to P. One need only demand that this middle unwinding operation be symmetric with respect to P', which is easily accomplished, because this unwinding operation really comes from a 3-dimensional process. The example in Section 4 is a bit different, because one translates part of ω up a little while leaving the other part alone, but it is also not difficult to handle. We leave the details as an exercise. (Keep in mind that the intermediate stages of the deformation do not have to be symmetric about P, only the end result.)

Proposition 7.20. *If our excellent package is symmetric, then there exists a bilipschitz reflection r on $\mathbb{R}^4$ across M, and there exist reflections r_j across each M^j which are uniformly bilipschitz. These reflections all agree with the standard reflection across P outside Ω. There is also a bilipschitz reflection across $\widetilde{M}$.*

Let τ denote the standard reflection across P, and for any α in any S_l define σ_α by

$$\sigma_\alpha = \psi_\alpha \circ \theta \circ \tau \circ \theta^{-1} \circ \psi_\alpha^{-1}. \tag{7.21}$$

This is the same as taking $\theta \circ \tau \circ \theta^{-1}$, which agrees with τ outside Ω and on a neighborhood of its boundary (by (3.5)) but does something complicated inside, and then transporting it to Ω_α using ψ_α. Our symmetry assumptions in Definition 7.18 imply that $\theta \circ \tau \circ \theta^{-1}$ also agrees with τ on a neighborhood of each $\overline{\Omega}_i$, and hence

$$\sigma_\alpha = \tau \text{ on } \mathbb{R}^4 \backslash Y_\alpha \text{ and on a neighborhood of } \partial Y_\alpha \tag{7.22}$$

(where Y_α is as defined just before (3.12)). Also,

$$\sigma_\alpha(Y_\alpha^+) = Y_\alpha^- \quad \text{and} \quad \sigma_\alpha(Y_\alpha^-) = Y_\alpha^+ \,, \tag{7.23}$$

by definition of $Y_\alpha^\pm$, and

$$\sigma_\alpha(x) = x \quad \text{for all } x \in \Sigma_\alpha \,, \tag{7.24}$$

because of the definition of Σ_α (just before (3.15)) and the fact that τ fixes every element of P.

Define r by taking r to be the identity on M, $r = \tau$ on $\mathbb{R}^4\backslash\Omega$, and $r = \sigma_\alpha$ on each Y_α. Define r_j to be the identity on M^j, $r_j = \tau$ on $\mathbb{R}^4\backslash\Omega$, and $r_j = \sigma_\alpha$ on each Y_α with $\alpha \in \mathcal{S}_l$, $l < j$, and $r_j = \tau$ on Ω_α when $\alpha \in \mathcal{S}_j$. Clearly r and the r_j's are involutions, and one can use the formulae (7.6) and (7.11) for the complementary components of M and M^j and (7.23) to show that r and the r_j's exchange the complementary components of M and M^j. It is not hard to check that r and the r^j's are continuous (and even smooth in the case of the r_j's), because of the way that the Ω_α's fit together, and because of (7.22). One should be a little bit careful about the continuity of r and the r_j's across M and the M^j's, but the only slightly tricky issue is the continuity of r at points in the singular set F of M. For this it is it is useful to observe that

$$r(\Omega_\alpha) = \Omega_\alpha \,, \tag{7.25}$$

for all α in any $\mathcal{S}_l$. This observation is easy to derive from the definition of r and the simpler fact that $r(Y_\alpha) = Y_\alpha$ for all α. (Note that the analogue of (7.25) for the r_j's holds as well.)

The uniform Lipschitz conditions are easy to check. The main point is that the σ_α's are uniformly Lipschitz. For the r_j's this is enough because the smoothness of the r_j's permits their uniform Lipschitzness to be derived from a bound on their gradients, which we get from the corresponding bounds for the σ_α's. In the case of r one must be a little more careful, since it is not smooth across M, but the continuity of r across M allows one to piece the local Lipschitz conditions together. Uniform bilipschitzness follows from uniform Lipschitzness and the fact that these mappings are involutions (and hence are their own inverses).

As usual, one can deal with $\widetilde{M}$ by treating separately the little pieces which look like M^j's (as in the definition of $\widetilde{M}$, just after (3.20)).

RECAPITULATION 7.26. Under mild additional conditions on our excellent package which are satisfied in the examples, we obtain that M, the M^j's, and $\widetilde{M}$ have exactly two complementary components which are uniform domains, and there are uniformly bilipschitz reflections across the M's.

8. The complementary components, part 2.

Throughout this section we use the excellent package described in Section 6, taking a modicum of care to ensure that the additional requirements of the preceding section ((7.2) and symmetry) are met. We shall use freely the definitions and notation of Section 3, and in particular we assume that M has been constructed as in Section 3. Let $V^{\pm}$ be the complementary domains of M, as in (7.6) and (7.7).

Theorem 8.1. a) *$V^{\pm}$ are uniform domains, and there is a bilipschitz reflection on $\mathbb{R}^4$ which equals the identity outside Ω, fixes every point on M, and interchanges V^+ and V^-. Neither V^+ nor V^- is quasiconformally equivalent to a ball (or a half-space).*

b) *There is a homeomorphism ν from $\mathbb{R}^4$ onto itself such that $\nu =$ the identity outside Ω, $\nu = \theta$ outside $\omega_1 \cup \omega_2$, $\nu(\omega_i) = \Omega_i$ for $i = 1, 2$, and ν maps P onto M.*

c) *There exist a constant $C > 0$ and a locally bounded function $\eta : [0, \infty) \longrightarrow [0, \infty)$ with $\lim_{t \to 0} \eta(t) = 0$ such that for each $x \in M$ and $r > 0$ there is an open set $W \subseteq \mathbb{R}^4$ with $B(x, r) \subseteq W \subseteq B(x, Cr)$ and a homeomorphism μ from $B(0,1)$ onto W which satisfy $\mu(B(0,1) \cap P) = M \cap W$,*

$$(8.2) \qquad |\mu(y) - \mu(z)| \leq r\, \eta(|y - z|), \quad \text{for all } y, z \in B(0,1),$$

and

$$(8.3) \qquad |\mu^{-1}(p) - \mu^{-1}(q)| \leq \eta(r^{-1}\, |p - q|), \quad \text{for all } p, q \in U.$$

Part c) is a stronger version of (**) for M which takes into account the ambient space as well. It implies that V^+ and V^- are *strongly uniform* domains in the sense of [HY]. This follows from the compactness result in [HY, Theorem 4.4].

One should keep in mind that the boundaries of these "Bing domains" are Ahlfors regular (at least if we choose ρ small enough, as in Lemma 3.45), uniformly rectifiable (in the sense of [DS4]), and have well-behaved geodesic distance functions (as in Theorem 1.12 and Lemma 3.69). One can construct much simpler examples which satisfy the properties in Theorem 8.1 using the methods of [Tu2], but the boundaries of these examples are not so well behaved.

The fact that $V^{\pm}$ are uniform domains and the existence of the bilipschitz reflection were proved in the preceding section (Lemma 7.15 and Proposition 7.20). It follows that $V^{\pm}$ cannot be quasiconformally equivalent to a ball, since, as pointed out in the introduction, well-known results about uniform domains then imply that the quasiconformal map would extend to a quasisymmetric map between the closures, which would contradict the fact that M is not quasisymmetrically equivalent to $\mathbb{R}^3$. This gives a).

Let us assume b) for the moment and derive c). We use the same argument as in the proof of Theorem 6.3. We choose a as in Lemma 3.65, and we consider separately the cases i)-v) in Lemma 3.44. The cases i), iii), and v) are covered by Lemma 3.65. We shall use the next lemma to deal with case iv), and we shall consider ii) after that.

Lemma 8.4. *There exist open subsets O_1 and O_2 of Ω such that $\Omega_i \subseteq O_i$ for $i = 1, 2$ and each $\overline{O}_i$ is homeomorphic to $\overline{B}(0,1)$ via a mapping which sends $M \cap O_i$ onto $P \cap B(0,1)$.*

Set $O_i' = U_i \times (-\varepsilon, \varepsilon)$, where U_i and ε are as in (6.1) and (6.1'). Clearly each $\overline{O}_i'$ is homeomorphic to $\overline{B}(0,1)$ via a mapping which takes P to itself, since the U_i's are closed topological 3-balls. Thus if ν is as in b) then $O_i = \nu(O_i')$, $i = 1, 2$, are open subsets of Ω such that $O_i \supseteq \Omega_i$ and each $\overline{O}_i$ is homeomorphic to $\overline{B}(0,1)$ via a mapping which sends M to P. This proves Lemma 8.4.

Now suppose that x, r satisfy iv) in Lemma 3.44. If the l from iv) equals 1, then we can take W to be one of the O_i's from Lemma 8.4. When $l > 1$ we can reduce to the $l = 1$ case using the self-similarity property (3.20). These choices of W admit homeomorphic parameterizations with the right kind of equicontinuity conditions because they all reduce to the two (uniformly continuous) models in Lemma 8.4 by self-similarity.

The remaining case, where x, r satisfies ii), is handled in exactly the same way as the corresponding step in the proof of Theorem 6.3 (with ν playing the role of F). That is, one must go through the song-and-dance of (6.11), but that part of the argument applies equally well to the current situation. This proves c) given b).

We are left with proving b). Recall that G is the decomposition of $\mathbb{R}^3$ which is associated to the present initial package as in Section 2 (just after Definition 2.2). (See also the discussion around Sublemma 3.40.) Let G' denote the decomposition of $\mathbb{R}^4$ which extends G trivially. That is, G' consists of the elements of G (viewed as subsets of $\mathbb{R}^4$) together with the singletons from $\mathbb{R}^4 \backslash P$. Our first task in proving b) is to come to grips with $\mathbb{R}^4/G'$.

Proposition 8.5. *There is a homeomorphism ξ from $\mathbb{R}^4/G'$ onto $\mathbb{R}^4$ which sends P/G onto P and which equals "the identity" outside $\omega_1 \cup \omega_2$.*

In other words, there is a nice extension of the homeomorphism given by Theorem 6.2 to $\mathbb{R}^4$. To prove this we use the following.

Lemma 8.6. *There is a continuous 1-parameter family f_t, $t \in [0,1]$, of continuous mappings from $\mathbb{R}^3$ onto itself such that f_0 is the identity, each f_t agrees with the identity outside $D_1 \cup D_2$ (where D_1 and D_2 come from our initial package), each f_t for $t < 1$ is a homeomorphism, and f_1 induces a homeomorphism from $\mathbb{R}^3/G$ onto $\mathbb{R}^3$.*

Thus the homeomorphism in Theorem 6.2 can be deformed to the identity in a natural way. This follows from Bing's construction of a homeomorphism as in Theorem 6.2. Specifically, Bing produces an increasing sequence of integers $\{j_i\}$ with $j_i \geq 1$ and homeomorphisms $\{T_i\}$ on $\mathbb{R}^3$ such that T_i maps the $\overline{D}_\alpha$'s with $\alpha \in \mathcal{S}_{j_i}$ to sets of diameter less that $\frac{1}{i}$ and $T_i = T_{i-1} \circ T_i'$, where T_i' takes each $\overline{D}_\beta$ to itself when $\beta \in \mathcal{S}_{j_{i-1}}$ and T_i' equals the identity outside of all these D_β's. Each of the T_i''s can be deformed continuously to the identity through homeomorphisms which also take each of these D_β's to themselves and equal the identity off of the D_β's. This follows from the construction; Bing obtains T_i' by sliding the D_α's around, $\alpha \in \mathcal{S}_{j_i}$, without ever doing anything outside the D_β's. (The point is to slide the D_α's around to make them have very small diameter.) This sequence $\{T_i\}$ converges to a mapping T which induces a homeomorphism from $\mathbb{R}^3/G$ onto $\mathbb{R}^3$. Indeed, by construction $T_l = T_i$ on the complement of the D_α's for $\alpha \in \mathcal{S}_{j_i}$ when $l \geq i$, and $T_l(D_\alpha) = T_i(D_\alpha)$ has diameter less than $1/i$

for each $\alpha \in \mathcal{S}_{j_i}$, and these properties imply that $\{T_i\}$ converges to a mapping which induces a homeomorphism from $\mathbb{R}^3/G$ onto $\mathbb{R}^3$.

To get the deformation described in Lemma 8.6 we take f_0 to be the identity, f_1 to be T, f_t to be T_i when $t = i/(i+1)$, and on $(i-1)/i < t < i/(i+1)$ we deform T_{i-1} into T_i through homeomorphisms by deforming the identity to T_i' through homeomorphisms which map each D_α to itself for $\alpha \in \mathcal{S}_{j_i}$ and which equal the identity off these D_α's. (For $i = 1$ this makes sense with T_0 taken to be the identity.) It is not hard to check that this deformation has the properties described in Lemma 8.6.

Proposition 8.5 is a straightforward consequence of Lemma 8.6. Let $b > 0$ be such that $\overline{D}_i \times [-b, b] \subseteq \omega_i$ for $i = 1, 2$. We define ξ by setting it to be the identity when $|x_4| > b$ and by taking ξ to be the obvious copy of f_t on the $x_4 = \pm b(1-t)$ 3-planes when $0 \le t \le 1$. That is, ξ should take these 3-planes to themselves, and be the same as f_t modulo the obvious vertical translation down to P and back. It is easy to check that this choice of ξ has the right properties. This proves Proposition 8.5.

To prove b) in Theorem 1.2 it is enough to produce a mapping $\zeta : \mathbb{R}^4 \to \mathbb{R}^4$ which agrees with θ outside $\omega_1 \cup \omega_2$, maps P onto M, and induces a homeomorphism from $\mathbb{R}^4/G'$ onto $\mathbb{R}^4$. Indeed, if we can build such a mapping ζ, then the desired ν will result from Proposition 8.5. We would like to simply take $\zeta = h$, where h is as in Lemma 3.21, but it is not completely clear that h has the right properties on $\mathbb{R}^4$, *i.e.*, $h^{-1}(\Omega_\alpha)$ might leak out into $\mathbb{R}^4 \backslash P$ further than we want even when $\alpha \in \mathcal{S}_l$ for l large. Rather than attempt some fine analysis we shall modify h a little bit brutally to avoid this problem. Note that h has some nice self-similarity properties that we do not need to replicate in ζ.

Given $\alpha \in \mathcal{S}_l$ define ω_α to be $h^{-1}(\Omega_\alpha)$. This is compatible with the choice of ω_j's in our initial package, because of Lemma 3.21 and (3.6), and we could have defined the ω_α's through the same kind of recursive constructions as in Section 3, but this amounts to the same thing as the present definition. Set $C_l^* = \cup_{\alpha \in \mathcal{S}_l} \overline{\omega}_\alpha$, in analogy with the definition of the defining sequence $\{C_l\}$ associated to the D_α's (just before (3.11)). Set $C^* = \cap_l C_l^*$, and let G^* denote the decomposition of $\mathbb{R}^4$ associated to the defining sequence $\{C_l^*\}$ in the manner described just after Definition 2.2.

Lemma 8.7. *h is constant on each element of G^*, and it induces a homeomorphism from $\mathbb{R}^4/G^*$ onto $\mathbb{R}^4$.*

The proof of this is practically the same as for the corresponding statement for $\mathbb{R}^3/G$ in Lemma 3.21. Let us briefly review the highlights. The ω_α's (and their closures) enjoy the same nesting properties as do the Ω_α's (as described in Lemma 3.10). Let $\mathcal{S}$ be, as before (just after (3.11)), the set of infinite sequences which take values in $\{1, \dots, n\}$. To each element s of $\mathcal{S}$ we associate a set A_s^* which is the intersection of the $\overline{\omega}_\alpha$'s which correspond to the ancestors $\alpha \in \mathcal{S}_l$ of the sequence s, just as for the D_α's (before Sublemma 3.40).

Sublemma 8.8. *$C^* = \bigcup_{s \in \mathcal{S}} A_s^*$, and the A_s^*'s are the connected components of C^*.*

This is the analogue of Sublemma 3.40 for the ω_α's, and the same proof works here, except that we should verify that the $\overline{\omega}_\alpha$'s are connected. A priori we have a problem, since h is not a homeomorphism, but in fact we have that

$$\omega_\alpha = h_l^{-1}(\Omega_\alpha), \quad \text{for all } \alpha \in \mathcal{S}_l, \tag{8.9}$$

where h_l is as in the proof of Lemma 3.21. This equality is an easy consequence of the definitions in the proof of Lemma 3.21, but let us write it out. Notice first that

$$h \circ h_j^{-1} = \text{the identity on } \mathbb{R}^4 \setminus \bigcup_{\delta \in \mathcal{S}_j} \Omega_\delta, \tag{8.10}$$

and

$$h \circ h_{l+1}^{-1}(\overline{\Omega}_\beta) = \overline{\Omega}_\beta, \quad \text{for all } \beta \in \mathcal{S}_{l+1}. \tag{8.11}$$

These follow immediately from (3.28), (3.29), and the definition of h as the limit of the h_k's. (Actually, one should think a little about the inclusion "$\supseteq$" in (8.11). To derive this inclusion from its counterpart in (3.29) one can use the compactness of $\overline{\Omega}_\beta$ and the uniform convergence of the h_k's to h.) Even though h is not a homeomorphism we can convert (8.11) into $(h \circ h_{l+1}^{-1})^{-1}(\overline{\Omega}_\beta) = \overline{\Omega}_\beta$ using (8.10) and the fact that the $\overline{\Omega}_\beta$'s, $\beta \in \mathcal{S}_{l+1}$, are disjoint (Lemma 3.10). Since $\overline{\Omega}_\beta \subseteq \Omega_\alpha$ when β is a child of α (by (3.9)) we conclude that

$$(h \circ h_{l+1}^{-1})^{-1}(\Omega_\alpha) = \Omega_\alpha \quad \text{when } \alpha \in \mathcal{S}_l \tag{8.12}$$

using also (8.10) again. From (3.27) we have that $h_{l+1} = g_l \circ h_l$, where g_l is as in the proof of Lemma 3.21, and from (3.25) we obtain $g_l(\Omega_\alpha) = \Omega_\alpha$ when $\alpha \in \mathcal{S}_l$. Thus (8.12) yields $(h \circ h_l^{-1})^{-1}(\Omega_\alpha) = (h \circ h_{l+1}^{-1} \circ g_l)^{-1}(\Omega_\alpha) = \Omega_\alpha$. This proves (8.9).

It follows that the $\overline{\omega}_\alpha$'s are connected (since each h_l is a homeomorphism), which is the fact we needed for Sublemma 8.8.

Recall the (bijective) mapping $f : \mathcal{S} \longrightarrow F$ defined just after (3.11). We have that

$$h(p) = f(s)\,, \qquad \text{for all } p \in A_s^* \text{ and } s \in \mathcal{S}\,. \tag{8.13}$$

This is the analogue of (3.41) in this situation, and it follows from chasing definitions. Because of Sublemma 8.8 this says exactly that h is constant on each component of C^*, and hence on the nontrivial elements of G^*. Thus h induces a continuous map from $\mathbb{R}^4/G^*$ into $\mathbb{R}^4$. It is not hard to see that it is actually a bijection and even a homeomorphism, using (8.13), (8.10), (8.11), (3.13), and the fact that the h_j's are all homeomorphisms. This proves Lemma 8.7.

Because of Lemma 8.7, the proof of Theorem 8.1.b) comes down to showing that the decompositions G^* and G' of $\mathbb{R}^4$ are equivalent in a way which does not disturb points in P. Let us first check that they "agree" on P. To do this we begin with the observation that

$$\omega_\alpha \cap P = D_\alpha\,, \qquad \text{for all } \alpha \in \mathcal{S}_l \tag{8.14}$$

(and any l). By (8.9) the left side is the same as $h_l^{-1}(\Omega_\alpha) \cap P$, and since $h_l(P) = M^l$ this is the same as $h_l^{-1}(\Omega_\alpha \cap M^l)$. Since $h_l(D_\alpha) = \Omega_\alpha \cap M^l$, by Lemma 3.21, we get (8.14). Thus if C is as in Sublemma 3.40 and C^* is as in Sublemma 8.8, then $C = C^* \cap P$, because the corresponding statement is true for the defining sequences $\{C_l\}$ and $\{C_l^*\}$, by (8.14). Moreover, if A_s and A_s^* are as in Sublemmas 3.40 and 8.8, then

$$A_s = A_s^* \cap P\,, \qquad \text{for all } s \in \mathcal{S}\,, \tag{8.15}$$

by the definitions of A_s and A_s^*. This makes precise the sense in which G^* and G' "agree" on P, since the A_s's and A_s^*'s are the only nontrivial elements of these decompositions.

Lemma 8.16. *There is a homeomorphism σ from $\mathbb{R}^4/G^*$ onto $\mathbb{R}^4/G'$ which "equals the identity" on P and on the complement of $\omega_1 \cup \omega_2$, and which sends ω_i/G^* to ω_i/G' for $i = 1, 2$.*

The statement of the lemma is an abuse of language whose meaning is hopefully clear to the reader. It is "justified" by (8.15) and the fact that both the decompositions G^* and G' are trivial (consisting only of singletons) on the complement of $\omega_1 \cup \omega_2$.

To prove Lemma 8.16 we shall construct a sequence of diffeomorphisms $\{\sigma_l\}_{l=1}^{\infty}$ on $\mathbb{R}^4$ with the following properties (for all l):

(8.17) $$\sigma_l = \text{the identity on } P \cup \left(\mathbb{R}^4 \backslash (\omega_1 \cup \omega_2)\right), \text{ and} \quad \sigma_l(\omega_i) = \omega_i \text{ , for } i = 1, 2\,;$$

(8.18) $$\sigma_{l+1} = \sigma_l \quad \text{on } \mathbb{R}^4 \backslash \bigcup_{\alpha \in \mathcal{S}_l} \omega_\alpha\,;$$

(8.19) $$\sigma_{l+1}(\omega_\alpha) = \sigma_l(\omega_\alpha)\,, \qquad \text{for all } \alpha \in \mathcal{S}_l\,;$$

(8.20) $$\text{every point in } \sigma_l(\overline{\omega}_\alpha) \text{ lies within } l^{-1} \text{ of } \overline{D}_\alpha\,, \quad \text{for all } \alpha \in \mathcal{S}_l \text{ when } l \geq 2\,.$$

We take σ_1 to be the identity, and to construct the later σ_l's we use an iterative construction based on the following.

Sublemma 8.21. *For each $\varepsilon > 0$ and any $\alpha \in \mathcal{S}_l$ (with l arbitrary) we can find a diffeomorphism of $\mathbb{R}^4$ onto itself which equals the identity on P and on the complement of ω_α and which maps $\overline{\omega}_\beta$ to a set which lies within ε of $\overline{D}_\beta$ for each of the two children β of α in $\mathcal{S}_{l+1}$.*

If $l = 0$, so that $\omega_\alpha = \Omega$ and the children of ω_α are simply ω_1 and ω_2, then this sublemma follows from the way that Ω, ω_1, and ω_2 were chosen (as smooth solid 4-dimensional tori which are cut in half by the 3-plane P in the standard way, etc.). In other words, we can just shrink the ω_i's in Ω as close to the D_i's as we want, without disturbing any points in P or outside of Ω. One can do this rather explicitly, but one could also use Lemma 4.1 (or rather a small variant of it).

For an arbitrary $\alpha \in \mathcal{S}_l$ we can reduce to the preceding case as follows. We know from (8.9) that $h_l(\omega_\alpha) = \Omega_\alpha = \psi_\alpha(\Omega)$, and we have that $h_l(\omega_\alpha \cap P) = h_l(D_\alpha) = \Omega_\alpha \cap P = \psi_\alpha(\Omega \cap P)$ by (8.14), (3.31), (3.8), and (3.7). We want to show that $h_l(\omega_\beta)$ is the same as the image under ψ_α of ω_1 or ω_2 when β is a child of α. If β is a child of α, then $\omega_\beta = h_{l+1}^{-1}(\Omega_\beta)$ by (8.9), and so $h_l(\omega_\beta) = h_l \circ h_{l+1}^{-1}(\Omega_\beta) = g_l^{-1}(\Omega_\beta)$ by the definition (3.27) of h_{l+1}, where g_l is as in the proof of Lemma 3.21. We also have $g_l^{-1}(\Omega_\beta) = \theta_\alpha^{-1}(\Omega_\beta)$ because of the definition of g_l (just before (3.25)). The fact that this is the same as the image under ψ_α of ω_1 or ω_2 follows from (3.22) and (3.6), using also the definitions of Ω_β and ψ_α (just before (3.7)) to get that $\psi_\alpha^{-1}(\Omega_\beta)$ is one of Ω_1 or Ω_2. Now that we understand $h_l(\omega_\beta)$ we can reduce Sublemma 8.21 for ω_α to the $l = 0$ case of the preceding paragraph using h_l. Note that this reduction of the general case to Ω involves mappings with large distortions, so that we are not getting good estimates (as a function of l and ε), but we do not care. This proves Sublemma 8.21.

It is now a simple matter to construct the σ_l's recursively. Given σ_l, one builds σ_{l+1} by composing σ_l on the right with diffeomorphisms on each of the ω_α's, $\alpha \in \mathcal{S}_l$, which are chosen as in Sublemma 8.21. Because these diffeomorphisms equal the identity on P and outside the corresponding ω_α's the properties (8.17), (8.18), and (8.19) are maintained. We get (8.20) for σ_{l+1} simply by choosing ε in Sublemma 8.21 to be suitably small. This choice of ε will depend on the modulus of continuity of σ_l, but we do not mind.

Once we have the σ_l's we simply take σ to be their limit as $l \to \infty$. This makes sense as an $\mathbb{R}^4$-valued mapping only as long as we stay away from C^*, but in fact the limit exists as a map from $\mathbb{R}^4/G^*$ onto $\mathbb{R}^4/G'$ because of (8.20). It is not hard to check that this choice of σ has the required properties. This proves Lemma 8.16.

Because of Lemmas 8.16 and 8.7 we can form the map $\zeta = h \circ \sigma^{-1}$ which gives a homeomorphism from $\mathbb{R}^4/G'$ onto $\mathbb{R}^4$. Lemmas 8.16 and 3.21 imply that ζ agrees with θ outside $\omega_1 \cup \omega_2$ and that it maps P onto M. By composing ζ with the inverse of the map promised in Proposition 8.5 we get a homeomorphism ν as in Theorem 8.1.b). This completes the proof of Theorem 8.1.

9. Analysis on these sets.

To what extent can we do analysis on the sets M, M^j, and $\widetilde{M}$? Because the geometry of these sets is approximately Euclidean we can hope that much of the usual analysis on Euclidean spaces also works on these sets. For this we should remember to require that ρ in Definition 3.2 satisfy $\rho^3 n < 1$, so that (3.46) and the conclusions of Lemma 3.45 hold.

We can get a lot of milage out of the fact that the M^j's are all smooth, M is smooth away from a compact singular set with 3-dimensional measure zero, and $\widetilde{M}$ is smooth away from a single point. For instance we can talk about smooth functions (away from the singular sets) and we can use Stokes' theorem (if we are a little bit careful about the singularities). We also have automatically the "local" results in real analysis, like the existence almost everywhere of Lebesgue points of locally integrable functions, points of density of measurable sets, and derivatives of Lipschitz functions. We can get more quantitative results from real analysis using the fact that our sets are Ahlfors regular and hence "spaces of homogeneous type" in the sense of [CW1], [CW2], with uniform bounds in the case of the M^j's. The uniform rectifiability of these sets implies additional analytical information, as in [Da], [DS2], [DS3], [DS4]. Of course the M's are much nicer than most regular or uniformly rectifiable sets, because they degenerate only near small sets, and in a moderate way.

What about Sobolev and Poincaré inequalities? To what extent can we control a function on M in terms of its gradient? If these sets were quasisymmetrically equivalent to $\mathbb{R}^3$ then they would have to be well behaved for Sobolev and Poincaré inequalities. Indeed, if a quasisymmetric parameterization were to exist, it would distort Hausdorff measure by only an "A_∞ weight", by the method of Gehring [Ge] (see also [DS1], [Se3]), and one could then obtain Sobolev and Poincaré inequalities from the results in [DS1] (since the A_∞ weight in question would have to be a strong A_∞ weight). This approach is not available here, but we can verify Sobolev and Poincaré inequalities directly using the bilipschitz balls provided by Proposition 3.70, as we shall do in the next section. (One could do with less than these bilipschitz balls, but there is not much point in that here. The method of [Se1] (see also [DS3, Section 6]) would work too, but it is unnecessarily indirect for the present situation.)

It would be interesting to find nice ways to see the strangeness of the M's in terms of analysis on them or on their complementary domains, *e.g.*, in terms of harmonic functions, or Clifford holomorphic functions (see [BDS]), or nonlinear potential theory.

10. Sobolev and Poincaré inequalities.

The main result of this section is that the sets M, M^j, or $\widetilde{M}$ from Section 3 satisfy Sobolev and Poincaré inequalities under mild assumptions. This will be a rather simple consequence of Proposition 3.70. Much of the structure of excellent packages will not really be needed, as in the comments which follow Definition 3.2, and there is nothing special about the dimension 3. For that matter, we could do with less than the bilipschitz balls provided by Proposition 3.70, but since we have them we may as well use them.

Let us begin with some general definitions. Let us call a subset of $\mathbb{R}^4$ a "singular submanifold" if it is a smooth embedded 3-dimensional submanifold away from a singular set which is closed and has zero 3-dimensional Hausdorff measure. This definition accommodates the sets M and $\widetilde{M}$, while the M^j's are everywhere smooth already. All of our integrals, L^p norms, etc., on singular submanifolds will be taken with respect to 3-dimensional Hausdorff measure, which we shall denote by "dx" or something similar. If f is a locally integrable function on a singular submanifold then its differential df makes sense as a distribution (or more precisely a current, as in [Fe]) which is defined away from the singular set. As a practical matter one should not take this business about currents too seriously here, because we shall always work with the differential on the smooth part of our set, in such a way that all the computations can be reduced to similar problems on a piece of $\mathbb{R}^3$ using local coordinates. We shall always assume that df is locally integrable away from the singular set, so that we can use the Euclidean metric to define $|df|$ and hence L^p norms and other integrals of $|df|$. When we say that "df is locally integrable" on our singular submanifold we shall mean that $|df|$ is actually locally integrable across the singular set, and hence has an unambiguous locally integrable extension to the whole singular submanifold, since the singular set has measure zero. This extension may differ from the natural distributional definitions of df near the singular set -*i.e.*, we could be dropping Dirac masses or other singular contributions to the "true" df- but we do not care. We shall always do our distribution theory away from the singular set and then extend brutally across the singular set.

Proposition 10.1. *Suppose that $E = M, \widetilde{M}$, or M^j for some j, that ρ in Definition 3.2 satisfies $\rho^3 n < 1$, and that our excellent package satisfies (3.69). There exist constants L_1 and C_1, depending only on our excellent package, so that if $p, q \in E$, f is a locally integrable function* *on* $E \cap B(p, L_1|p-q|)$, *p and q are Lebesgue points for f, and df is locally integrable on* $E \cap B(p, L_1|p-q|)$ (*in the sense just described when* $E = M$ *or* $\widetilde{M}$), *then*

$$(10.2)\quad |f(p) - f(q)| \leq C_1 \int_{E \cap B(p, L_1|p-q|)} \left(\frac{1}{|p-z|^2} + \frac{1}{|q-z|^2} \right) |df(z)|\, dz\,.$$

The statement that p is a Lebesgue point for f means that

$$(10.3)\qquad \lim_{r \to 0} r^{-3} \int_{E \cap B(p,r)} |f(x) - f(u)|\, du = 0\,.$$

The "du" here refers to 3-dimensional Hausdorff measure, as indicated above. Almost all points are Lebesgue points.

To prove Proposition 10.1 let us first review the situation for $\mathbb{R}^3$. Suppose that h is a smooth function on a ball B in $\mathbb{R}^3$ with radius r. Then

$$(10.4)\qquad r^{-6} \int_B \int_B |h(u) - h(v)|\, du\, dv \leq C\, r^{-3} \int_B r\, |dh(w)|\, dw\,,$$

where C does not depend on h, B, or r. The proof of this is quite easy. We can express $h(u) - h(v)$ in terms of an integral of dh over the line segment which connects u to v, and (10.4) is obtained by averaging this formula over all u and v and applying Fubini's theorem. This Poincaré-type inequality (10.4) also holds when h is merely locally integrable and has locally integrable distributional first derivatives, because of standard approximation arguments (as in the proof of [St, p. 122, Proposition 1]).

Now suppose that x and y lie in $\overline{B}$, h is locally integrable and has locally integrable first derivatives on B, and that x and y are Lebesgue points for h in the sense of (10.3) with $E = B$. (If x or y lies in ∂B, so that they are not in the putative domain of h, then we simply assume that h is also defined at x and y in such a way that (10.3) holds.) Then

$$(10.5)\qquad |h(x) - h(y)| \leq C \int_B \left(\frac{1}{|x-z|^2} + \frac{1}{|y-z|^2} \right) |dh(z)|\, dz\,.$$

This is also well-known, but let us quickly go through a proof. Let us assume for simplicity that B is centered at the origin and has radius one. Set $B_k(x) = B((1-2^{-k})x, 2^{-k})$ for $k \geq 0$, and define $B_k(y)$ similarly. Thus $B_0(x) = B_0(y) = B$ and $B_{k+1} \subseteq B_k \subseteq B$ for all k (by the triangle inequality). If we let $\mathrm{Av}\,(h, Z)$ denote the average of h over the subset Z of B (with positive measure), then our assumption that x and y are Lebesgue points implies that $\lim_{k\to\infty} \mathrm{Av}\,(h, B_k(x)) = h(x)$ and similarly for y. Hence

$$
\begin{aligned}
|h(x) - h(y)| \leq & \sum_{k=0}^{\infty} |\mathrm{Av}\,(h, B_k(x)) - \mathrm{Av}\,(h, B_{k+1}(x))| \\
& + \sum_{k=0}^{\infty} |\mathrm{Av}\,(h, B_k(y)) - \mathrm{Av}\,(h, B_{k+1}(y))|\,,
\end{aligned} \tag{10.6}
$$

since $B_0(x) = B_0(y)$. We also have that

$$
|\mathrm{Av}\,(h, B_k(x)) - \mathrm{Av}\,(h, B_{k+1}(x))| \leq C \int_{B_k(x)} 2^{-2k}\, |dh(z)|\, dz\,, \tag{10.7}
$$

and similarly with x replaced by y, because of (10.4) applied to B_k. It is easy to derive (10.5) from these inequalities.

Proposition 10.1 follows immediately from (10.5) and Proposition 3.70. We are also using here the comment made shortly after the statement of Proposition 3.70 to the effect that the bilipschitz 3-ball W in Proposition 3.70 can be chosen to be smooth away from p and q, and the bilipschitz parameterization of W can be taken to be smooth away from these points. This permits us to avoid technical issues concerning the distribution theory. Other than that we are simply using the bilipschitz invariance of (10.2) and (10.3) in a brutal way.

Proposition 10.8. *Suppose that $E = M, \widetilde{M}$, or M^j for some j, that ρ in Definition 3.2 satisfies $\rho^3 n < 1$, and that our excellent package satisfies (3.69). Then there exist constants L_2 and C_2, depending only on our excellent package, so that if B is a ball with radius r centered on E and f is a locally integrable function on $E \cap L_2 B$ such that df is also locally integrable there (in the sense described before Proposition 10.1 when $E = M$ or $\widetilde{M}$), then we have the Poincaré inequality*

$$
r^{-6} \int_{E\cap B} \int_{E\cap B} |f(x) - f(y)|\, dx\, dy \leq C_2\, r^{-3} \int_{E\cap L_2 B} r\, |df(z)|\, dz\,. \tag{10.9}
$$

If f is a locally integrable function on E with compact support and if df is locally integrable on E, then

$$|f(x)| \leq C \int_E \frac{1}{|z-x|^2} \, |df(z)| \, dz \,, \tag{10.10}$$

for some C (depending only on our excellent package) and almost all $x \in E$.

To prove this let us first check that

$$\int_{B(z,t)\cap E} \frac{1}{|x-z|^2} \, dx \leq C\, t \,, \tag{10.11}$$

for all $z \in E$ and $t > 0$. Using the Ahlfors regularity of E we get that

$$\int_{(B(z,2s)\setminus B(z,s))\cap E} \frac{1}{|z-x|^2} \, dx \leq C\, s \,, \tag{10.12}$$

for all $s > 0$, and (10.11) follows from this by summing the obvious geometric series.

Once we have (10.11) we can derive (10.9) from Proposition 10.1 by simply averaging (10.2) over p and q and using Fubini's theorem. The pointwise inequality (10.10) is an immediate consequence of (10.2). (Just take y to be far far away.)

Let us now check that Sobolev embeddings work for the sets E as in Proposition 10.8. Consider first the potential operator I_1 on functions on E defined by

$$I_1(g)(x) = \int_E \frac{1}{|x-z|^2} \, g(z) \, dz \,. \tag{10.13}$$

This operator has exactly the same $L^p \to L^q$ mapping properties on any regular set of dimension 3 as on $\mathbb{R}^3$ itself, *i.e.*, it maps $L^p(E)$ into $L^q(E)$ when $1 < p < 3$ and $1/q = 1/p - 1/3$. This is not hard to prove -the point is that the two situations are essentially the same at the level of this kind of measure theory- and one can simply mimic the proof of the corresponding result on $\mathbb{R}^3$ ([St, p. 119, Theorem 1]). This is really just a minor variation on [St, p. 121, Comment 1.4]. Alternatively, one could invoke theorems from the real method of interpolation of Banach spaces or other general results.

The usual Sobolev embedding theorems on E follow immediately from the $L^p \to L^q$ mapping properties for the potential operator I_1 and (10.10), at least for $p > 1$. This method breaks down for $p = 1$, but one can establish isoperimetric inequalities for these sets using the Poincaré inequality (10.9) and a covering lemma. (A very similar argument was used in [DS1].)

If df lies in L^p for some $p > 3$, then one can modify f on a set of measure zero to get a function which is Hölder continuous of order $1 - 3/p$. This is an easy consequence of (10.2) and Hölder's inequality. If $df \in L^3$ then f lies in $\mathrm{BMO}(E)$, because of (10.9).

The bottom line is that Propositions 10.1 and 10.8 permit us to verify that many of the usual results about functions on $\mathbb{R}^3$ which satisfy Sobolev space conditions also work on M, $\widetilde{M}$, and the M^j's. Of course the preceding list is not exhaustive.

11. A remark about polyhedra.

Proposition 11.1. *A finite d-dimensional polyhedron P (in some $\mathbb{R}^n$) which is a topological manifold (without boundary) satisfies the analogue of (**) for compact sets (i.e., the conditions of Definition 1.7 hold when r is sufficiently small).*

These polyhedra can be pretty strange, because of the results of Edwards and Cannon on the double suspensions of spheres (as mentioned in Sections 1 and 2).

Proposition 11.1 is no surprise, but it seems worthwhile to record it in view of the gap between (**) and the existence of a quasisymmetric parameterization established by Theorem 1.12, and since we know that there are interesting examples of these polyhedra.

Let us prove the proposition. Let P be as above, and note that P must have pure dimension d. Since P is a polyhedron we can give it the structure of a simplicial complex. That is, we can realize P as a (finite) union of d-dimensional simplices in such a way that the collection of all these simplices together with all their faces (of any dimension, including 0 (vertices)) have the property that when any two of them intersect, either one is a face of the other or the two intersect in a common face. We shall call a simplex in P "distinguished" if it is one of these basic simplices or one of their faces (as opposed to a random simplex floating around in $\mathbb{R}^n$). All simplices in this discussion are assumed to be closed, and ∂A will be used to denote the geometric (or simplicial) boundary of A (as opposed to the uninteresting topological boundary of A as a subset of $\mathbb{R}^n$). If A is just a vertex, then we interpret ∂A to be the empty set.

Finite polyhedra obviously have a lot of homogeneity to them, and the proof of Proposition 11.1 merely requires a precise formulation of this homogeneity. We begin with a preliminary fact.

Let A, X be distinguished simplices in P, with $A \subseteq X$. Then A is a face of X, and we can order the vertices $v_1, \ldots, v_l$ of X in such a way that $v_1, \ldots, v_j$ are the vertices of A. Thus A is the convex hull of $v_1, \ldots, v_j$ and X is the convex hull of $v_1, \ldots, v_l$. The $j-1$-plane $Q(A)$ determined by A can be described as the set of points of the form $\sum_{i=1}^{j} \lambda_i v_i$ where $\sum_i \lambda_i = 1$ and the λ_i's are real numbers. (Points in A correspond to restricting ourselves to λ_i's which are nonnegative.) Let $Q_0(A)$ denote the set of points of the form $\sum_{i=1}^{j} \lambda_i v_i$ where $\sum_i \lambda_i = 0$. This is just a translation of $Q(A)$ which contains the origin. Notice that $Q(A)$ is preserved by translations by elements of $Q_0(A)$.

Let $H(A, X)$ be the set of points of the form $\sum_i \lambda_i v_i$ with $\sum_i \lambda_i = 1$ and $\lambda_i \geq 0$ when $i > j$. Notice that $X \subseteq H(A, X)$. If A is just a vertex then $H(A, X)$ is the cone with vertex A generated by X, while in general it is a product of the plane determined by A with a cone. Note that $H(A, X)$ is preserved by translations of elements of $Q_0(A)$. One can also check that $H(A, X)$ is preserved by dilations by positive constants centered at elements of $Q(A)$, *i.e.*, mappings of the form $z \mapsto a + t(z - a)$ for $a \in Q(A)$ and $t > 0$. We shall eventually use these symmetries of $H(A, X)$ to make precise the homogeneity of P, but first we establish the following.

Lemma 11.2. *Let A, X be as in the preceding paragraphs, and let $a \in A$ and $p \in H(A, X) \backslash X$ be given. Then there is a point x on the line segment that joins a and p with the property that x lies in a face of X which does not contain (all of) A. (It could be that $x = a \in \partial A$).*

By assumptions, $a = \sum_i \alpha_i v_i$ and $p = \sum_i \lambda_i v_i$, where $\sum_i \alpha_i = \sum_i \lambda_i = 1$, $\alpha_i \geq 0$ when $i \leq j$, $\alpha_i = 0$ when $i > j$, and $\lambda_i \geq 0$ when $i > j$. Choose t to be the first real number ≥ 0 such that $(1-t)\alpha_i + t\lambda_i = 0$ for some $i \leq j$ ($t = 0$ is possible), and set $x = \sum_i ((1-t)\alpha_i + t\lambda_i) v_i$. Our choice of t implies that $(1-t)\alpha_i + t\lambda_i \geq 0$ for all $i \leq j$, and this is true when $i > j$ as well, by our assumptions. It is easy to check that x lies in a face of X which does not contain A. Also, $t < 1$, because

$p \notin X$ (and hence $\lambda_i < 0$ for some $i \leq j$). This proves the lemma.

Given a distinguished simplex A in P let $S(A)$ denote the star of A, *i.e.*, the union of all the distinguished simplices in P which contain A as a face. Because P has pure dimension d one may simply take the union of the d-dimensional distinguished simplices which contain A as a face. (The d-dimensional ones will contain all the others.) If A is a vertex then $S(A)$ is a neighborhood of A (in P) with nice properties. If A has positive dimension, then every point in $A \backslash \partial A$ lies in the interior of $S(A)$ (relative to P), but this is not necessarily true for points in ∂A. Let $H(A)$ be the union of $H(A, X)$ for all distinguished simplices X in P which contain A, and note that $S(A) \subseteq H(A)$.

For a distinguished simplex A in P let $C(A)$ denote the union of all distinguished simplices X which do not contain A. Thus $C(A)$ contains ∂A in particular, and $C(A)$ is approximately the same as the complement of $S(A)$. In fact $C(A)$ is the union of the complement of $S(A)$ and the faces of the d-simplices in $S(A)$ which do not contain A.

There is a constant k such that

$$\operatorname{dist}(a, C(A)) \geq k \operatorname{dist}(a, \partial A), \tag{11.3}$$

for any distinguished simplex A in P and all $a \in A$. This comes down to the fact that if X is a distinguished simplex in P, and if X does not contain A, then either X is disjoint from A, X is contained in ∂A, or X meets A in a face of ∂A and makes a definite angle with A. Keep in mind also that there are only finitely many A's and X's around, so that we can easily choose k to be independent of them. When A is a vertex the correct version of (11.3) is that $\operatorname{dist}(A, C(A))$ is bounded from below.

Let us call a ball $B(z, r)$ centered on P "good" with respect to a distinguished simplex A in P if $z \in A$ and $B(z, r) \cap C(A) = \varnothing$. The key property of a good ball is that

$$B(z, r) \cap P = B(z, r) \cap H(A) \tag{11.4}$$

when $B(z, r)$ is good relative to A. Clearly $B(z, r) \cap P = B(z, r) \cap S(A)$ when $B(z, r)$ is good, and so we get one inclusion in (11.4) from the fact that $S(A) \subseteq H(A)$. If we have a point p in $B(z, r) \cap H(A)$ which does not lie in $S(A)$, then Lemma 11.2 produces a point x on the line segment which joins p and z (so that $x \in B(z, r)$) with the property that x lies in a distinguished simplex in P which does not contain A. This means that $x \in C(A)$, which contradicts the assumption that $B(z,r)$ is good. This proves (11.4).

If $B(y,s)$ is another ball which is good relative to A then the obvious translation and dilation which sends $B(z,r)$ onto $B(y,s)$ must send $B(z,r) \cap P$ onto $B(y,s) \cap P$. This follows from (11.4), because each $H(A,X)$ is preserved by this translation and dilation (see the paragraph just before Lemma 11.2), and so $H(A)$ is also preserved by these mappings. Thus the intersection of a good ball with P is equivalent (by a translation and a dilation) to one of finitely many models. To prove Proposition 11.1 we need to show that we can ignore the bad balls.

Lemma 11.5. *Given any constant $L > 0$, there exist positive constants K, ε (depending on L and P) with the property that if $x \in P$ and $0 < r < \varepsilon$ then there is a ball $B(\xi, t)$ such that $B(\xi, Lt)$ is good with respect to some simplex, $B(\xi,t) \supset B(x,r)$, and $t \leq Kr$.*

Let x and r be given, and let A be a distinguished simplex in P which contains x. The ball $B(x, Lr)$ is itself good if $Lr \leq k \operatorname{dist}(x, \partial A)$, where k is as in (11.3), and so we assume that this inequality is not true. This means that there is a point $y \in \partial A$ such that $B(x,r) \subset B(y,Cr)$ for a suitable constant $C = C(L)$. Thus y lies in a lower-dimensional simplex, and we can repeat the argument to conclude that either $B(y, LCr)$ is good or $B(y,Cr)$ is contained in $B(z, C^2 r)$ for some z in a lower dimensional simplex. Repeating this as many times as necessary (but at most d times) we reduce to the case of vertices. A ball centered at a vertex is good as soon as its radius is small enough. Lemma 11.5 follows easily from this.

Let us analyze the the structure of the $H(A)$'s some more. For each distinguished simplex A in P we have that $H(A)$ is a topological manifold. Indeed, near a point in $A \backslash \partial A$ $H(A)$ looks like P, because of (11.4) and the existence of good points, and therefore $H(A)$ is a topological manifold near such points. This implies that all points of $H(A)$ are manifold points, because $H(A)$ is invariant under translations by elements of $Q_0(A)$ and dilations centered at points in $Q(A)$ (as discussed just before Lemma 11.2). In fact we get that there is a constant L so that if $a \in A$ and $r > 0$ then there is a relatively open set U of $H(A)$ with $B(x,r) \cap H(A) \subseteq U \subseteq B(x, Lr) \cap H(A)$ such that U is a topological d-ball and its closure is homeomorphic to the closed unit d-ball. This follows easily from the invariance properties of $H(A)$ just mentioned. (That is, we choose U once for some fixed ball, and then we use the translations and dilations to extend this choice to other balls.) We can also choose these sets U so that they are all translates and dilates of each other for a fixed A. Hence there are only finitely many models total, since there are only finitely many A's. We can also choose the constant L so that it does not depend on A, since there are only finitely many A's.

We can now finish the proof of Proposition 11.1. Let $x \in P$ and $r > 0$ be given. Lemma 11.5 implies that if r is small enough, then there is a ball $B(\xi, t)$ such that $B(\xi, Lt)$ is good with respect to some simplex A, $B(\xi, t) \supset B(x, r)$, and $t \leq Kr$. From the observations of the preceding paragraph we obtain that there is relatively open set U in $H(A)$ which is a topological d-ball and satisfies $B(\xi, t) \cap H(A) \subseteq U \subseteq B(\xi, Lt) \cap H(A)$. Since $B(\xi, Lt)$ is good we actually have that $B(\xi, Lt) \cap H(A) = B(\xi, Lt) \cap P$. As mentioned above, we can also choose U so that its closure is homeomorphic to a closed ball and so that U can be reduced by translations and dilations to one of a finite set of models. This implies that P satisfies the compact version of (**), as promised. (Of course it is the finiteness of the set of the models for U which gives us the uniform estimates as in (**), as opposed to (*). Do not forget that continuous maps between compact sets are uniformly continuous.)

12. Concluding remarks.

Geometric topology provides a lot of technology for building homeomorphic parameterizations of a set, and it provides many interesting examples, but little is known about quantitative estimates on the geometric complexity of these parameterizations. The most basic example that I know is the type of mapping constructed by Bing in [B1], for which optimal estimates are (apparently) still unknown (but see [B5] and the remarks at the end of Section 2 of [FS]). The construction in Section 3 gives an interesting formulation to this problem, namely, what kind of estimates can be satisfied by a homeomorphic parameterization of the set M from Section 6 by $\mathbb{R}^3$? We know from [B1] that such a parameterization exists, and we know that it cannot be quasisymmetric, but it seems to be unknown whether M can be parameterized by a Hölder continuous map whose inverse is also Hölder continuous (for instance), let alone the possible range of orders of Hölder continuity (if any). I am also unaware of any estimates (other than uniform continuity) for the homeomorphisms between the strange polyhedral spheres of Edwards and Cannon and standard Euclidean spheres.

One can wonder whether Condition (**), even in combination with Ahlfors regularity, implies the existence of local homeomorphic parameterizations with a universal choice (or family of choices) of modulus of continuity ω as in Definition 1.7. For instance, do Condition (**) and Ahlfors regularity imply the existence of homeomorphic local parameterizations which are Hölder continuous, of some universal order, or even of just some positive order? This leads back to questions about the type of mapping produced in [B1] for the set M from Section 6, or the kind of local coordinates that might exist for the strange polyhedral spheres of Edwards and Cannon. (I am not optimistic.) Note that there is a positive result when $d = 2$, because of Theorem 1.6.

One can also ask whether the combination of (**) and Ahlfors regularity (of the same dimension) imply the existence of local homeomorphic parameterizations which satisfy Sobolev space conditions, as well as their inverses. (One should be a little careful in the formulation of Sobolev space conditions for a map into a metric space. I prefer to use maximal functions, as in [Se2].) In the case of quasisymmetric parameterizations there are pretty strong results of this type, with the Sobolev exponent p larger than the dimension, because of a method of Gehring [Ge] (see also [DS1] and [Se3]). The general case would be more complicated (if true at all), because one would have to modify the parameterization. Note that the 2-dimensional case is again special, because of Theorem 1.6. In all dimensions one can get "uniform rectifiability" (in the sense of [DS4]) under even more general conditions, by [DS5]. Uniform rectifiability implies the existence of some non-homeomorphic parameterizations with otherwise very good estimates, *i.e.*, there are bilipschitz parameterizations of large pieces of the set, and the set can be put inside a bigger one which admits a controlled parameterization that allows a limited amount of crossing. (See also [DS2].)

Geometric topology is particularly successful at parameterizing a set if one is permitted to "stabilize" first by taking a Cartesian product with the real line or some $\mathbb{R}^k$. Perhaps there is a general theorem to the effect that (*), possibly in combination with Ahlfors regularity, implies (**) after stabilizing. This should be compared with the result of Ferry [F2] for (†) mentioned in the introduction. A positive conjecture is supported by the examples discussed in Sections 4 and 5. In each of those situations the non-manifold M becomes homeomorphic to $\mathbb{R}^4$ after taking a Cartesian product with the real line. (See [K, p. 87, Theorem 1] for the set discussed in Section 4, [B4] for the set discussed in Section 5, and [D, Section 11] for both cases and others.) This fact together with self-similarity allows us to obtain (**) for the product by the same arguments as used in Section 6 to prove Theorem 6.3. Other examples can be produced as in of [D, Section 11].

One can also ask whether stabilization makes it easier to get quasisymmetric parameterizations. That is, if M is the non-manifold constructed as in Sections 4 or 5, then does $M \times \mathbb{R}$ admit a quasisymmetric parameterization by $\mathbb{R}^4$? I am pessimistic, for reasons like those in Remark 6.19.

See also [Se5] for some related topics.

Acknowledgements. I would like to thank Juha Heinonen for many valuable consultations. We have been discussing these kinds of questions for some time, and it was he who told me about [FS]. I am also grateful to a splintered cabal of topologists for revealing to me a fraction of their ancient knowledge.

References.

[Ar] Armentrout, S., Cellular decompositions of 3-manifolds that yield 3-manifolds. *Memoirs Amer. Math. Soc.* **107** (1971).

[A1] Assouad, P., Espaces Métriques, Plongements, Facteurs, Thèse de Doctorat, Université de Paris XI, 1977.

[A2] Assouad, P., Étude d'une dimension métrique liée à la possibilité de plongement dans $\mathbb{R}^n$. *C. R. Acad. Sci. Paris* **288** (1979), 731-734.

[A3] Assouad, P., Plongements Lipschitziens dans $\mathbb{R}^n$. *Bull. Soc. Math. France* **111** (1983), 429-448.

[AV] Alestalo, P., Väisälä, J., Quasisymmetric embeddings of products of cells into the Euclidean space. *Ann. Acad. Sci. Fenn. Ser. A I Math.* **19** (1994), 375-392.

[B1] Bing, R. H., A homeomorphism between the 3-sphere and the sum of two solid horned spheres. *Ann. of Math.* **56** (1952), 354-362.

[B2] Bing, R. H., Upper semicontinuous decompositions of E^3. *Ann. of Math.* **65** (1957), 363-374.

[B3] Bing, R. H., A decomposition of E^3 into points and tame arcs such that the decomposition space is topologically different from E^3. *Ann. of*

Math. **65** (1957), 484-500.

[B4] Bing, R. H., The Cartesian product of a certain nonmanifold and a line is E^4. *Ann. of Math.* **70** (1959), 399-412.

[B5] Bing, R. H., Shrinking without lengthening. *Topology* **27** (1988), 487-493.

[Bi] Bižaca, Ž., An explicit family of exotic Casson handles. *Proc. Amer. Math. Soc.* **123** (1995), 1297-1302.

[BDS] Brackx, F., Delanghe, R., Sommen, F., *Clifford Analysis*, Pitman, 1982.

[B^∞] Bryant, J., Ferry, S., Mio, W., Weinberger, S., Topology of homology manifolds. *Bull. Amer. Math. Soc.* **28** (1993), 324-328.

[C1] Cannon, J., The recognition problem: What is a topological manifold?. *Bull. Amer. Math. Soc.* **84** (1978), 832-866.

[C2] Cannon, J., The characterization of topological manifolds of dimension $n \geq 5$. *Proceedings I.C.M., Helsinki* (1978), 449-454.

[C3] Cannon, J., Shrinking cell-like decompositions of manifolds: codimension 3. *Ann. of Math.* **110** (1979), 83-112.

[CW1] Coifman, R. Weiss, G., *Analyse Harmonique Non-commutative sur Certains Espaces Homogènes.* Lecture Notes in Math. **242**, Springer-Verlag, 1971.

[CW2] Coifman, R., Weiss, G., Extensions of Hardy spaces and their use in analysis. *Bull. Amer. Math. Soc.* **83** (1977), 569-645.

[Cu] Curtis, M., An imbedding theorem. *Duke Math. J.* **24** (1957), 349-352.

[D] Daverman, R., *Decompositions of Manifolds.* Academic Press, 1986.

[Da] David, G., *Wavelets and Singular Integrals on Curves and Surfaces.* Lecture Notes in Math. **1465**, Springer-Verlag, 1991.

[DS1] David, G., Semmes, S., Strong A_∞-weights, Sobolev inequalities, and quasiconformal mappings. In *Analysis and Partial Differential Equations*, edited by C. Sadosky, Lecture Notes in Pure and Appl. Math. **122**, Marcel Dekker, 1990.

[DS2] David, G., Semmes, S., *Singular integrals and rectifiable sets in $\mathbb{R}^n$: au-delà des graphes lipschitziens.* Astérisque **193**, Soc. Math. France, 1991.

[DS3] David, G., Semmes, S., Quantitative rectifiability and Lipschitz mappings. *Trans. Amer. Math. Soc.* **337** (1993), 855-889.

[DS4] David, G., Semmes, S., *Analysis of and on Uniformly Rectifiable Sets.* Math. Surveys and Monographs **38**, Amer. Math. Soc., 1993.

[DS5] David, G., Semmes, S., to exist.

[DoS] Donaldson, S., Sullivan, D., Quasiconformal 4-manifolds. *Acta Math.* **163** (1989), 181-252.

[E] Edwards, E., The topology of manifolds and cell-like maps. *Proceedings I.C.M., Helsinki* (1978), 111-127.

[Fe] Federer, H., *Geometric Measure Theory*, Springer-Verlag, 1969.

[F1] Ferry, S., Topological finiteness theorems for manifolds in Gromov-Hausdorff space. *Duke Math. J.* **74** (1994), 95-106.

[F2] Ferry, S., Counting simple-homotopy types of topological manifolds. Preprint.

[Fr] Freedman, M. H., The topology of four-dimensional manifolds. *F. Diff. Geometry* **17** (1982), 357-453.

[FS] Freedman, M., Skora, R., Strange actions of groups on spheres. *J. Diff. Geom.* **25** (1987), 75-98.

[Ge] Gehring, F. W., The L^p integrability of the partial derivatives of a quasiconformal mapping. *Acta Math.* **130** (1973), 265-277.

[GP] Greene, R., Petersen V, P., Little topology, big volume. *Duke. Math. J.* **67** (1992), 273-290.

[GPW] Grove, K., Petersen V, P., Wu, J.-Y., Geometric finiteness theorems via controlled topology. *Invent. Math.* **99** (1990), 205-213; erratum, **104** (1991), 221-222.

[HK] Heinonen, J., Koskela, P., Definitions of quasiconformality. *Invent. Math.* **120** (1995), 61-79.

[HY] Heinonen, J., Yang, S., Strongly uniform domains and periodic quasiconformal maps. *Ann. Acad. Sci. Fenn. Ser. A I Math.* **20** (1995), 123-148.

[K] Kirby, R., *The Topology of 4-manifolds.* Lecture Notes in Math. **1374**, Springer-Verlag, 1989.

[P] Petersen V, P., Gromov-Hausdorff convergence of metric spaces. *Proc. Symp. Pure Math.* **54** (1993), 489-504.

[Q1] Quinn, F., Resolutions of manifolds, and the topological characterization of manifolds. *Invent. Math.* **72** (1983), 267-284.

[Q2] Quinn, F., An obstruction to the resolution of homology manifolds. *Michigan Math. J.* **301** (1987), 285-292.

[Se1] Semmes, S., Differentiable function theory on hypersurfaces in $\mathbb{R}^n$ (without bounds on their smoothness). *Indiana Math. J.* **39** (1990), 985-1004.

[Se2] Semmes, S., Chord-arc surfaces with small constant II: Good parameterizations. *Advances in Math.* **88** (1991), 170 -199.

[Se3] Semmes, S., Bilipschitz mappings and strong A_∞ weights. *Ann. Acad. Sci. Fenn. Serie A I Math.* **18** (1993), 211-248.

[Se4] Semmes, S., Finding structure in sets with little smoothness. To appear in *Proc. I.C.M., Zürich* (1994), Birkhäuser.

[Se5] Semmes, S., On the nonexistence of bilipschitz parameterizations and geometric problems about A_∞ weights. To appear in *Revista Mat. Iberoamericana.*

[SS] Siebenmann, L., Sullivan, D., On complexes that are Lipschitz manifolds. In *Geometric Topology*, J. Cantrell, ed., Academic Press, 1979, 503-525.

[St] Stein, E. M., *Singular Integrals and Differentiability Properties of Functions.* Princeton Univ. Press, 1970.

[To1] Toro, T., Surfaces with generalized second fundamental form in L^2 are Lipschitz manifolds. *J. Diff. Geom.* **39** (1994), 65-102.

[To2] Toro, T., Geometric conditions and existence of bilipschitz parameterizations. *Duke Math. J.* **77** (1995), 193-227.

[Tu1] Tukia, P., The planar Schönflies theorem for Lipschitz maps. *Ann. Acad. Sci. Fenn. Ser. A I Math.* **5** (1980), 49-72.

[Tu2] Tukia, P., A quasiconformal group not isomorphic to a Möbius group. *Ann. Acad. Sci. Fenn. Ser. A I Math.* **6** (1981), 149-160.

[TV] Tukia, P., Väisälä, J., Quasisymmetric embeddings of Euclidean spaces. *Ann. Acad. Sci. Fenn. Ser. A I Math.* **5** (1980), 97-114.

[V1] Väisälä, J., Quasisymmetric maps of products of curves into the plane. *Rev. Roumaine Math. Pures Appl.* **33** (1988), 147-156.

[V2] Väisälä, J., Metric duality in Euclidean spaces. Preprint.

Recibido: 15 de febrero de 1.995

Stephen Semmes*
Department of Mathematics
Rice University
Houston, Texas 77251, U.S.A.
semmes@math.rice.edu

and

I.H.E.S.
35 Route de Chartres
91440 Bures-sur-Yvette, FRANCE

* supported by the U.S. National Science Foundation.

REVISTA MATEMÁTICA IBEROAMERICANA
VOL. 12, N.° 2, 1996

Etude de la vitesse de convergence de l'algorithme en cascade dans la construction des ondelettes d'Ingrid Daubechies

Sylvain Durand

Résumé. L'objet de cet article est l'étude de la vitesse de convergence des algorithmes de résolution d'une équation à deux échelles. Il s'agit d'algorithmes de point fixe, souvent appelés "algorithmes en cascade", qui sont utilisés dans la construction des ondelettes. Nous étudions la vitesse de convergence dans les espaces de Lebesgue et de Besov, et montrons que la qualité de convergence dépend de deux facteurs indépendants. Le premier, qui va de soi, est la régularité de la fonction d'échelle qui est la solution de l'équation. Le second facteur (qui est la découverte essentielle de ce travail) concerne des propriétés algébriques spécifiques de la fonction servant à initialiser l'algorithme. celle-ci doit satisfaire des conditions analogues à celles de Strang-Fix.

Abstract. The aim of this paper is the study of the convergence of algorithms involved in the resolution of two scale equations. They are fixed point algorithms, often called "cascade algorithm", which are used in the construction of wavelets. we study their speed of convergence in

Lebesgue and Besov spaces, and show that the quality of the convergence depends on two independent factors. The first one, as we could foresee, is the regularity of the scaling function which is the solution of the equation. The second factor (that is the essential discovery of this work) concerns specific algebraic properties of the function used to initialize the algorithm. This function must satisfy conditions analogous to Strang-Fix conditions.

0. Introduction.

Dans ce travail, nous étudions la vitesse de convergence des algorithmes de résolution d'une équation d'échelle,

$$\varphi(x) = \sum_{l \in \mathbb{Z}^n} c_l \, \varphi(2x - l)\,, \tag{1}$$

où la fonction φ que l'on cherche à construire appartient à l'espace $L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$ et satisfait $\int \varphi(x)\, dx = 1$. Pour construire la fonction d'échelle φ, on utilise un algorithme itératif appelé *algorithme en cascade*. On définit l'opérateur T de $L^2(\mathbb{R}^n)$ dans lui même par

$$T(f)(x) = \sum_{l \in \mathbb{Z}^n} c_l \, f(2x - l)\,. \tag{2}$$

Puis, partant d'une fonction arbitraire f_0 dans $L^2(\mathbb{R}^n)$, on définit la suite $\{f_j\}_{j \in \mathbb{N}}$ par

$$f_j = T(f_{j-1}) = T^j(f_0)\,, \qquad \text{pour tout } j \in \mathbb{N}\,. \tag{3}$$

Nous allons montrer comment la qualité de convergence de cet algorithme de point fixe, est reliée aux propriétés de la fonction limite φ et du terme initial f_0.

Notre première motivation était la construction des ondelettes. C'est dans ce même but que l'algorithme a été utilisé par I. Daubechies dans [4]. Celui-ci est cependant également utilisé dans d'autres finalités. Nous avons donc essayé d'énoncer nos théorèmes dans le cadre le plus général possible.

On définit plus souvent l'équation d'échelle par

$$\varphi(x) = \sum_{l \in \mathbb{Z}^n} c_l \, \varphi(kx - l)\,,$$

pour un entier $k \geq 2$. Les théorèmes énoncés dans cet article se généralisent sans difficultés à cette équation, bien qu'ils soient toujours énoncés lorsque l'entier k est égal à 2, comme cela est le cas pour les ondelettes.

Les équations d'échelle ont été introduites, pour la première fois, en 1956, par G. de Rham [6], dans le but de construire des fonctions continues, nulle part dérivables. Puis ces équations ont été récemment réintroduites dans le cadre des algorithmes d'interpolation. Nous pouvons citer, dans ce domaine, les travaux de W. Dahmen, A. Cavarretta et C. Micchelli [3], [2], [1], G. Deslauriers et S. Dubuc [9], [7], [8], N. Dyn, A. Gregory et D. Levin [11]. Ces algorithmes ont des applications dans le Computer Aided Design (C.A.D.) car ils permettent de lisser les angles des surfaces polyhédriques. Contrairement à ce qu'il en est pour les travaux de G. de Rham, il s'agit souvent de construire des courbes les plus régulières possibles. Cette divergence dans les objectifs va de pair avec le caractère instable de la régularité de la fonction φ.

Le problème de la convergence de l'algorithme en cascade est par contre beaucoup plus stable, lorsque le problème de l'existence et de l'unicité de la solution de l'équation d'échelle a été résolu. Signalons à ce propos que, si l'on omet la condition $\varphi \in L^1(\mathbb{R}^n)$, alors l'espace des solutions de l'équation d'échelle est de dimension infinie. Si l'on suppose $\varphi \in L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$, il est, au plus, de dimension 1 et, lorsque cela est le cas, φ est définie de façon unique par la seconde condition $\int \varphi(x)\, dx = 1$, [10].

Le problème concernant la convergence $L^2(\mathbb{R}^n)$ est donc mal posé dans $L^2(\mathbb{R}^n)$. C'est pourquoi, dans tous les énoncés concernant la convergence $L^2(\mathbb{R}^n)$, on doit imposer à f_0 une autre propriété (qui sera, en général, plus forte que $f_0 \in L^1(\mathbb{R}^n)$). Sinon f_0 pourrait être un autre point-fixe de T ce qui conduirait à $T^j(f_0) = f_0$ qui ne converge pas vers φ.

Nous montrons, dans cet article, que la qualité de la convergence dépend de deux facteurs indépendants et montrons quel est le rôle joué par chacun des deux facteurs. Le premier, qui va de soi, est la régularité de la fonction d'échelle φ. Le second facteur concerne des propriétés algébriques spécifiques de la fonction f_0. Pour avoir une bonne convergence, celle-ci doit satisfaire des propriétés analogues à celles de Strang-Fix,

$$\partial^\alpha \hat{f}_0(2l\pi) = 0\,, \qquad \text{pour tout } \ l \in (\mathbb{Z}^n)^*\,,$$

et

$$\partial^\alpha \hat{f}_0(0) = \partial^\alpha \hat{\varphi}(0)\,,$$

où les $\alpha \in \mathbb{N}^n$ sont donnés par la régularité de φ.

Nous nous ramenons à une suite d'opérateurs définis par changement d'échelle à l'aide d'un noyau intégrale. On montre que l'on a

$$f_j = E_j(\varphi),$$

où l'opérateur E_j est défini par

$$E_j(g)(x) = 2^{nj} \int K(2^j x, 2^j y)\, g(y)\, dy$$

et le noyau K dépend du choix de f_0 et de φ. La convergence de f_j vers φ est donc une conséquence immédiate de la convergence de E_j vers l'opérateur d'identité I.

Nous énonçons les théorèmes dans le cadre de la convergence dans les espaces de Lebesgue $L^p(\mathbb{R}^n)$ pour $1 \leq p < +\infty$, puis dans les espaces de Besov $B_p^{s,q}(\mathbb{R}^n)$ qui sont les plus généraux. L'objet de cet article est de présenter les principaux théorèmes démontrés dans [10]. A l'exception du théorème sur la convergence dans $L^2(\mathbb{R}^n)$, nous nous contenterons ici de donner les résultats.

1. Nouvel énoncé du problème.

Cet énoncé s'appuie sur l'existence d'une fonction $\tilde{\varphi} \in L^2(\mathbb{R}^n)$ telle que, pour φ donnée,

$$\langle \varphi(x-l), \tilde{\varphi}(x) \rangle = \delta_{0l}, \qquad \text{pour tout } \; l \in \mathbb{Z}^n, \tag{4}$$

où $\delta_{0l} = 1$ si $l = 0$, 0 sinon, et $\langle \cdot, \cdot \rangle$ désigne le produit scalaire usuel sur $L^2(\mathbb{R}^n)$.

Dans des cas très généraux, il est toujours possible de construire une telle fonction.

Lemme 1. *On suppose que $\varphi \in L^2(\mathbb{R}^n)$ vérifie les hypothèses suivantes:*

1) *sa transformée de Fourier $\hat{\varphi}$ appartient à $C^\infty(\mathbb{R}^n)$.*

2) *il existe un entier $N > 0$ et un réel $\delta > 0$ tels que*

$$\sum_{|l| \leq N} |\hat{\varphi}(\xi + 2l\pi)|^2 \geq \delta,$$

pour tout $\xi \in [0, 2\pi]^n$.

Alors il existe une fonction $\tilde{\varphi}$ *dans la classe de Schwartz* $\mathcal{S}(\mathbb{R}^n)$ *telle que*

$$\int \bar{\varphi}(x-k)\,\tilde{\varphi}(x-l)\,dx = \delta_{kl}\,, \qquad \textit{pour tous } \; k, l \in \mathbb{Z}^n\,.$$

Pour démontrer le lemme, on doit assurer

$$\sum_{l \in \mathbb{Z}^n} \bar{\hat{\varphi}}(\xi + 2l\pi)\,\hat{\tilde{\varphi}}(\xi + 2l\pi) = 1\,.$$

On cherche $\hat{\tilde{\varphi}}$ sous la forme $\hat{\varphi}\,\theta$ où $\theta \in C_0^\infty(\mathbb{R}^n)$. Or on sait, par hypothèse, que

$$\sum_{l \in \mathbb{Z}^n} |\hat{\varphi}(\xi + 2l\pi)|^2 \geq \delta\,, \qquad \text{si } \; \xi \in [0, 2\pi]^n\,.$$

Il en résulte que

$$\sum_{l \in \mathbb{Z}^n} |\hat{\varphi}(\xi + 2l\pi)|^2\,\theta(\xi + 2l\pi) \geq \delta\,, \qquad \text{sur } \; [0, 2\pi]^n\,,$$

si on prend $\theta \geq 0$ partout et $\theta \geq 1$ sur un compact assez grand. Ensuite, on considère la somme

$$\sigma(\xi) = \sum_{l \in \mathbb{Z}^n} |\hat{\varphi}(\xi + 2l\pi)|^2\,\theta(\xi + 2l\pi)$$

qui est une fonction de $C_0^\infty(\mathbb{R}^n)$, 2π-périodique en chaque variable et qui vérifie $\sigma(\xi) \geq \delta$ sur $\mathbb{R}^n$. On pose finalement

$$\hat{\tilde{\varphi}} = \hat{\varphi}(\xi)\,\frac{\theta(\xi)}{\sigma(\xi)}$$

et on a, par construction

$$\sum_{l \in \mathbb{Z}^n} \bar{\hat{\varphi}}(\xi + 2l\pi)\,\hat{\tilde{\varphi}}(\xi + 2l\pi) = 1\,.$$

Ce résultat nous permet de reformuler le problème de la convergence de l'algorithme en cascade. On peut, en effet, énoncer le "lemme-définition" suivant.

Lemme 2. *Si φ vérifie les conditions* 1) *et* 2), *alors il existe une fonction $\tilde{\varphi}$ dans la classe de Schwartz $\mathcal{S}(\mathbb{R}^n)$ telle que, pour toute fonction f_0 dans $L^2(\mathbb{R}^n)$, on ait*

$$f_j = \sum_{l\in\mathbb{Z}^n} \langle \varphi, \tilde{\varphi}_{jl}\rangle \, f_{jl} \ ,$$

avec $f_{jl}(x) = 2^{nj/2} f_0(2^j x - l)$ et $\tilde{\varphi}_{jl}(x) = 2^{nj/2}\tilde{\varphi}(2^j x - l)$.

La démonstration du Lemme 2 est très simple. On commence par observer qu'il existe une suite α_{jl}, $j \in \mathbb{N}, k \in \mathbb{Z}^n$, telle que, pour toute fonction $f_0 \in L^2(\mathbb{R}^n)$, on ait

$$f_j = T^j(f_0) = \sum_{l\in\mathbb{Z}^n} \alpha_{jl}\, f_{jl} \ .$$

Ceci s'obtient par récurrence à partir de la définition de T. Pour terminer la preuve, il suffit donc de déterminer les coefficients α_{jl}. Pour cela, on utilise le fait que φ soit point fixe de l'opérateur T. On a donc, en prenant $f_0 = \varphi$,

$$\varphi = \sum_{l\in\mathbb{Z}^n} \alpha_{jl}\, \varphi_{jl} \ .$$

Le lemme 1 nous fournit finalement $\alpha_{jl} = \langle \varphi, \tilde{\varphi}_{jl}\rangle$.

On utilise maintenant les notations suivantes,

$$K(x,y) = \sum_{l\in\mathbb{Z}^n} f_0(x-l)\,\bar{\tilde{\varphi}}(y-l)\,, \tag{5}$$

ce qui entraîne

$$f_j = 2^{nj}\int K(2^j x, 2^j y)\,\varphi(y)\,dy\ . \tag{6}$$

Cette expression nous pousse à étudier la suite d'opérateurs E_j définis de la façon suivante.

Définition 1. *Pour tout entier j, on note E_j l'opérateur défini sur $L^2(\mathbb{R}^n)$ et à valeurs dans $L^2(\mathbb{R}^n)$, qui, à une fonction g, associe la fonction*

$$E_j(g)(x) = 2^{nj} \int K(2^j x, 2^j y)\, g(y)\, dy\,. \tag{7}$$

Nous oublions maintenant la définition particulière de K et nous nous posons le problème général de trouver des conditions suffisantes (et si possible nécessaires) sur le noyau K, pour que la suite d'opérateurs E_j converge vers l'identité. La convergence de f_j vers φ en sera une conséquence.

REMARQUE 1. Le fait de pouvoir prendre $\tilde{\varphi}$ dans la classe de Schwartz nous assure une régularité arbitraire pour le noyau K en variable y.

REMARQUE 2. La propriété 2) du Lemme 1, résulte de la propriété, en apparence plus simple suivante: il n'existe pas de $\xi_0 \in \mathbb{R}^n$ tel que $\hat{\varphi}(\xi_0 + 2l\pi) = 0$, pour tout $l \in \mathbb{Z}^n$.

REMARQUE 3. Si $\{\varphi(x-l),\ l \in \mathbb{Z}^n\}$ est une base inconditionnelle de l'espace V_0 qu'elle engendre, on a nécessairement $\sum_{l\in\mathbb{Z}^n} |\hat{\varphi}(\xi + 2l\pi)|^2 \geq \delta > 0$ presque partout. La seule difficulté pour obtenir la propriété 2), est de transformer ce presque partout en partout. C'est le cas si la série écrite est uniformément convergente sur $[0, 2\pi]^n$. Ceci a lieu pour toutes les ondelettes d'I. Daubechies, ainsi que pour toutes les ondelettes qui sont utilisées dans la pratique.

REMARQUE 4. En utilisant un résultat de P. G. Lemarié-Rieusset [12], on peut montrer que, si φ est une fonction d'échelle à support compact minimal dans une analyse multirésolution de dimension 1, il est possible dans $C_0^\infty(\mathbb{R}^n)$. Le fait de prendre $\tilde{\varphi}$ à support compact, peut simplifier les démonstrations.

2. Convergence dans les espaces de Lebesgue $L^p(\mathbb{R}^n)$.

Nous commençons par étudier la convergence dans l'espace de Lebesgue $L^2(\mathbb{R}^n)$.

Lemme 3. *Soient ε, R et C trois réels strictement positifs. On suppose que le noyau K vérifie*

1) $\sup_{t\in\mathbb{R}^n} \iint_{|x-t|\le R} |K(x,y)|^2\,(1+|x-y|)^{n+\varepsilon}\,dx\,dy \le C\,,$

2) $\sup_{t\in\mathbb{R}^n} \iint_{|y-t|\le R} |K(x,y)|^2\,(1+|x-y|)^{n+\varepsilon}\,dx\,dy \le C\,,$

3) $\int K(x,y)\,dy = 1$, *presque pour tout.*

Alors, pour toute fonction g dans $L^2(\mathbb{R}^n)$, $E_j(g)$ converge vers g en norme $L^2(\mathbb{R}^n)$. De plus, si $\int K(x,y)\,dy = 1$ appartient à $L^\infty(\mathbb{R}^n)$ alors la condition 3) est nécessaire.

Nous commençons par montrer que l'opérateur E_0 est borné sur $L^2(\mathbb{R}^n)$. Pour cela, on utilise l'inégalité de Cauchy-Schwarz.

$$\begin{aligned}\|E_0(g)\|_2^2 &= \int \Big|\int K(x,y)\,g(y)\,dy\Big|^2 dx\\ &\le \int \Big(\int |K(x,y)|^2\,(1+|x-y|)^{n+\varepsilon}\,dy\Big)\\ &\qquad \cdot \Big(\int |g(z)|^2\,(1+|x-z|)^{-n-\varepsilon}\,dz\Big)\,dx\,.\end{aligned}$$

On décompose l'intégrale sur des cubes unitaires:

$$\begin{aligned}\|E_0(g)\|_2^2 &\le \sum_{l\in\mathbb{Z}^n}\Big(\int_{[l,l+1]}\Big(\int |K(x,y)|^2\,(1+|x-y|)^{n+\varepsilon}\,dy\Big)\\ &\qquad\qquad \cdot\Big(\int |g(z)|^2\,(1+|x-z|)^{-n-\varepsilon}\,dz\Big)\,dx\Big)\\ &\le \sum_{l\in\mathbb{Z}^n}\Big(\int_{[l,l+1]}\Big(\int |K(x,y)|^2\,(1+|x-y|)^{n+\varepsilon}\,dy\Big)\,dx\Big)\\ &\qquad\qquad \cdot \sup_{t\in[l,l+1]}\Big(\int |g(z)|^2\,(1+|t-z|)^{-n-\varepsilon}\,dz\Big)\end{aligned}$$

avec $[l,l+1] = [l_1,l_1+1]\times\cdots\times[l_n,l_n+1]$. En utilisant la première hypothèse du lemme, on constate que, à une constante multiplicative près, ceci est inférieur à

$$\sum_{l\in\mathbb{Z}^n}\ \sup_{t\in[l,l+1]}\int |g(y)|^2(1+|t-y|)^{-n-\varepsilon}\,dy\,,$$

ou encore

$$\int |g(y)|^2 \Big(\sum_{l \in \mathbb{Z}^n} \sup_{t \in [l, l+1]} (1 + |t-y|)^{-n-\varepsilon} \Big) \, dy \, .$$

Finalement, on a montré que

$$\int \Big| \int K(x,y) \, g(y) \, dy \Big|^2 dx \leq C \int |g(y)|^2 \, dy \, ,$$

c'est à dire $\|E_0(g)\|_2 \leq C \, \|g\|^2$. Et comme $L^2(\mathbb{R}^n)$ est un espace homogène, on a aussi

$$\|E_j(g)\|_2 \leq \|g\|_2 \, .$$

Pour terminer la démonstration, on peut alors se concentrer sur le cas des fonctions C^∞ à support compact. Le reste se fait par densité.

On doit montrer que $\|E_j(g) - g\|_2$ tend vers zéro lorsque j tend vers l'infini, pour toute fonction $g \in \mathcal{C}_0^\infty(\mathbb{R}^n)$. Comme $\int K(x,y) \, dy = 1$, on a

$$E_j(g)(x) - g(x) = 2^{nj} \int K(2^j x, 2^j y) \, (g(y) - g(x)) \, dy \, .$$

Et comme $g \in \mathcal{C}_0^\infty(\mathbb{R}^n)$, on a , pour tout réel $0 < s \leq 1$,

$$|g(y) - g(x)| \leq C \, |y - x|^s \, ,$$

d'où

$$|E_j(g)(x) - g(x)| \leq C \, 2^{nj} \int |K(2^j x, 2^j y)| \, |y - x|^s \, dy \, .$$

On note maintenant A un réel positif tel que le support de g soit inclus dans $[-A, A]^n$. Et on traite séparément les cas où $x \in [-A, A]^n$ et $\mathbb{R}^n \setminus [-A, A]^n$. Lorsque $x \in \mathbb{R}^n \setminus [-A, A]^n$, on a $g(x) = 0$ et, par conséquent, $g(y) - g(x)$ s'annule pour $y \in \mathbb{R}^n \setminus [-A, A]^n$. On a donc

$$\begin{aligned} \|E_j(g) - g\|_2^2 \leq C & \int_{[-A,A]^n} 2^{nj} \Big(\int |K(2^j x, 2^j y)| \, |y - x|^s \, dy \Big)^2 dx \\ & + C \int_{\mathbb{R}^n \setminus [-A,A]^n} 2^{nj} \Big(\int_{[-A,A]^n} |K(2^j x, 2^j y)| \, |y - x|^s \, dy \Big)^2 dx \\ & := \mathrm{I}_A + \mathrm{J}_A \, . \end{aligned}$$

Pour la première intégrale, on utilise, une nouvelle fois, l'inégalité de Cauchy-Schwarz.

$$\begin{aligned}
\mathrm{I}_A &= \int\limits_{[-A,A]^n} 2^{2nj}\Big(\int |K(2^j x, 2^j y)|\, |y-x|^s \, dy\Big)^2 dx \\
&\le 2^{2nj} \int\limits_{[-A,A]^n} \Big(\int |K(2^j x, 2^j y)|^2 \,(1+2^j\,|x-y|)^{n+\varepsilon}\, dy\Big) \\
&\qquad\qquad \cdot \Big(\int |y-x|^{2s}\,(1+2^j\,|x-y|)^{-n-\varepsilon}\, dy\Big)\, dx\,.
\end{aligned}$$

On impose $s < \varepsilon/2$ et on majore le deuxième facteur de l'intégrand, de la façon suivante

$$2^{-nj}\int 2^{-2sj}\,|y|^{2s}\,(1+|y|)^{-n-\varepsilon}\, dy \le C\, 2^{-nj}\, 2^{-2sj}\,.$$

On a donc

$$\begin{aligned}
\mathrm{I}_A &\le C\, 2^{-2sj}\, 2^{nj} \iint\limits_{x\in[-A,A]^n} |K(2^j x, 2^j y)|^2\,(1+2^j\,|x-y|)^{n+\varepsilon}\, dy\, dx \\
&\le C\, 2^{-2sj}\, 2^{-nj} \iint\limits_{x\in[-2^j A, 2^j A]^n} |K(x,y)|^2\,(1+|x-y|)^{n+\varepsilon}\, dy\, dx\,.
\end{aligned}$$

Et d'après la première hypothèse du lemme

$$\mathrm{I}_A \le C\, 2^{-2sj}\,.$$

Passons au reste de l'intégrale. On doit majorer

$$\mathrm{J}_A = \int\limits_{x\notin[-A,A]^n} 2^{nj}\Big(\int\limits_{y\in[-A,A]^n} |K(2^j x, 2^j y)|\, |y-x|^s\, dy\Big)^2 dx\,.$$

En reprenant les mêmes calculs que pour I_A, on arrive à

$$\mathrm{J}_A \le C\, 2^{-2sj}\, 2^{-nj} \iint\limits_{y\in[-2^j A, 2^j A]^n} |K(x,y)|^2\,(1+|x-y|)^{n+\varepsilon}\, dy\, dx\,.$$

Et, d'après la seconde hypothèse du lemme, on a

$$J_A \leq C\, 2^{-2sj},$$

et

$$\|E_j(g) - g\|_2 \leq C\, 2^{-sj}.$$

Si maintenant on se place dans le cas particulier où le noyau K est défini par (5), on obtient, comme corollaire du Lemme 3, un premier théorème.

Théorème 1. *On suppose que $\varphi \in L^2(\mathbb{R}^n)$ vérifie les conditions 1) et 2) du Lemme 1. Si $f_0 \in L^2(\mathbb{R}^n)$ est à support compact et vérifie $\sum_{l\in\mathbb{Z}^n} f_0(x-l) = 1$ presque partout, alors f_j converge vers φ au sens de la norme $L^2(\mathbb{R}^n)$. De plus, si $|f_0(x)| \leq (1+|x|)^{-n-\varepsilon}$, $(\varepsilon > 0)$, alors la condition $\sum_{l\in\mathbb{Z}^n} f_0(x-l) = 1$ est nécessaire.*

Si l'identité $\sum_{l\in\mathbb{Z}^n} f_0(x-l) = 1$ n'est pas vérifiée, il nous reste cependant la convergence faible de l'algorithme.

Proposition 1. *On suppose que $\varphi \in L^2(\mathbb{R}^n)$ vérifie les conditions 1) et 2) du Lemme 1. Si $f_0 \in L^2(\mathbb{R}^n)$ est à support compact et vérifie $\int f_0(x)\,dx = 1$, alors f_j converge vers φ faiblement dans $L^2(\mathbb{R}^n)$.*

Le Lemme 3 se généralise, sans difficultés, au cadre de la convergence $L^p(\mathbb{R}^n)$ pour $1 \leq p < +\infty$. On peut donc énoncer un deuxième théorème.

Théorème 2. *On suppose que $\varphi \in L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$ vérifie les hypothèses 1) et 2) du Lemme 1. Si $f_0 \in L^p(\mathbb{R}^n)$ est à support compact et $\sum_{l\in\mathbb{Z}^n} f_0(x-l) = 1$, presque pour tout, alors f_j converge vers φ au sens de la norme $L^p(\mathbb{R}^n)$.*

La démonstration du Lemme 3 s'appuyant sur la densité de l'espace $C_0^\infty(\mathbb{R}^n)$ dans $L^p(\mathbb{R}^n)$, pour $1 \leq p < +\infty$, nous ne savons pas généraliser ce résultat au cas $p = +\infty$. Il nous reste, en fait, uniquement la convergence faible. Pour obtenir la convergence forte, on doit faire une hypothèse supplémentaire qui porte sur la régularité de φ.

3. Convergence dans les espaces de Besov.

Nous commençons par étudier la convergence uniforme de l'algorithme. Le théorème suivant, qui se démontre toujours en ramenant à la convergence des opérateurs E_j, nous donne la vitesse de convergence de l'algorithme en fonction de la régularité de φ.

Théorème 3. *Soit $s \in \mathbb{R}_+^*$. On suppose que $\varphi \in C^s(\mathbb{R}^n)$ vérifie les conditions* 1) *et* 2) *du Lemme* 1. *Si $f_0 \in L^\infty(\mathbb{R}^n)$ vérifie*

$$\sum_{l\in\mathbb{Z}^n} |f_0(x-l)|\,|x-l|^s \leq C < +\infty$$

et

$$\partial^\beta \hat{f}_0(2l\pi) = \partial^\beta \hat{\varphi}(2l\pi)\,,$$

pour $\beta \in \mathbb{N}^n$ telle que $|\beta| = \beta_1 + \cdots + \beta_n < s$ et $l \in \mathbb{Z}^n$, alors on a

$$\|f_j - \varphi\|_\infty \leq C\,2^{-sj}\,.$$

Pour la démonstration, on utilise un développement de Taylor-Young de φ à l'ordre s, dans (6).

$$E_j(\varphi)(x) - \varphi(x) = \sum_{1\leq|\beta|<s} 2^{-|\beta|j}\Big(\int K(2^jx,y)\,\frac{(y-2^jx)^\beta}{\beta!}\,dy\Big)\partial^\beta\varphi(x) + O\Big(2^{-sj}\int |K(2^jx,y)|\,|y-2^jx|^s\,dy\Big)\,.$$

Cette expression est intéressante car elle nous montre comment la vitesse de convergence de l'algorithme est ralentie par les termes

$$\mu_\beta(2^jx) = \int K(2^jx,y)\,\frac{(y-2^jx)^\beta}{\beta!}\,dy\,.$$

Pour les annuler, on doit satisfaire

$$\partial^\beta \hat{f}_0(2l\pi) = \partial^\beta \hat{\varphi}(2l\pi)\,, \qquad \text{pour tout } l \in \mathbb{Z}^n\,.$$

Comme, on a $\partial^\beta\hat{\varphi}(2l\pi) = 0$, pour tous $l \in (\mathbb{Z}^n)^*$ et $\beta \in \mathbb{N}^n$ telle que $|\beta| \leq s$, on doit prendre f_0 telle que

(8) $\partial^\beta \hat{f}_0(2l\pi) = 0\,,$ pour tous $l \in (\mathbb{Z}^n)^*$ et $\beta \in \mathbb{N}^n$ telle que $|\beta| \leq s$,

et

$$(9) \quad \int x^\beta f_0(x)\,dx = \int x^\beta \varphi(x)\,dx\,, \quad \text{pour tout } \beta \in \mathbb{N}^n \text{ telle que } |\beta| \leq s\,.$$

On a donc uniquement besoin de connaître les moments de φ.

Si cette dernière condition n'est plus vérifiée, mais si (8) l'est encore, les fonctions μ_β sont encore constantes. Cette condition est suffisante pour que l'algorithme converge dans l'espace de Hölder (si f_0 est régulière) comme le montre le théorème suivant. Soit $s_0 \in \mathbb{R}_+^*$, on note $[s_0]$ le plus petit entier strictement inférieur à s_0.

Théorème 4. *On suppose que φ appartient à $C^{s_0}(\mathbb{R}^n)$ et vérifie les conditions* 1) *et* 2) *du Lemme* 1. *Si $f_0 \in C^{s_0}(\mathbb{R}^n)$ vérifie*

1) $|\partial^\beta f_0(x)| \leq C\,(1+|x|)^{-n-|\beta|}$, *pour tout $\beta \in \mathbb{N}^n$ telle que $|\beta| < s_0$,*

2) $|f_0(x)| \leq C\,(1+|x|)^{-n-\varepsilon}$, *pour un réel $\varepsilon > 0$,*

3) $|\partial^\beta f_0(x) - \partial^\beta f_0(x')| \leq C\,|x-x'|^{-s_0+[s_0]} \sup_{x,x'}(1+|t|)^n$, *pour tout $\beta \in \mathbb{N}^n$ telle que $|\beta| = [s_0]$,*

4) $\sum_{l\in\mathbb{Z}^n}(x-l)^\beta f_0(x-l)$ *est constante, pour tout $\beta \in \mathbb{N}^n$ telle que $|\beta| \leq [s_0]$,*

5) $\sum_{l\in\mathbb{Z}^n} f_0(x-l) \equiv 1$,

alors f_j converge vers φ dans $C^s(\mathbb{R}^n)$ pour tout $s < s_0$.

La démonstration de ce théorème fait appel à la théorie des opérateurs d'intégrales singulières. Il en est de même pour le théorème suivant qui en est une généralisation aux espaces de Besov $B_p^{s,q}(\mathbb{R}^n)$. Cette famille d'espaces comprend les espaces de Hölder $C^s(\mathbb{R}^n) = B_\infty^{s,\infty}(\mathbb{R}^n)$ et les espaces de Sobolev $H^s(\mathbb{R}^n) = B_2^{s,2}(\mathbb{R}^n)$.

Théorème 5. *Soit $s_0 \in \mathbb{R}_+^*$. On suppose que φ appartient à $B_p^{s,q}(\mathbb{R}^n)$, pour tout $s < s_0$ $(1 \leq p < +\infty,\ 1 \leq q \leq +\infty)$ et satisfait les conditions* 1 *et* 2 *du Lemme* 1. *Si $f_0 \in C^{s_0}(\mathbb{R}^n)$ est à support compact et*

$$\sum_{l\in\mathbb{Z}^n}(x-l)^\beta f_0(x-l) \quad \textit{est constante pour } |\beta| < s_0$$

$(\sum_{l\in\mathbb{Z}^n} f_0(x-l) = 1)$, *alors f_j converge vers φ dans $B_p^{s,q}(\mathbb{R}^n)$, pour tout $s < s_0$.*

REMARQUE 5. Cet énoncé n'est pas entièrement satisfaisant car on impose à f_0 une régularité hölderienne d'ordre s_0, alors que l'on recherche la convergence de l'algorithme dans les espaces $B_p^{s,q}(\mathbb{R}^n)$ qui correspondent à une régularité d'ordre $s < s_0$ en un sens un peu moins précis. Il n'est pas naturel, pour étudier la convergence d'une suite dans un espace donné, de prendre son terme initial dans un autre espace. En l'ocurrence, nous n'avons que $B_p^{s,q}(\mathbb{R}^n) \subset C^{s-n/p}(\mathbb{R}^n)$.

En utilisant la Remarque 4, nous avons obtenu un énoncé plus précis qui est cependant uniquement valable en dimension 1.

Théorème 6. *Soit $s_0 \in \mathbb{R}_+^*$, $1 \leq p < +\infty$ et $1 \leq q \leq +\infty$. On suppose que φ appartient à $B_p^{s,q}(\mathbb{R}^n)$ pour tout $s < s_0$ et φ est une fonction d'échelle à support compact minimal dans une analyse multirésolution. Si $f_0 \in B_p^{s,q}(\mathbb{R}^n)$ pour tout $s < s_0$ et vérifie*

$$\int |f_0(x)|^p (1 + |x|)^p dx \leq C < +\infty$$

et

$$\sum_{l \in \mathbb{Z}^n} (x - l)^\beta f_0(x - l)$$

est constante pour tout $\beta \in \mathbb{N}^n$ telle que $|\beta| < s_0$ ($\sum_{l \in \mathbb{Z}^n} f_0(x - l) = 1$), alors l'algorithme converge dans $B_p^{s,q}(\mathbb{R}^n)$, pour tout $s < s_0$.

4. Exemples.

Ce dernier paragraphe est consacré aux exemples d'applications des théorèmes énoncés dans les paragraphes précédents. Nous commentons quelques choix standards pour la fonction f_0 et en donnons des nouveaux.

EXEMPLE 4.1. Nous commençons par montrer comment la mauvaise convergence se manifeste lorsque f_0 ne satisfait pas les bonnes conditions. Supposons, par exemple, que f_0 appartienne à $C_0^\infty(\mathbb{R}^n)$ mais ne vérifie pas la condition $\sum_{l \in \mathbb{Z}^n} f_0(x - l) \equiv 1$. Le corollaire 1 nous assure la convergence faible de l'algorithme. Mais, d'après le Théorème 1, nous savons qu'il n'y a pas convergence forte.

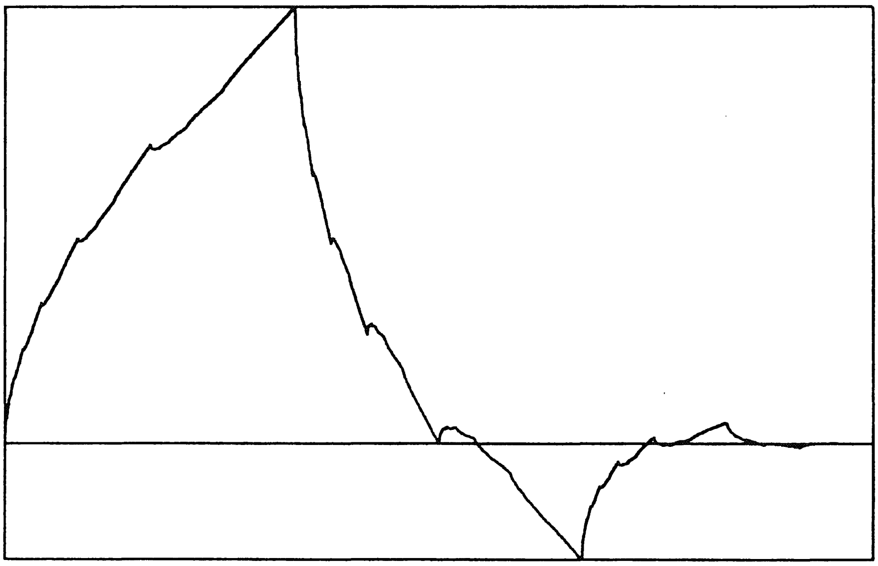


Figure 1. La fonction φ_3 (voir texte)

Les figures 2 et 3 sont une illustration de l'identité

$$\begin{aligned} f_j(x) &= \varphi(x)\Big(\int K(2^j x, y)\,dy\Big) + O(2^{-sj}) \\ &= \varphi(x)\Big(\sum_{l\in\mathbb{Z}} f_0(x-l)\Big) + O(2^{-sj}), \end{aligned}$$

obtenue à partir de 6 et de $|\varphi(x) - \varphi(y)| \leq C\,|x-y|^s$.

Dans cet exemple, φ_3 est l'ondelette d'I. Daubechies, solution de l'équation

$$\begin{aligned} \varphi(x) &= \frac{1+\sqrt{3}}{4}\,\varphi(2x) + \frac{3+\sqrt{3}}{4}\,\varphi(2x-1) \\ &\quad + \frac{3-\sqrt{3}}{4}\,\varphi(2x-2) + \frac{1-\sqrt{3}}{4}\,\varphi(2x-3). \end{aligned} \tag{10}$$

La fonction φ_3 appartient à $C^s(\mathbb{R})$ avec $s = 2 - \log_2(1+\sqrt{3}) \sim 0.55\cdots$, [5].

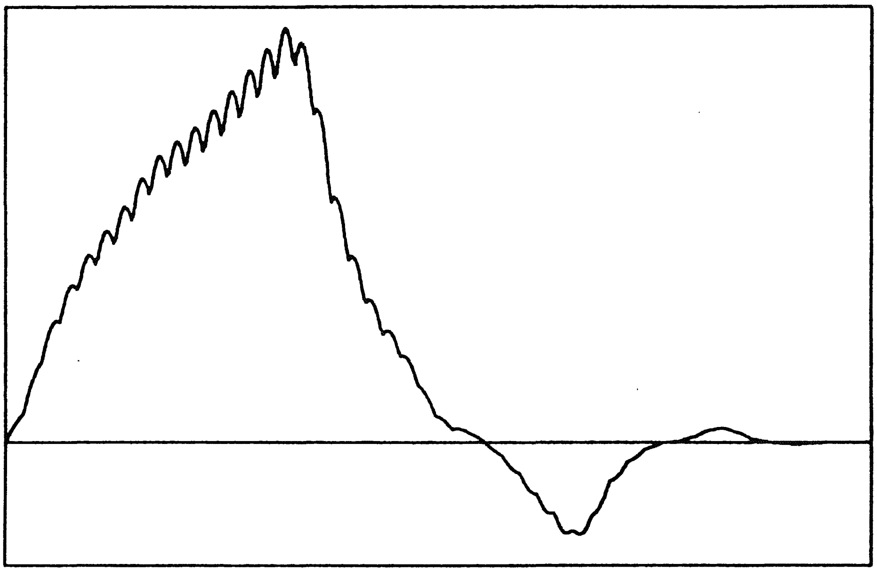


Figure 2. f_4 lorsque $f_0 \in C_0^\infty(\mathbb{R}^n)$ mais ne satisfait pas $\sum_{l \in \mathbb{Z}^n} f_0(x-l) = 1$. Il y a uniquement convergence faible.

Figure 3. f_8 lorsque $f_0 \in C_0^\infty(\mathbb{R}^n)$ mais ne satisfait pas $\sum_{l \in \mathbb{Z}^n} f_0(x-l) = 1$.

EXEMPLE 4.2. En introduisant l'algorithme en cascade, I. Daubechies a utilisé, comme fonction f_0, $\chi_{[-\pi/2,\pi/2]}$ la fonction indicatrice de l'intervalle $[-\pi/2, \pi/2]$. On a alors immédiatement $\sum_{l\in\mathbb{Z}} f_0(x-l) \equiv 1$ ce qui nous assure, d'après le Théorème 2, la convergence de l'algorithme dans les espaces $L^p(\mathbb{R})$.

Si l'on cherche à construire φ_3, la convergence uniforme a lieu avec une vitesse en $O(2^{-sj})$. Cependant, si la régularité de la fonction d'échelle que l'on cherche à construire dépasse 1, la vitesse de convergence ne pourra pas dépasser 2^{-j} car on n'a pas $\sum_{l\in\mathbb{Z}}(x-l)\, f_0(x-l) = \int y\, \varphi(y)\, dy$.

EXEMPLE 4.3. Dans [17], L. F. Villemoes a initialisé l'algorithme avec une fonction de la classe de Schwartz $\mathcal{S}(\mathbb{R})$ normalisée par $\int f_0 = 1$ et dont le support de la transformée de Fourier est dans $[-2\pi, 2\pi]$. On obtient

$$\partial^\alpha f_0(2l\pi) = 0\,, \qquad \text{pour tous } l \in \mathbb{Z}^*\,,\ \alpha \in \mathbb{N}\,,$$

et l'algorithme converge dans les espaces de Besov $B_p^{s,q}(\mathbb{R})$ dès que φ appartient à $B_p^{s+\varepsilon,q}(\mathbb{R})$ pour $\varepsilon > 0$. Cependant, comme cela était le cas dans l'exemple précédent, la vitesse de convergence dans spaces de Lebesgue ne peut, en général, pas dépasser 2^{-j} car on n'a pas l'égalité des moments,

$$\int x^\alpha f_0(x)\, dx = \int x^\alpha \varphi(x)\, dx\,.$$

EXEMPLE 4.4. Un moyen pour obtenir cette identité nous est donné par I. Daubechies et J. Lagarias dans [5]. Celui-ci fait appel à un calcul des valeurs de φ et de ses dérivées sur les entiers.

Si $\varphi \in C^{N+\varepsilon}(\mathbb{R})$ où N est entier et $0 < \varepsilon < 1$, on prend pour f_0, la fonction spline de degré $2N+1$ qui satisfait la propriété,

$$\text{(11)} \qquad \partial^k f_0(l) = \partial^k \varphi(l)\,, \qquad \text{pour tous } k = 0, \ldots, N,\ l \in \mathbb{Z}\,.$$

Les valeurs de $\partial^k\varphi(l)$ sont déterminées par résolution du système,

$$\partial^k \varphi(l) = \sum_{m\in\mathbb{Z}} c_{2l-m}\, 2^k\, \partial^k \varphi(m)\,, \qquad \text{pour tout } l \in \mathbb{Z}\,.$$

Avec un tel choix de fonction initiale, l'algorithme converge dans $L^\infty(\mathbb{R})$, à la vitesse $O(2^{-(N+\varepsilon)})$.

Pour démontrer cette propriété, I. Daubechies et J. Lagarias utilisent une méthode différente de la notre. On peut cependant retrouver ce résultat en appliquant le Théorème 3, ceci bien-sûr en supposant que φ satisfait les hypothèses du théorème. On peut facilement vérifier que f_0 satisfait les conditions,

$$\sum_{l\in\mathbb{Z}}(x-l)^k f_0(x-l) = \sum_{l\in\mathbb{Z}}(x-l)^k \varphi(x-l).$$

Au delà d'une rapidité de convergence optimale, ce choix de fonction initiales nous offre une autre propriété intéressante. les valeurs, sur la grille $2^{-j}\mathbb{Z}$, de f_j et de ses dérivées jusqu'à l'ordre m, correspondent à celle de φ et de ses dérivées.

$$(12) \qquad \partial^k f_j(2^{-j}l)=\partial^k\varphi(2^{-j}l), \quad \text{pour tous } k=0,\dots,m\,,\ l\in\mathbb{Z}\,.$$

Cette méthode a cependant l'inconvénient d'imposer, à l'utilisateur, le calcul préalable des valeurs des dérivées de φ sur les entiers. Ceci n'est, en fait, pas nécessaire, comme le montre l'exemple suivant.

EXEMPLE 4.5. Pour initialiser l'algorithme, il est possible d'employer des fonctions splines, d'une autre manière. On suppose encore que $\varphi \in C^{N+\varepsilon}(\mathbb{R}^n)$ avec $N\in\mathbb{N}$ et $0<\varepsilon<1$. Mais on prend, cette fois-ci, pour f_0, la fonction spline d'ordre N définie de la façon suivante.

On a

$$f_0(x) = \sum_{l\in\mathbb{Z}} a_l\, \chi_{[0,1]}^{*(N+1)}(x-l)\,,$$

où $\chi_{[0,1]}^{*(N+1)}$ est la fonction indicatrice de l'intervalle $[0,1]$ convolée N fois par elle même, et la suite $\{a_l\}_{l\in\mathbb{Z}}$ est choisie telle que

$$f_0(l)=\varphi(l)\,, \qquad \text{pour tout } l\in\mathbb{Z}\,.$$

La fonction f_0 satisfait alors les conditons habituelles et d'après le Théorème 3, il y a convergence dans $L^\infty(\mathbb{R}^n)$ à la vitesse $2^{-(N+\varepsilon)j}$.

On a eu besoin, préalablement, de calculer uniquement les valeurs de φ sur $\mathbb{Z}$ sans se soucier des dérivées de φ.

On a cependant perdu, par rapport au choix précédent, la propriété $\partial^\alpha f_j(2^{-j}l) = \partial^\alpha\varphi(2^{-j}l)$, pour tout $l\in\mathbb{Z}$ pour $\alpha=0,\dots,N$. (Nous n'avons, de plus, conservé que la convergence dans $C^N(\mathbb{R})$, alors que l'autre choix nous fournissait la convergence dans $C^{N+\varepsilon}(\mathbb{R})$. Pour remédier à ce problème, il suffit de prendre des splines d'ordre $N+1$.)

Les deux exemples que nous venons de donner, présentent également un inconvénient en commun: il est indispensable de connaître un majorant de l'indice de régularité de φ pour obtenir une vitesse de convergence optimale. Il faut adapter le choix de N à celui de φ, ce qui suppose une étude préalable de la régularité de la fonction d'échelle.

Nous présentons, dans le paragraphe suivant, un sixième exemple qui nous permet de contourner cette difficulté.

EXEMPLE 4.6. Un autre choix possible pour initialiser l'algorithme en cascade, nous est proposé par F. Paiva [16].

Commençons par remarquer que, lorsque φ est à support compact, le fait de ne pas prendre f_0 à support compact ne pose pas de problème. En effet, la fonction f_0 n'intervient que dans une deuxième étape de l'algorithme en cascade qui consiste à interpoler la suite $\{\langle\varphi, \tilde{\varphi}_{jl}\rangle\}_{l\in\mathbb{Z}^n}$ à l'aide de $2^{nj/2} f_0(2^j x - l)$. Cette suite est finie, et la longueur du support de φ nous est donné par la longueur N du filtre m_0. Pour déterminer $f_j(x)$, pour tout réel x donné sur le support $[0, N-1]$ de φ, nous n'avons alors besoin que de la restriction de f_0 à l'intervalle $[-(N-1), 2^j(N-1)]$. Tout se passe donc comme si f_0 était à support compact, alors qu'en théorie, elle ne l'est pas. Et on peut même prendre $\hat{f}_0$ à support compact, ce qui n'aurait pas été possible si f_0 l'était.

F. Paiva considère une fonction g de la classe de Schwartz dont le support de la transformée de Fourier est dans $[-2\pi, 2\pi]$, et qui vérifie

$$\sum_{l\in\mathbb{Z}} \hat{g}(\xi + 2l\pi) \equiv 1\,.$$

On a donc $\partial^\alpha \hat{g}(2l\pi) = 0$ pour $l \in \mathbb{Z}, \alpha \in \mathbb{N}$ sauf pour $\hat{g}(0)$ qui est égal à 1, et on a $g(l) = \delta_{0l}$ pour $l \in \mathbb{Z}$ (on peut, par exemple, prendre $\hat{g} = |\hat{\Phi}|^2$ où Φ est la fonction d'échelle de [13]). On pose ensuite

$$f_0(x) = \sum_{l\in\mathbb{Z}} \varphi(l)\, g(x-l)\,. \tag{13}$$

Cette fonction f_0 vérifie les conditions

$$\sum_{l\in\mathbb{Z}} (x-l)^\alpha f_0(x-l) = \int y^\alpha\, \varphi(y)\, dy\,.$$

Il y a donc convergence dans $L^\infty(\mathbb{R})$ à la vitesse $O(2^{-sj})$ et dans $C^{s-\varepsilon}(\mathbb{R})$ pour tout $\varepsilon > 0$, dès que φ appartient à $C^s(\mathbb{R})$.

Nous retrouvons, de plus, la propriété du premier exemple,

$$f_j(2^{-j}l) = \varphi(2^{-j}l)\,, \qquad \text{pour tout } l \in \mathbb{Z}\,.$$

Mais nous n'avons plus la même propriété pour les dérivées.

References.

[1] Cavaretta, A. S., Dahmen, W., Micchelli, C., A., Stationary subdivision. *Mem. Amer. Math. Soc.* **453** (1991), 1-186.

[2] Dahmen, W., Micchelli, C. A., Stationnary subdivision, fractals and wavelets. Technical report, I.B.M., 1989.

[3] Dahmen, W., Micchelli, C. A., Using the refinement equation for evaluating integrals of wavelets. *SIAM J. Numer. Anal.* **30** (1993), 507-537.

[4] Daubechies, I., Orthonormal bases of compactly supported wavelets. *Comm. Pure Appl. Math.* **41** (1988), 909-996.

[5] Daubechies, I., Lagarias, J., Two-scale difference equations II, local regularity, infinite products of matrices and fractals. *SIAM J. Math. Anal.* **23** (1992), 1031-1079.

[6] de Rham, G., Sur une courbe plane. *J. de Math.* **35** (1956), 25-42.

[7] Deslauriers, G., Dubuc, S., Interpolation dyadique. In *Fractal, dimensions non entières et applications.* Masson, 1987.

[8] Deslauriers, G., Dubuc, S., Symetric iterative interpolation processes. *Constr. Approx.* **5** (1989), 49-68.

[9] Dubuc, S., Interpolation through an iterative scheme. *Journal Math. Anal. & Applications* **114** (1986), 185-204.

[10] Durand, S., Etude de la vitesse de convergence de l'algorithme en cascade intervenant dans la construction des ondelettes. Thesis, Université Paris IX-Dauphine, 1993.

[11] Dyn, N., Gregory, J. A., Levin, D., Analysis for curve design. *Constr. Approx.* **7** (1991), 127-147.

[12] Lemarié, P. G., Fonctions à support compact dans les analyses multirésolutions. *Revista Mat. Iberoamericana* **7** (1991), 157-182.

[13] Lemarié, P. G., Meyer, Y., Ondelettes et bases hilbertiennes. *Revista Mat. Iberoamericana* **2** (1986), 1-18.

[14] Meyer, Y., *Ondelettes et Opérateurs I. Ondelettes.* Herman, 1990.

[15] Meyer, Y., *Ondelettes et Opérateurs II. Opérateurs de Calderón-Zygmund.* Herman, 1990.

[16] Paiva, F., Thesis, Université Paris IX-Dauphine, 1994.

[17] Villemoes, L. F., Wavelet analysis of refinement equations. *SIAM J. Math. Anal.* **25** (1994), 1433-1460.

Recibido: 14 de octubre de 1.994
Revisado: 25 de octubre de 1.995

Sylvain Durand
CEREMADE
Université Paris IX-Dauphine
FRANCE
durand@ceremade.dauphine.fr

Revista Matemática Iberoamericana
Vol. 12, N.° 2, 1996

Boundary Harnack Principle for separated semihyperbolic repellers, harmonic measure applications

Z. Balogh and A. Volberg

0. Introduction, notations and main results.

0.1. Boundary Harnack Principle.

There is an extensive literature on Boundary Harnack Principle (BHP). The results are of the following type: let Ω be a domain in $\mathbb{R}^n$ with a certain geometric property of $\partial\Omega$ and let u, v be positive harmonic functions on Ω vanishing on $V \cap \partial\Omega$ (V is an open set), then there exists a constant $C = C(\Omega, V, K)$ such that

$$\frac{u(x)/v(x)}{u(y)/v(y)} \le C\,, \qquad \text{for } x, y \in K \cap \Omega\,, \tag{0.1}$$

where K is a compact in V.

For domains with Lipschitz boundary this was proved independently by Ancona [A1] and Wu [W]. These results have been extended later by Bass and Burdzy [BB] to Hölder domains and L-harmonic functions, where L is a uniformly elliptic operator with bounded coefficients. Here we need a stronger version of BHP similar to the one in nontangentially accessible domains (NTA) which is due to Jerison and Kenig [JK]. Namely they proved a Hölder estimate:

$$(0.2) \qquad \left|\frac{u(x)/v(x)}{u(y)/v(y)} - 1\right| \le C(\Omega, V, K)\,|x-y|^{\epsilon}$$

for $x, y \in K \cap \Omega$ if Ω is an NTA domain in $\mathbb{R}^n$.

A byproduct of their approach and important ingredient in proving (0.2) is a fundamental property of harmonic measure in NTA domains:

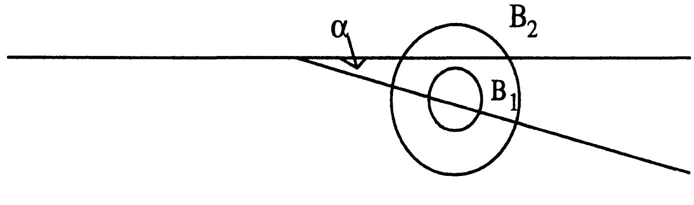
$$(0.3) \qquad \omega(B(x, 2R)) \le C\,\omega(B(x, R)),$$

where $B(x, R)$ denotes *the euclidean ball* and $x \in \partial\Omega$. Inequality (0.3) is called *the doubling condition.*

Note that simply connected NTA domains in $\mathbb{R}^2$ are just quasidiscs and (0.2), (0.3) are standard.

We are going to extend these results to a wider class, namely the class of John domains. We send the reader to [Po] or [NV] for the detailed account on this class of domains. Also for terminology such as *John constant*, etc., see [Po], [NV].

Johnness is a very natural generalization of NTA property. However (0.3) and (0.2) are false even for simply connected John domains (*i.e. John discs*). Here is a simple example:



In this example Ω is the complement of three segments meeting at the origin. Then $\omega_\Omega(B_1) \sim r^{\pi/(\pi-\alpha)}$ and $\omega_\Omega(B_2) \sim r \gg r^{\pi/(\pi-\alpha)}$.

However, harmonic measure of John discs satisfies a certain doubling condition. One needs to replace *euclidean balls* by *the balls in internal metric* ρ ($\equiv$ *infimum diameters of curves*): for $Q \in \partial\Omega$, let

$$B_\rho(Q, r) = \{x \in \bar{\Omega} : \text{there exists } \gamma_{Q,x} \text{ such that } x, Q \in \gamma_{Q,x}, \; \gamma_{Q,x} \setminus \{Q, x\} \subset \Omega, \text{ and } \operatorname{diam}\gamma_{Q,x} \le r\}.$$

Using the properties of the Riemann mapping (see [Po]) one can now prove:

Proposition 1 *Let Ω be a John disc. Then there exists $C_\Omega < \infty$ such that for all $Q \in \partial\Omega$, $r > 0$,*

$$\omega(\Omega, B_\rho(Q, 2r)) \leq C_\Omega\, \omega(\Omega, B_\rho(Q, r))\,. \tag{0.4}$$

To formulate results of the first part of our paper let us introduce the following notion of uniformly John domain.

A domain Ω on the Riemann sphere is called *a uniformly John domain* if for any two points $x_1, x_2 \in \Omega$ there exists a curve $\gamma = \gamma_{x_1,x_2}$ connecting x_1 to x_2 and lying in Ω such that

i) for all $\xi \in \gamma$, $\operatorname{dist}(\xi, \partial\Omega) \geq c_1 \operatorname{dist}(\xi, \{x_1, x_2\})$;

ii) $\operatorname{diam}\gamma \leq C_2\, \rho(x_1, x_2)$.

In this definition "dist" and "diam" are understood in the spherical metric.

John discs are uniformly John domains by the results of Gehring, Hag and Martio [GHM] and Näkki, J. Väisälä [NV]. One can choose the hyperbolic geodesics to serve as γ_{x_1,x_2}.

Our main goal is to extend (0.2) and (0.3) to a broader class of domains. The class of John domains would be a good candidate if we replace the spherical metric by the internal metric. The change of metric is partially justified by the above example and proposition.

But it turns out that (0.2) and (0.3), even with the corresponding change of metric are generally false. This is shown by the example after Theorem 3.2. (Our example concerns (0.2) but may be used as well to disprove (0.3).)

To have the extension of (0.2) and (0.3) one wishes to have a certain approximate "self- similarity" of the boundary which is given, for example by uniform Johnness (see Proposition 2).

One can imagine a uniformly John domain as a domain which is uniformly thick at "every scale"(see Proposition 2 below). We also need the complement to be uniformly thick in the sense of potential theory. Here comes the widely used condition of uniformly perfectness of the boundary (UP). Let us recall that a set E is called *uniformly perfect with* UP *constant* $\alpha > 0$ if

$$\operatorname{cap}(B(x,r) \cap E) \geq \alpha \operatorname{cap}(B(x,r)), \quad \text{for all } x \in E, \text{ for all } r \leq \operatorname{diam} E .$$

Here are some of our results:

Theorem 3.1. *Let Ω be a uniformly John domain with uniformly perfect boundary. Then harmonic measure satisfies the doubling condition in the sense that* (0.4) *holds.*

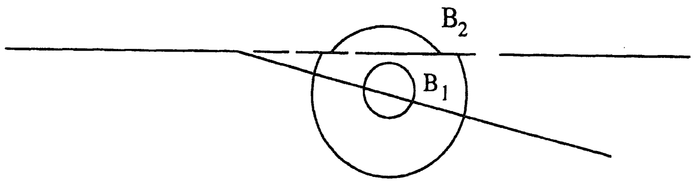
Theorem 3.2. *Let Ω be a uniformly John domain with uniformly perfect boundary. And let u, v be harmonic functions as in* (0.2). *Then for $x, y \in K$ we have*

$$\left| \frac{u(x)/v(x)}{u(y)/v(y)} - 1 \right| \leq C(\Omega, V, K)\, \rho(x,y)^{\varepsilon} .$$

The above formulation of the result is stated in Section 3 as Corolary 3.3 and it follows immediately from Theorem 3.2.

Neither uniform Johnness nor UP is necessary for (0.4). It is not difficult to construct the corresponding examples. We do not know necessary and sufficient conditions neither for (0.4) nor for (0.5). On the other hand in our dynamical applications we always have the UP property as a natural feature. Uniform Johnness is also available for a wide class of holomorphic dynamical systems.

On the other hand, in a certain sense Theorems 3.1, 3.2 are sharp. At least there is no hope that they hold for an arbitrary John domain. The next example is an easy modification of the example above. We are grateful to Juha Heinonen for it -our initial one was more complicated. Domain Ω now is the complement of three segments united with $\cup(x_n^1, x_n^2)$, where $(x_n^2 - x_n^1)/x_n^1 \to 0$. It is a John domain and it is not uniformly John. The ball B_1 has harmonic measure of the order $r^{1+\varepsilon}$. As to B_2 it has a "mushroom" in its upper part. The size of this mushroom is $O(r)$. So the doubling condition (0.4) fails. By similar considerations the conclusion of Theorem 3.2 also fails for this domain. The conclusion: *Change of metric helps only if we can localize Ω in the sense which we are going to describe.*



The main difficulty of the proofs comes from the "generic" infinite connectivity of Ω. To overcome this, we use an important feature of uniformly John domains. That is: a geometric localization.

This property is similar to the one proven by P. Jones in [Jo] for NTA domains and it was used essentially by Jerison and Kenig in [JK] to prove (0.3). We would like to mention that there is also an abstract theory of BHP (on graphs, Riemann surfaces; see [A2]) which also suggests the use of localization.

To formulate this property, let Ω be a John domain and $Q \in \partial\Omega$, $r > 0$. We say that Ω admits (K, M, N)-localization in Q at the scale r if there exist John domains $\{\Omega_Q^\ell(r)\}_{\ell=1,\dots,N}$ such that

1) $\Omega_Q^\ell(r) \subset \Omega$, and John constants of $\Omega_Q^\ell(r)$ are bounded by K;

2) $\cup_\ell \Omega_Q^\ell(r) \supset B_\rho(Q, r)$;

3) $\operatorname{diam} \Omega_Q^\ell(r) \leq M\, r$;

4) $\Omega_Q^i(r) \cap \Omega_Q^j(r) = \varnothing$, $i \neq j$.

The domain Ω is called *John localizable* if it is (K, M, N)-localizable for all $Q \in \partial\Omega$ and for all r, $0 < r < \operatorname{diam} \Omega$ with *uniform bounds* on K, M and N.

For details on this property we refer to [BV1] and [BV2], where the following assertion was proved:

Proposition 2. *A John domain is uniformly John if and only if it is John localizable.*

0.2. Applications to holomorphic dynamical systems.

In the second part of the paper we are going to apply our results to study the harmonic measure on Julia sets of a large class of dynamical systems.

Let us remind that a generalized polynomial-like map (GPL) is a triple (f, V, U) where U is a topological disc and V is the finite union of topological discs with disjoint closures and $f : V \to U$ is a branched (or regular) covering. The limit set ($\equiv$ *Julia set*) is $J_f = \partial K_f$, where $K_f = \bigcap_{n\geq 0} f^{-n}(U)$ is the filled Julia set. The dynamics is called *semihyperbolic* if

1) there are no parabolic points on J_f;

2) all critical points on J_f are separated from their orbit:

$$\operatorname{dist}(c, orb(c)) \geq \Delta > 0, \qquad \text{for all } c \in \operatorname{Crit}(f) \cap J_f .$$

The reader may find the detailed account on semihyperbolicity *e.g.* in [CJY], [Ma], [BV1], [BV2]. What we need is the important result of [CJY] which we formulate in a form convenient for us:

Theorem A ([CJY]). *For a generalized polynomial-like system* (f,V,U), *the domain* $A_\infty(f) := \bar{\mathbb{C}} \setminus K_f)$ *is a John domain if* f *is semihyperbolic.*

Another important fact is a recent result by Mañé and da Rocha [MR], Hinkkanen [H] and Eremenko [E].

Theorem B. *Let* (f, V, U) *be a generalized polynomial-like map. Then* J_f *is uniformly perfect.*

These results were proven in the context of rational dynamics but the proofs can be carried over without major modifications. Or, one can use the fact that a GPL is quasiconformally (qc) conjugated to a polynomial. Johnness and the UP property are quasiconformally invariant.

We may ask when a GPL (f, U, V) gives rise to a uniformly John $A_\infty(f)$. For general semihyperbolic GPL this is not true (see [BV2]). On the other hand the uniform Johnness of $A_\infty(f)$ holds for the class of so called *separated semihyperbolic* GPL. We are going to recall this definition from [BV2].

Let us remind that K_f denotes the filled Julia set of f. For $x \in K_f$ we denote by K_x the component of connectivity of K_f containing x. We split the critical points in K_f in two parts:

$$C_1 = \{c_1 \in K_f : \ c_1 \text{ is critical point of } f, \ \ K_{c_1} = \{c_1\}\},$$
$$C_2 = \{c_2 \in K_f : \ c_2 \text{ is critical point of } f, \ \ K_{c_2} \neq \{c_2\}\}.$$

After this splitting we have:

Definition 1. *A semihyperbolic* GPL (f, U, V) *is called separated if there is* $\delta > 0$ *such that*

$$\operatorname{dist}(\operatorname{orb}(c_1), K_{c_2}) > \delta,$$

for any $c_1 \in C_1$ *and* $c_2 \in C_2$.

The following result from [BV2] shows that uniformly John domains appear naturally in dynamics:

Theorem C ([BV2]).

1. *If* J_f *is is totally disconnected, then the domain* $A_\infty(f) := \bar{\mathbb{C}} \backslash K_f$ *is a uniformly John domain if and only if* $A_\infty(f)$ *is a John domain.*

2. *Suppose that the generalized polynomial-like system* (f, V, U) *is semihyperbolic. The domain* $A_\infty(f)$ *is uniformly John if and only if* f *is separated semihyperbolic.*

So uniformly John domains can be met rather often. In the second part of this paper we are going to use Theorem 3.2 to prove a certain *rigidity result* of the harmonic measure on Julia sets of separated semihyperbolic GPL. To explain this result let us recall that a GPL (g, U, V) is called *maximal* if $\omega_g = m_g$, where ω_g is the harmonic measure on J_g e valuated at infinity and m_g denotes the measure of maximal entropy. We can now recall from [BPV]:

Definition 2. *The* GPL (f, U, V) *is called conformally maximal if* f *is conformally conjugated in a neighborhhod of* J_f *to a* GPL $(g, \tilde{U}, \tilde{V})$ *which is maximal* .

It is clear that for conformal maximality of (f, U, V) it is necessary to have $\omega_f \approx m_f$. By *rigidity* of harmonic measure, we understand the sufficiency of this condition. Our result in this direction is:

Theorem 5.5. *Let (f, V, U) be a generalized polynomial-like map which is separated semihyperbolic. Then, (f, U, V) is conformally maximal if and only if harmonic measure is absolutely continuous with respect to measure of maximal entropy.*

In particular if, in addition (f, V, U) is polynomial-like in the sense of Douady and Hubbard then f is conformally equivalent to a polynomial map if and only if harmonic measure is absolutely continuous with respect to measure of maximal entropy.

Note that infinite connectivity of $A_\infty(f)$ is a main difficulty again. Particular cases of this result can be found in [LyV] (hyperbolic case), [BPV] (J_f is a Cantor set). Our result is related to results of Lopes [Lo] and Mañe, Da Rocha [MR].

Commenting again the first part let us mention that throughout it we play with ideas virtually present in [A1], [W], [JK]. This is why Section 1 unites the results which have *exactly* the same proofs as corresponding results of [JK]. We give only the statements and references. Unfortunately two key results of [JK] cannot be imitated in our setting. The reason is that for NTA domains internal metric is equivalent to euclidean and for *uniformly John domains* this is no longer the case. Section 2 contains the proof of these stubborn results. In Section 3, Theorems 3.1 and 3.2 are proved.

In the second part of the paper we consider applications in dynamics. In Section 4 we apply Theorem 3.2 to construct a lifting of the harmonic measure to a one sided shift space.

In Section 5 we use this shift model of the harmonic measure to prove Theorem 5.5 .

Commenting on the assumptions of Theorem 5.5, we believe that the result is true for general f (without any condition of separated semihyperbolicity.)

The results of this article were partially announced in [BV3].

1. Estimates of harmonic measure and Green's function in John domains.

Throughout this section domains are John domains with John constants at most K and having uniformly perfect boundary with UP constant at least α.

Lemma 1.1. *There exists $\beta > 0$ and $r_0 > 0$ such that for all $Q \in \partial\Omega$ and $r < r_0$; and every positive harmonic function u in Ω, if u vanishes continuously on $B(Q,r) \cap \partial\Omega$, then there exists $M = M(\alpha)$ such that for $z \in \Omega \cap B(Q,r)$*

$$u(z) \le M \left(\frac{|z-Q|}{r}\right)^{\beta} \max_{\xi \in \Omega \cap \partial B(Q,r)} u(\xi). \tag{1.1}$$

The proof is exactly the same as for [JK, Lemma 4.1].

Lemma 1.2. *Let Ω be a John domain with John center A and diameter D. Let $Q \in \partial\Omega$, $d > 0$ and $\Omega' \subset \Omega$ be a domain with uniformly perfect $\partial\Omega'$ such that $\Omega \setminus \Omega' \subset B(Q, d/100)$. If u is positive harmonic in Ω' and u vanishes continuously on $B(Q,d) \cap \partial\Omega'$, then*

$$u(z) \le M' \, u(A) \tag{1.2}$$

for all $z \in B(Q, d/2) \cap \Omega'$. Here $M' = M'(d/D, K, \alpha)$.

The proof is exactly the same as for [JK, Lemma 4.4]. See also Carleson's article [C, p. 398].

In what follows $\delta(z) = \operatorname{dist}(z, \partial\Omega)$ and $G(\cdot, P)$ denotes Green's function for the domain Ω.

Lemma 1.3. *Let Ω be a John domain. Let $Q \in \partial\Omega$. Then there exists $M = M(K,\alpha)$ such that for every $r > 0$, $z \in B(Q, r/2)$ and $P \in \Omega \setminus B(z, \delta(z)/2)$*

$$G(z,P) \le M \, \omega(B_\rho(Q,r), P).$$

The proof is exactly the same as for [JK, (4.3)].

Lemma 1.4. *Let Ω be a John domain with John center A and diameter D. Let $d > 0, Q, Q' \in \partial\Omega, P \in \Omega \setminus (B_\rho(Q, 2d) \cup B(Q', 2d))$, $|P - A| \ge d$. Let Ω' be a subdomain of Ω with uniformly perfect $\partial\Omega'$ and such that $\Omega \setminus \Omega' \subset B(Q', d/100)$. Then there exists $M' = M'(D/d, K, \alpha) < \infty$ such that:*

$$G_{\Omega'}(A,P) \le M' \cdot \omega(\Omega', B_\rho(Q,d), P).$$

The independence of P is important here.

REMARK 1.5. The statement of the lemma stays true if we replace "B_ρ" by "B" everywhere.

2. Estimates of harmonic measure and Green's functions in uniformly John domains.

We first prove some assertions concerning arbitrary John domains. We have not been able to follow the lines of [JK], so the proofs are given. As before K is a John constant and $\alpha > 0$ is a UP constant of $\partial\Omega$.

Lemma 2.1. *Let Ω be a John domain with John center A and diameter D. Let $d > 0, Q, Q' \in \partial\Omega$, $P \in \Omega \backslash (B_\rho(Q, 20\,d) \cup B(Q', 20\,d))$, $|P - A| \geq 10\,d$. Let Ω' be a subdomain in Ω with uniformly perfect $\partial\Omega'$ and such that $\Omega \backslash \Omega' \subset B(Q', d/100)$. Then there exists $M' = M'(D/d, K, \alpha) < \infty$ independent of P such that:*

$$\omega(\Omega', B_\rho(Q, d), P) \leq M' \, G_{\Omega'}(A, P).$$

PROOF. We consider only the case $\Omega = \Omega'$. The general case follows exactly the same lines but it is more tedious. Let $B(Q) = B(Q, 3d/2)$, $B(P) = B(P, d/4)$. Lemma 1.2 implies that (consider $\Omega \backslash B(P, \delta(P)/4)$):

$$G(z, P) \leq M_1' \, G(A, P), \qquad z \in \Omega \setminus B(P). \tag{2.1}$$

Points P and Q are $20\,d$-far apart in internal metric, but may not be even $2d$-far apart in euclidean metric. Still, if they are far enough in euclidean sense, namely, if

$$B(Q) \cap B(P) = \varnothing, \tag{2.2}$$

we carry out the proof of [JK, (4.6)] as follows. Let φ be a C^∞-function supported by $B(Q)$ such that $\varphi \equiv 1$ exactly on $B(Q, d)$. Then using (2.1) and the estimate $|\Delta\varphi| \leq C \cdot d^{-2}$ we get

$$\omega(B(Q, d), P) \leq \int \varphi \, d\omega(\cdot, P) = |\langle \varphi, \Delta G(\cdot, P) \rangle|$$

$$
\begin{aligned}
&= \left| \int G(z,P)\, \Delta\varphi \, dA(z) \right| \\
&\le M'\, G(A,P) \int_{B(Q)} d^{-2}\, dA(z) \\
&\le M'\, G(A,P)\,,
\end{aligned}
$$

which is even more than we wanted to prove: the left part is harmonic measure of $B(Q,d)$ and $B(Q,d) \supset B_\rho(Q,d)$.

Now assume the negation of (2.2):

$$B(Q) \cap B(P) \neq \varnothing\,. \tag{2.3}$$

We cannot modify the proof by modifying φ above. If we try $\varphi \equiv 1$ on $B_\rho(Q,d)$ and with support in $B_\rho(Q,3d/2)$ (which would be appropriate) we cannot claim that $|\Delta\varphi| \sim d^{-2}$. This might not be true.

Given (2.3) let $\mathcal{O}(P)$ denote the component of $\Omega \cap B(P,5d)$ which contains P. Let $\Gamma(P)$ be the part of $\partial\mathcal{O}(P)$ which is not in $\partial\Omega$; $\Gamma(P)$ consists of parts of the circle $\partial B(P,5d)$ which separate P from A. By our assumptions

$$\mathcal{O}(P) \cap B_\rho(Q,d) = \varnothing. \tag{2.4}$$

Let P' denote an arbitrary point on $\Gamma(P)$. By (2.3)

$$B(P') := B(P', d/4) \cap B(Q) = \varnothing\,.$$

Thus, the first part of the proof shows that

$$\omega(B_\rho(Q,d), P') \le M'\, G(A,P')\,. \tag{2.5}$$

Now the minimum principle is applicable to the harmonic function

$$M'\, G(A,z) - \omega(B_\rho(Q,d), z)\,, \qquad z \in \mathcal{O}(P)\,.$$

It is non-negative on $\partial\mathcal{O}(P)$ by (2.4) and (2.5). Evaluating it at $P \in \mathcal{O}(P)$ we finish the proof of Lemma 2.1.

REMARK 2.2. The statement of the lemma holds if we replace everywhere "B_ρ" by "B".

Lemma 2.3. *Let Ω be a John domain with center A and diameter D. Let $0 < d < D/10$, $Q \in \bar{\Omega}$, $|Q - A| \geq 2d$. Let Ω_1 be a John domain inside $\Omega \cap B_\rho(Q, d/4)$ with the same John and UP constants. We assume that*

$$\operatorname{dist}_\rho(Q, \Omega \cap \partial\Omega_1) \geq c_1\, d\,.$$

Let u, v be two positive harmonic functions in Ω continuous in $\bar{\Omega}$ and vanishing on $B(Q, d) \cap \partial\Omega$. Let $\Omega_1' := \Omega_1 \cap B_\rho(Q, c_1 d/(4K))$. We put

$$M = \max_{\Omega} \frac{u}{v}\,, \qquad m = \min_{\Omega} \frac{u}{v}\,, \qquad \ell = \frac{M}{m} - 1\,,$$
$$M_1 = \max_{\Omega_1'} \frac{u}{v}\,, \qquad m_1 = \min_{\Omega_1'} \frac{u}{v}\,, \qquad \ell_1 = \frac{M_1}{m_1} - 1\,.$$

There exists $q \in (0,1)$, $q = q(d/D, K, \alpha, c_1)$, such that

$$\ell_1 \leq q\,\ell\,.$$

PROOF. Put $\Gamma = (\partial\Omega) \setminus B(Q, d)$. Without loss of generality we may assume that

$$v(\xi) \leq u(\xi) \leq (1 + \ell)\, v(\xi)\,, \qquad \xi \in \Gamma\,. \tag{2.6}$$

We are going to prove that for all $\xi \in \Omega_1'$ at once: either

$$v(\xi) \leq u(\xi) \leq (1 + q\ell)\, v(\xi)\,, \tag{2.7}$$

or

$$(1 + (1 - q)\ell)\, v(\xi) \leq u(\xi) \leq (1 + \ell)\, v(\xi)\,. \tag{2.8}$$

Any of these two relationships proves the lemma.

The left inequality (2.7) and the right one (2.8) are clear from (2.6) as

$$u = v = 0 \;\text{ on }\; (\partial\Omega) \cap B(Q, d) = (\partial\Omega) \setminus \Gamma\,.$$

Let us denote

$$\Gamma_1 = \left\{\xi \in \Gamma :\; u(\xi) \leq \left(1 + \frac{\ell}{2}\right) v(\xi)\right\}, \qquad \Gamma_2 = \Gamma \setminus \Gamma_1\,.$$

We put

$$v_i(z) = \int_{\Gamma_i} v(\xi)\, d\omega(\Omega, \xi, z)\,, \qquad i = 1, 2\,.$$

We first consider the case

$$v_1(A) \geq v_2(A)\,. \tag{2.9}$$

Starting from (2.9) we will prove the right inequality (2.7). The negation of (2.9) leads to the left inequality (2.8).

First step. We are going to construct a special disc on which (2.9) is nearly satisfied.

Let A_1 denote the John center of Ω_1. As by assumption $\operatorname{diam}\Omega_1 \geq c_1 d$ we get $\operatorname{dist}(A_1, \partial\Omega_1) \geq c_1 d/(2K)$. Let us consider a John arc in Ω connecting A and A_1. It intersects $\partial\Omega_1$ in, say, A_2. We have $|A_2 - A_1| \geq \operatorname{dist}(A_1, \partial\Omega_1) \geq c_1 d/(2K)$ and John property implies $\delta(A_2) := \operatorname{dist}(A_2, \partial\Omega) \geq c_1 d/(2K^2)$. Put

$$\beta = \min\left\{\frac{c_1 d}{4K^2}, \frac{c_1 d}{10}\right\}$$

and consider $B(A_2, \beta)$. By Harnack's inequality and (2.9)

$$v_1(\eta) \geq \gamma\, v_2(\eta)\,, \qquad \text{for all } \eta \in B(A_2, \beta)\,, \tag{2.10}$$

where $0 < \gamma = \gamma(d/D, K, c_1)$.

Second step. Let us prove that

$$u(\eta) \leq (1 + q'\ell)\, v(\eta)\,, \qquad \text{for all } \eta \in B(A_2, \beta)\,, \tag{2.11}$$

for a certain $q' \in (0,1)$, $q' = q'(d/D, K, c_1)$.

Apply Poisson formula to the function u in Ω:

$$\begin{aligned} u(\eta) &= \int_\Gamma u(\xi)\, d\omega(\Omega, \xi, \eta) \\ &= \int_{\Gamma_1} + \int_{\Gamma_2} \\ &\leq \left(1 + \frac{\ell}{2}\right) \int_{\Gamma_1} v(\xi) + \cdots + (1 + \ell) \int_{\Gamma_2} v(\xi) \cdots \\ &:= \left(1 + \frac{\ell}{2}\right) v_1(\eta) + (1 + \ell)\, v_2(\eta)\,. \end{aligned}$$

Now $(v_1+v_2)/v=1$ and $v_1 \geq \gamma v_2$ on $B(A_2,\beta)$. Thus (2.11) follows.

If (2.11) were true in Ω_1' $(:= \Omega_1 \cap B_\rho(Q, c_1 d/4))$ we would be already done.

To push η in (2.11) from $B(A_2,\beta)$ to Ω_1' we need several steps more. Let $\partial_{\text{dist}}\Omega_1 = (\partial\Omega_1)\cap\Omega$.

Third step.

$$v(\eta) \geq \gamma' v(A), \qquad \eta \in B(A_2,\beta), \tag{2.12}$$

$$v(\eta) \leq M' v(A), \qquad \eta \in \bar{\Omega}_1 . \tag{2.13}$$

Here $0 < \gamma' = \gamma'(d/D, K, c_1)$, $M' = M'(d/D, K, \alpha) < \infty$. First inequality is just Harnack's inequality. The second one follows by Lemma 1.2.

Forth step. Let $\{x_i\}_{i=1}^N$ be a maximal $\beta/2$-net on $\partial_{\text{dist}}\Omega_1$ in the sense of metric $\rho_1 = \rho_{\Omega_1}$ on $\bar{\Omega}_1$. Clearly

$$N = N(K, c_1, d) < \infty . \tag{2.14}$$

Let $B^i_{\rho_1}$ denote $B_{\rho_1}(x_i, \beta/2)$.

We apply Lemmas 2.1 and 1.4 to Ω_1 and $\eta \in \Omega_1'$ playing the role of P, x_i playing the role of Q. Then

$$\omega(\Omega_1, B^i_{\rho_1}, \eta) \asymp G_{\Omega_1}(A_1, \eta), \qquad \eta \in \Omega_1' , \tag{2.15}$$

$$\omega(\Omega_1, B(A_2,\beta), \eta) \asymp G_{\Omega_1}(A_1, \eta), \qquad \eta \in \Omega_1' . \tag{2.16}$$

The constants implicitly involved here depend only on $d/D, K, \alpha$ and c_1.

Fifth step. We apply Poisson formula to u, (as in the second step) but now in Ω_1. For $\eta \in \Omega_1'(= \Omega_1 \cap B_\rho(Q, c_1 d/(4K)))$.

$$\begin{aligned} u(\eta) &= \int_{\partial_{\text{dist}}\Omega_1} u(\xi)\, d\omega(\Omega_1, \xi, \eta) \\ &= \int_{B(A_2,\beta)\cap\partial\Omega_1} u(\xi)\, d\omega(\Omega_1,\xi,\eta) + \int_{(\partial_{\text{dist}}\Omega_1)\setminus B(A_2,\beta)} u(\xi)\, d\omega(\Omega_1,\xi,\eta) \\ &\leq (1+q'\ell)\int_{\dots} v(\xi) + \cdots + (1+\ell)\int_{\dots} v(\xi) \cdots \end{aligned} \tag{2.17}$$

$$= (1 + q'\ell)\, v^1(\eta) + (1 + \ell)\, v^2(\eta)\,.$$

We are going to show that there exists $0 < \gamma'' = \gamma''(d/D, K, \alpha, c_1)$ such that

$$v^1(\eta) \geq \gamma''\, v^2(\eta)\,, \qquad \text{for all } \eta \in \Omega_1'\,. \tag{2.18}$$

To do this we proceed as follows. By (2.13)

$$\begin{aligned} v^2(\eta) &\leq \int_{\partial_{\text{dist}}\Omega_1} v(\xi)\, d\omega(\cdots) \\ &\leq \sum_{i=1}^{N} \int_{B_i} \cdots \\ &\leq M'\, v(A) \sum_{i=1}^{N} \omega(\Omega_1, B^i_{\rho_1}, \eta)\,. \end{aligned}$$

Now (2.14), (2.15) and (2.16) imply that for $M'' = M''(d/D, K, \alpha, c_1) < \infty$

$$v^2(\eta) \leq M''\, v(A)\, \omega(\Omega_1, B(A_2, \beta), \eta)\,. \tag{2.19}$$

On other hand by (2.12)

$$v^1(\eta) = \int\limits_{B(A_2,\beta)\cap\partial_{\text{dist}}\Omega_1} v(\xi) d\omega(\cdots) \geq \gamma'\, v(A)\, \omega(\Omega_1, B(A_2, \beta), \eta)\,.$$

Together with (2.19) this gives (2.18).

Taking into account that $(v^1 + v^2)/v = 1$ and (2.17), (2.18) we obtain the required inequality

$$u(\eta) \leq (1 + q\ell)\, v(\eta)\,, \qquad \eta \in \Omega_1'\,,$$

which is the right inequality in (2.7).

Completely similarly the negation of (2.9) leads to the proof of the left part of (2.8). Lemma 2.3 is completed.

From now on Ω is uniformly John, that is, by Proposition 2 it is localizable. Reminding that $\Omega^\ell_Q(r)$, $\ell = \overline{1, k(Q)}$, are local John domains with properties 1)-4) (see the introduction) we denote their John centers by $A^\ell_Q(r)$, and let K be their John constant. By G_Ω we denote Green's function with pole at A.

Clearly all $\partial\Omega^\ell_Q(r)$ are uniformly perfect (as $\partial\Omega$ was) and we denote by α, $\alpha > 0$, the least UP constant for them.

Lemma 2.4. *Let Ω be uniformly John with center A and let $0 < r \le \operatorname{diam}\Omega$, $Q \in \partial\Omega$. Then there exists $M = M(K,\alpha)$ such that*

$$\omega(\Omega, B_\rho(Q, r/2), A) \le M \sum_{\ell=1}^{k(Q)} G_\Omega(A^\ell_Q(r)).$$

PROOF. We may put $\operatorname{diam}\Omega = 1$. Let $k = k(Q)$. Fix $r > 0$ and let $\Gamma = B(Q,r) \cap \partial\Omega$. Let $\{Q_1, \ldots, Q_S\}$ be a maximal r-net of Γ in the metric ρ. Clearly $S = S(K) < \infty$ because Ω is a John domain.

In what follows $C_1, C_2, \ldots$ are large constants depending on K and α. Put $\Gamma_0 = \{z \in \Gamma : \rho(Q,z) < r/2\}$, $\tilde\Gamma = \{z \in \Gamma : \rho(Q,z) > C_1 r\}$. Let $\{Q_1, \ldots, Q_c\}$ be a subnet, $c \le S$, consisting of points $Q_i \in \tilde\Gamma$.

Let us consider the finite family $\mathcal{F} = \{\Omega^\ell_{Q_i}(C_2 r)\}$, $i = \overline{1,C}$ $\ell = \overline{1, k(Q_i)}$. Let $\tilde\Omega$ be their union. We put $\tilde B = B(Q,r) \cap \tilde\Omega$. We have to delete somehow this set because Green's function on it cannot be controlled by $\sum_\ell G_\Omega(A^\ell_Q(r))$.

First step. Is to prove that

$$z \in B(Q,r) \text{ and } \rho(z,Q) > 2C_1 r \quad \text{implies} \quad z \in \tilde B. \tag{2.20}$$

In fact, let $\mathcal{O}$ be a component of $B(Q,r) \cap \Omega$ containing z and let $P \in \partial\Omega \cap \partial\mathcal{O}$. As $\rho(P,z) < 2r$ we see that $\rho(P,Q) > C_1 r$, that is $P \in \tilde\Gamma$. Let Q_i be a ρ-closest to P from our net. Then $\rho(z, Q_i) < r + 2r = 3r$ and $z \in \Omega^\ell_{Q_i}(3r)$. Thus $z \in \tilde\Omega$ and (2.20) is proved.

Second step. Let

$$\Omega_0 = \Omega, \quad \Omega_1 = \Omega_0 \setminus \overline{\bigcup_\ell \Omega^\ell_{Q_1}(C_2 r) \cap B(Q,r)}, \quad \ldots,$$
$$\Omega_k = \Omega_{k-1} \setminus \overline{\bigcup_\ell \Omega^\ell_{Q_{k-1}}(C_2 r) \cap B(Q,r)}, \quad \ldots, \quad \Omega_c = \Omega \setminus \overline{\tilde B}.$$

Denote by $G(z), G_1(z), \ldots, G_k(z), \ldots, G_c(z)$ Green's functions of corresponding domains with pole at A. They are ordered:

$$G_c(z) \le \cdots \le G_1(z) \le G_1(z) \le G(z). \tag{2.21}$$

Harmonic functions $\omega_k(z) = \omega(\Omega_k, \Gamma_0, z) = \omega(\Omega_k, \Gamma_0 \cap \partial\Omega_k, z)$ are ordered in the same way

$$\omega_c(z) \leq \cdots \leq \omega_1(z) \leq \omega(z).$$

Let us prove first that

$$\omega_c(A) \leq M(K,\alpha) \sum_{\ell=1}^{k} G(A_Q^\ell(2r)). \tag{2.22}$$

To this end note that $B(Q,r) \cap \Omega_c \subset B_\rho(Q, 2C_1 r)$. This is just another way to state (2.20). Recall that $B_\rho(Q, 2C_1 r)$ is covered by

$$\bigcup_{\ell=1}^{k} \Omega_Q^\ell(2M_0\, C_1 r)$$

and apply Lemma 1.2 to $G(z)$ in each of this domains separately. Then

$$z \in B(Q,r) \cap \Omega_c \text{ implies } G(z) \leq M(K,\alpha) \sum_{\ell=1}^{k} G(A_Q^\ell(2M_0\, C_1 r)). \tag{2.23}$$

We may now go along the lines of [JK, (4.6)]. In fact, (2.21) and Harnack's inequality imply

$$G_c(z) \leq M \sum_{\ell=1}^{k} G(A_Q^\ell(r)), \qquad z \in B(Q,r) \cap \Omega_c. \tag{2.24}$$

Consider a C^∞-function φ supported by $B(Q,r)$ and such that $\varphi \equiv 1$ exactly on $B(Q, r/2)$. Then

$$\begin{aligned}
\omega_c(A) &= \omega(\Omega_c, B_\rho(Q, \tfrac{r}{2}), A) \\
&\leq \omega(\Omega_c, B(Q, \tfrac{r}{2}), A) \\
&\leq \int \varphi\, d\omega_c(\xi, A) \\
&= |\langle \varphi, \Delta G_c \rangle| \\
&= \Big| \int_{B(Q,r)} \Delta\varphi\, G_c(z, A)\, dA(z) \Big|
\end{aligned}$$

$$\leq M \sum_{\ell=1}^{k} G(A_Q^\ell(r))\, r^{-2} \int_{B(Q,r)} dA(z)$$

$$= M \sum_{\ell=1}^{k} G(A_Q^\ell(r)).$$

And (2.22) is proved. The key estimate was (2.23).

REMARK. We cannot hope to have (2.23) with Ω instead of $\Omega_($

Just because $\Omega \cap B(Q,r)$ may contain the points which tremely far from Q in ρ-metric. For such points there is no estimate $G(z)$ by $G(A_Q^\ell(Cr)$ if r is small. This is the main c and the main difference from NTA domains. This is why we the procedure of excluding such points and this is why we nee(

Third step.

$$\omega_{k-1}(A) \leq M(K,\alpha)\,\omega_k(A). \tag{2.25}$$

We denote $u_1 = \omega_k$, $u_2 = \omega_{k-1}$, thus having $u_1 \leq u_2$. Denote t matrix

$$\{\omega_{m,l}\}_{m,\ell=1}^{k(Q_k)} = \{\omega_k(\Omega_{Q_k}^m(C_2r) \cap \partial B(Q,r), \quad A_{Q_k}^\ell(C_2r))\}_{m,}^{k((}$$

As C_2 grows the centers $A_{Q_k}^\ell(C_2r)$ are getting more and more f from $\partial B(Q,r)$. In particular, the entries of this matrix can b(as small as we wish: C_2 rules that. To choose appropriate C notice that

$$\text{(2.26)} \quad \xi \in \partial B(Q,r) \cap \Omega_{Q_k}^\ell(C_2r) \quad \text{implies} \quad u_2(\xi) \leq C_3\, u_2(A_{Q_k}^\ell($$

This is by Lemma 1.2 applied to $\Omega_{Q_k}^\ell(r)$ and $d \sim C_2r$. Cons depends on John and UP constants of $\Omega_{Q_k}^\ell(C_2r)$, that is inde of C_2. We choose C_2 so large that $\|(I - C_3\mathcal{U})^{-1}\| \leq 2$.

To compare $u_2(A_{Q_k}^\ell(C_2r))$ and $u_1(A_{Q_k}^\ell(C_2r))$ we write d Poisson formula for u_2 in Ω_k:

$$u_2(A_{Q_k}^\ell(C_2r)) = u_1(A_{Q_k}^\ell(C_2r)) + \int_{\Omega_{k-1} \cap \partial\Omega_k} u_2(\xi)\, d\omega_k(\xi, A_{Q_k}^\ell($$

We use (2.26) and the choice of $\mathcal{U}$ to conclude

$$u_2(A^\ell_{Q_k}(C_2 r)) \le 2\,u_1(A^\ell_{Q_k}(C_2 r)). \tag{2.27}$$

Put

$$1 \ll C_2' \ll C_2\,, \qquad 1 \ll C_2'' \ll C_2'\,,$$
$$\mathcal{O}' = \Big(\bigcup_\ell \Omega^\ell_{Q_k}(C_2' r)\Big)\bigcap \Omega_{k-1}, \qquad \mathcal{O}'' = \Big(\bigcup_\ell \Omega^\ell_{Q_k}(C_2'' r)\Big)\bigcap \Omega_{k-1}\,,$$
$$T' = \partial\mathcal{O}' \cap \Omega_{k-1}\,, \qquad T'' = \partial\mathcal{O}'' \cap \Omega_{k-1}\,.$$

Clearly

$$\begin{cases} \operatorname{dist}(T', Q_k) \ge C_2' r\,, \\ C_2'' r \le \operatorname{dist}(T'', Q_k) \le \operatorname{dist}_\rho(T'', Q_k) \le \dfrac{1}{100}\, C_2' r\,. \end{cases} \tag{2.28}$$

Let $\{x_j\}_{j=1}^n$ be a maximal $C_2'' r$-net of T' in the metric of $\rho_{\mathcal{O}'}$. Clearly $n = n(K, C_2') < \infty$. Put $B_j = B_{\rho_{\mathcal{O}'}}(x_j, C_2'' r)$. Poisson formula for u_2 in $\mathcal{O}'$ gives that if $\xi \in T''$ then

$$\begin{aligned} u_2(\xi) &= \int_{T'} u_2(z)\, d\omega_{\mathcal{O}'}(z, \xi) \\ &= \sum_1^n \int_{B_j} \cdots \\ &\le C_3 \sum_\ell u_2(A^\ell_{Q_k}(C_2 r)) \sum_1^n \omega_{\mathcal{O}'}(B_j, \xi) \\ &\le 2\,C_3 \sum_\ell u_1(A^\ell_{Q_k}(C_2 r)) \sum_1^n \omega_{\mathcal{O}'}(B_j, \xi)\,. \end{aligned} \tag{2.29}$$

We used (2.26), (2.27) and Lemma 1.2 with $d \sim C_2 r$. Let a^ℓ_k be a point of intersection of T' with the John arc connecting $A^\ell_{Q_k}(C_2 r)$ with Q_k. Let S^ℓ_k denote a disc from the family $\{B_j\}$ containing a^ℓ_k, and $\tilde{S}^\ell_k$ denote $B(a^\ell_k, \delta(a^\ell_k)/2)$. Notice that

$$\delta(a^\ell_k) \ge C(K)\, C_2' r \ge 10\, C_2'' r$$

if C_2' is large enough in comparison to C_2''. In particular $S_k^\ell \subset \tilde{S}_k^\ell$. Clearly (using the notations $A_k^\ell := A_{Q_k}^\ell(C_2 r)$) if $\xi \in T''$, then

$$\begin{aligned} u_1(\xi) &= \int u_1(z)\, d\omega_{\mathcal{O}'\cap\Omega_k}(z,\xi) \\ &\geq C(K)\sum_\ell u_1(a_k^\ell)\,\omega_{\mathcal{O}'\cap\Omega_k}(\tilde{S}_k^\ell,\xi) \\ &\geq C(K, C_2/C_2')\sum_\ell u_1(A_k^\ell)\,\omega_{\mathcal{O}'\cap\Omega_k}(\tilde{S}_k^\ell,\xi)\,. \end{aligned} \tag{2.30}$$

Now we apply Lemmas 1.4 and 2.1 to the domain $\mathcal{O}'$ and the fact that $d \sim C_2'' r$ to obtain

$$\omega_{\mathcal{O}'}(B_s,\xi) \leq C(C_2'/C_2'', K, \alpha)\,\omega_{\mathcal{O}'}(B_t,\xi)\,,$$

for all $\xi \in T''$, and for all s, t.

Thus, if $\xi \in T''$ then

$$u_2(\xi) \leq C(C_2'/C_2'', K, \alpha)\, n \sum_\ell u_1(A_k^\ell)\,\omega_{\mathcal{O}'}(\tilde{S}_k^\ell,\xi)\,. \tag{2.31}$$

If we could prove that

$$\omega_{\mathcal{O}'}(\tilde{S}_k^\ell,\xi) \leq 2\,\omega_{\mathcal{O}'\cap\Omega_k}(\tilde{S}_k^\ell,\xi), \qquad \xi \in T'', \tag{2.32}$$

then (2.30)-(2.31)-(2.32) combined would imply

$$u_2(\xi) \leq C(C_2/C_2'', K, \alpha)\, u_1(\xi)\,, \qquad \text{if } \xi \in T''\,.$$

Applying this and the maximum principle to the component of $\Omega \setminus T''$ which contains A we would get (2.25).

So we are left to prove (2.32).

Let us put

$$w_2(\xi) = \omega_{\mathcal{O}'}(\tilde{S}_k^\ell,\xi)\,, \qquad w_1(\xi) = \omega_{\mathcal{O}'\cap\Omega_k}(\tilde{S}_k^\ell,\xi)\,.$$

Then $w_1 \leq w_2$. To prove the inverse inequality let us consider both functions in $\mathcal{O}' \cap \Omega_k$ and let $\gamma_k = \mathcal{O}' \cap (\partial(\mathcal{O}' \cap \Omega_k))$. From the construction we conclude that $\gamma_k \subset \partial B(Q,r)$ and

$$\operatorname{dist}_{\rho_{\mathcal{O}'}}(\gamma_k, Q_k) \leq 2r\,.$$

Clearly

$$w_2(\xi) \le w_1(\xi) + \omega_{\mathcal{O}'\cap\Omega_k}(\gamma_k, \xi)\, m\,, \tag{2.33}$$

where $m = \max\{w_2(z) : z \in \gamma_k\}$.

Introduce $v(z) = \omega(\mathcal{O}', B_{\rho_{\mathcal{O}'}}(Q_k, C(K)C_2''r), z)$. On γ_k we have by α-uniform perfectness that if $z \in \gamma_k$

$$v(z) \ge \gamma \ge \gamma\, \omega_{\mathcal{O}'\cap\Omega_k}(\gamma_k, z)\,,$$

where $\gamma = \gamma(\alpha) > 0$. On the rest of $\partial(\mathcal{O}' \cap \Omega_k)$ both functions vanish. Thus

$$\xi \in \mathcal{O}' \cap \Omega_k \quad \text{implies} \quad v(\xi) \ge \gamma\, \omega_{\mathcal{O}'\cap\Omega_k}(\gamma_k, \xi)\,. \tag{2.34}$$

Applying Lemmas 1.4 and 2.1 to $\mathcal{O}'$ and the fact that $d \sim C_2''r$ and taking into account that the radius of $\tilde{S}_k^\ell$ is at least $C(K)\, C_2'\, r$ we get that if $\xi \in T''$, then

$$\begin{aligned} v(\xi) &\le M'(C_2'/C_2'', K, \alpha)\, \omega_{\mathcal{O}'}(\tilde{S}_k^\ell, \xi) \\ &= M'(C_2'/C_2'', K, \alpha)\, w_2(\xi)\,. \end{aligned} \tag{2.35}$$

Uniting (2.33)-(2.35) we obtain that if $\xi \in T''$, then

$$w_2(\xi) \le w_1(\xi) + \gamma^{-1}\, M'(C_2'/C_2'', K, \alpha)\, m\, w_2(\xi)\,.$$

By an obvious extremal length estimate

$$m = \max\{w_2(z) : z \in \gamma_k\}$$

can be made as small as we wish by taking C_2' very large. We keep the ratio C_2'/C_2'' bounded as we make C_2' so large that

$$m\,\gamma^{-1}\, M'(C_2'/C_2'', K, \alpha) < \frac{1}{2}\,.$$

Then, (2.32) follows and the lemma is completed.

3. Doubling condition and BHP for uniformly.

3.1. John domains.

As always ρ denotes the internal metric.

Theorem 3.1. *Let Ω be a uniformly John domain with constant K and center A, and let $\partial\Omega$ be uniformly perfect with constant α. Then there exists $M = M(K, \alpha)$ such that*

$$\omega(\Omega, B_\rho(Q, 2r), A) \leq M(K, \alpha)\, \omega(\Omega, B_\rho(Q, r), A).$$

PROOF. This follows from Lemma 2.4 if we apply the classical Harnack inequality to $G_\Omega(z) := G_\Omega(z, A)$ for passing from $G_\Omega(A_Q^\ell(4r))$ to $G_\Omega(A_Q^\ell(2r))$, $\ell = 1, \dots, k(Q)$. We just need to apply Lemma 1.3 to finish the estimate.

Recall that in the introduction we showed that it was essential to replace euclidean balls by B_ρ. The second essential thing was to consider uniformly John property (equivalent to John localizability). The examples given in the introduction show that for general John domains the doubling property in ρ-metric (and in euclidian metric as well) fails to be true.

Theorem 3.2. *Let Ω be a uniformly John domain with constant K and let $\partial\Omega$ be uniformly perfect with constant α. Let $Q \in \partial\Omega$, $R > 0$, and let u, v be two positive harmonic functions in Ω vanishing continuously on $B(Q, 4R) \cap \partial\Omega$. Then there exist $M = M(K, \alpha)$ and $\varepsilon = \varepsilon(K, \alpha) > 0$ such that*

$$\Big|\frac{u(\xi)/v(\xi)}{u(\eta)/v(\eta)} - 1\Big| \leq M\Big(\frac{\rho(\xi, \eta)}{R}\Big)^\varepsilon \max_{x, y \in B(Q, 3R)} \Big(\Big|\frac{u(x)/v(x)}{u(y)/v(y)}\Big| + 1\Big),$$

for all $\xi, \eta \in B(Q, R)$.

PROOF. We are going to use Proposition 2. from the introduction and Lemma 2.3.

Let $Q_1 \in \partial\Omega$ be a closest point to ξ. Then $Q_1 \in B(Q, 2R)$. Let $r' = |Q_1 - \xi|$. If $\rho(\xi, \eta) \leq r'/2$, put $r = 2r'$; if $\rho(\xi, \eta) > r'/2$, put $r = 4\,\rho(\xi, \eta)$. Consider the local John domain $\Omega_1 := \Omega_{Q_1}^\ell(r)$ containing

ξ, η; $\ell \in \overline{1, k(Q_1)}$. There exists $M = M(K)$ (this is exactly M from (3) in the introduction) such that $\operatorname{diam} \Omega_1 \leq Mr$, that is $\Omega_1 \subset B_\rho(Q_1, Mr)$. Choose a local John domain $\Omega_2 := \Omega^\ell_{Q_1}(M^2 r)$ containing Ω_1. Again, $\operatorname{diam} \Omega_2 \leq M^3 r$, that is $\Omega_2 \subset B_\rho(Q_1, M^3 r)$. Choose $\Omega_3 := \Omega^\ell_{Q_1}(M^4 r)$ containing Ω_2, etc. We stop when we choose Ω_n such that

$$n = \max\{k : \; M^{2k-1} r \leq R\},$$

that is

$$n \geq \frac{1}{2} \frac{\log(R/r)}{\log M}.$$

Now we are left to apply Lemma 2.3 $n-1$ times to pairs of nested John domains $\Omega_n \supset \Omega_{n-1} \supset \cdots \supset \Omega_2 \supset \Omega_1$. For example, considering the pair $\Omega_k \supset \Omega_{k-1}$ we apply the lemma with $M^{2k-2} r \leq D \leq M^{2k-1} r$, $d = M^{2k-3} r$, $c_1 = 1/M$. If we denote by ℓ_k the quantity

$$\max_{\xi, \eta \in \Omega_k} \left| \frac{u(\xi)/u(\eta)}{v(\xi)/v(\eta)} - 1 \right|,$$

we obtain

$$\ell_{k-2} \leq q\,(d/D, K, \alpha, c_1)\, \ell_k = q(M, \alpha)\, \ell_k\,, qquadq = q(M, \alpha) \in (0,1)\,.$$

Finally

$$\ell_1 \leq q^{n/2-1}\, \ell_n\,,$$

which results in the estimate

$$\max_{\xi, \eta \in \Omega_1} \left| \frac{u(\xi)/v(\xi)}{u(\eta)/v(\eta)} - 1 \right| \leq \left(\frac{r}{R}\right)^\varepsilon \max_{x, y \in B(Q, 3R)} \left(\left| \frac{u(x) v(x)}{u(y)/v(y)} \right| + 1 \right).$$

If r was chosen to be $4\,\rho(\xi, \eta)$ this is what we want. Otherwise $\rho(\xi, \eta) \leq r/4 = |Q_1 - \xi|/2 = \operatorname{dist}(\xi, \partial\Omega)/2$ and we apply the usual Harnack's inequality in the disc $B(\xi, 3|Q_1 - \xi|/4)$ lying entirely in Ω_1. Then

$$\left| \frac{u(\xi)/v(\xi)}{u(\eta)/v(\eta)} - 1 \right| \leq C\, \frac{\rho(\xi, \eta)}{r} \max_{B(\xi, 3|Q_1 - \xi|/4)} \left| \frac{u(x)/v(x)}{u(y)/v(y)} - 1 \right|,$$

where C is absolute.

Combining two inequalities above we finish the proof in this case too.

Corollary 3.3. *Let Ω be a domain containing ∞. Let $\mathcal{U}$ be an open set containing $\mathbb{C}\setminus\Omega$ and u, v be two continuous functions in $\mathcal{U}$; positive, harmonic in $\mathcal{U}\cap\Omega$ and vanishing on $\mathbb{C}\setminus\Omega$. Suppose also that Ω is uniformly John with constant K and $\partial\Omega$ is uniformly perfect with constant α. Then if C is a compact subset of $\mathcal{U}$ we have for all $\xi,\eta\in C$ that*

$$\left|\frac{u(\xi)/v(\xi)}{u(\eta)/v(\eta)} - 1\right| \le M(C,K,\alpha)\,(\rho(\xi,\eta))^{\varepsilon}, \tag{3.1}$$

where $\varepsilon = \varepsilon(C,K,\alpha) > 0$.

4. Generalized polynomial-like maps. Geometric coding tree. Shift model for harmonic measure.

Throughout this section (f, V, U) is a generalized polynomial-like system (see the introduction), $J = J_f$. Let us remind an important notion from [PS]. One constructs the geometric coding tree as follows. Let $z_0 \in U \setminus J$. Let $z^1, \ldots, z^d$ be its preimages. Let γ^i be curves joining z_0 to z^j, $j = 1, \ldots, d$, such that

$$\overline{\mathrm{orb}(c)} \cap \bigcup_{j=1}^{d} \gamma^j = \varnothing$$

for any $c \in \mathrm{Crit}\,(f)$.

Let $\Sigma = \{1, \ldots, d\}^{\mathbb{N}}$ be the one sided shift space with σ denoting the shift to the left, and ρ_Σ be the standard metric on Σ:

$$\rho_\Sigma(\alpha,\beta) = e^{-k(\alpha,\beta)},$$

where $k(\alpha,\beta)$ is the least integer n for which $\alpha_n \neq \beta_n$. For each sequence α we put $\gamma^1(\alpha) = \gamma^{\alpha_1}$. Suppose that for every m, $1 \le m \le n$ and all $\alpha \in \sum$ the curves $\gamma^m(\alpha) : [0,1] \to U$ are already defined in such a way that $f(\gamma^m(\alpha)) = \gamma^{m-1}(\sigma(\alpha))$ and $\gamma^m(\alpha)(0) = \gamma^{m-1}(\alpha)(1)$. Define $\gamma^{n+1}(\alpha)$ by taking respective preimages of $\gamma^n(\sigma(\alpha))$. Put $z_n(\alpha) = \gamma^n(\alpha)(1)$. The graph $T = T(z_0, \gamma^1, \ldots, \gamma^d)$ with vertices z_0, $z_n(\alpha)$ and edges $\gamma^n(\alpha)$ is called a *geometric coding tree with root at* z_0. Given $\alpha \in \Sigma$, the subgraph composed by $z_0, z_n(\alpha), \gamma^n(\alpha)$ is called *α-branch* and is denoted by $b(\alpha)$. The branch $b(\alpha)$ is called *strongly convergent* if $\{\gamma^n(\alpha)\}$ converges to a point as $n \to \infty$.

It is easily seen from [CJY, Theorem 2.1] or from [BV1], [BV2] that the following simple proposition holds

Lemma 4.1. *Let f be semihyperbolic, then there exist $C_1 < \infty$, $0 < \theta_1 < 1$, such that*

$$\operatorname{diam} \gamma^n(\alpha) \le C_1\, \theta_1^n\,, \qquad \alpha \in \Sigma\,. \tag{4.1}$$

In particular each $b(\alpha)$ strongly converges to a point of J. (For a general f strong convergence of $b(\alpha)$ holds for all α except a set of zero Hausdorff dimension in (Σ, ρ_Σ).) Let $\pi(\alpha) = \lim z_n(\alpha)$ be a point of convergence of $b(\alpha)$. Let $r_n(\alpha) := |z_n(\alpha) - \pi(\alpha)|$.

Lemma 4.2. *Let f be semihyperbolic. Then*

$$\#\{z_n(\beta):\ \beta \in \Sigma\,,\ |z_n(\beta) - \pi(\alpha)| \le k\, r_n(\alpha)\} \le C(k, f)\,, \tag{4.2}$$

for any $\alpha \in \Sigma$ and any n.

PROOF. Put $x = \pi(\alpha)$. Choose M_0 in such a way that

$$\frac{C_1\, \theta_1^{M_0}}{1 - \theta_1} < \frac{\varepsilon}{2}\,, \tag{4.3}$$

where ε is from [CJY, Theorem 2.1] (Theorem B of [BV1], [BV2]).

Denote $\beta = \sigma^{n-M_0}(\alpha)$. Then $\pi(\beta) = f^{n-M_0}(x)$. Let $b_k(\alpha) = \cup_{i \ge k} \gamma^i(\alpha)$. Using (4.3) we have that $\operatorname{diam} b_{M_0}(\beta) \le C_1 \theta_1^{M_0}/(1-\theta_1) \le \varepsilon/2$. Therefore $b_{M_0}(\beta) \subseteq U_0 := B(f^{n-M_0}(x), \varepsilon/2)$. Denote by W_{n-M_0} the component of $f^{-(n-M_0)}(B(f^{n-M_0}(x), \varepsilon/2))$ which contains x. We conclude that $b_n(\alpha) \subseteq W_{n-M_0}$, thus $z_n(\alpha) \in W_{n-M_0}$, and so

$$\operatorname{diam} W_{n-M_0} \ge r_n(\alpha)\,.$$

Also $f^{n-M_0} : W_{n-M_0} \longrightarrow U_0$ is a branched covering of degree at most D (see [CJY, Theorem 2.1]) and the same is true with $2U_0$ if we replace W_{n-M_0} by a corresponding component of $f^{-(n-M_0)}(2U_0)$. Applying Lemma C from [BV1], [BV2] we see that W_{n-M_0} is τ-thick at x, that is $W_{n-M_0} \supset B(x, \tau\, r_n(\alpha))$.

Our purpose now is to enlarge W_{n-M_0} to the size of $B(x, k\, r_n(\alpha))$. To do this consider $U_1 = B(f^{n-M_0-M}(x), \varepsilon/2)$ for a certain M to be chosen later. Let U_2 be a component of $f^{-M}(U_0)$ containing $f^{n-M_0-M}(x)$. Denote by W_{n-M_0-M} the component of $f^{-(n-M_0-M)}(U_1)$ which contains x. We have the covering $f^{n-M_0-M} : W_{n-M_0-M} \longrightarrow U_1$ which sends W_{n-M_0} to U_2. This is the covering of degree at most D (see [CJY, Theorem 2.1]) and the same is true for $2U_1$ if W_{n-M_0-M} is replaced by a corresponding component of $f^{-(n-M_0-M)}(2U_1)$. Applying Lemma C from [BV1], [BV2] or using [HR] we see that with constants independent of n we have

$$\frac{\operatorname{diam} W_{n-M_0-M}}{\operatorname{diam} W_{n-M_0}} \sim \frac{\operatorname{diam} U_1}{\operatorname{diam} U_2} .$$

But from [CJY, Theorem 2.1] it follows that we can make $\operatorname{diam} U_2$ as small as we wish by choosing M large. Since W_{n-M_0-M} is τ-thick at x given k we can choose $M = M(k, f)$ so large that

$$B(x, k\, r_n(\alpha)) \subset W_{n-M_0-M} \,.$$

The degree of the map $f^{n-M_0-M} : W_{n-M_0-M} \longrightarrow U_1$ is bounded by D independent of n. So

$$\begin{aligned}\#\{z_n(\beta) : z_n(\beta) \in B(x, k\, r_n(\alpha))\} \\ &\le D\, \#\{z_{M+M_0}(\beta) : \ z_{M+M_0}(\beta) \in U_1\} \\ &\le D\, d^{M+M_0} = C(k, f)\,.\end{aligned}$$

Now we are in a position to prove

Theorem 4.3. *Let f be semihyperbolic. Then $\pi : \Sigma \longrightarrow J$ has the following properties:*

1) *π is Hölder continuous,*

2) *π is onto,*

3) *$\#\{\pi^{-1}(x)\} \le K(f)$ for any $x \in J$.*

PROOF. Hölder continuity is obvious from Lemma 4.1. Also 3) follows immediately from Lemma 4.2. "Onto" part is also easy.

One just applies the criterion of accessibility obtained by Przytycki in [P]. Or one can proceed as follows. Given $Q \in J$ let $z_{n(k)}(\alpha_k) \to Q$. We may assume that $\alpha_k \to \alpha$ in Σ. By Lemma 4.1 $|z_{n(k)}(\alpha_k) - \pi(\alpha_k)| \le C_1\, \theta_1^{n(k)}$. Thus $|Q - \pi(\alpha)| \le |Q - z_{n(k)}(\alpha_k)| + |z_{n(k)}(\alpha_k) - \pi(\alpha_k)| + |\pi(\alpha_k) - \pi(\alpha)| \to 0$, when $k \to \infty$. So $Q = \pi(\alpha)$ and the proof is completed.

REMARK. In fact, more is true. If we assume that f is separated semihyperbolic we can claim that the branches $b(\alpha)$ approximate the endpoint $\pi(\alpha) = Q$ by exhausting all the prime ends with impression Q.

PROOF. To see this we are going to prove that for any local domain $\Omega^i_Q(r)$ there is a point $z_n(\alpha) \in \Omega^i_Q(r)$. Because of the qc conjugacy it is enough to prove this for polynomials.

Suppose the contrary: there exists a local domain, say $\Omega^i_Q(r)$ free from preimages $z_n(\alpha)$. Consider the function u, which is harmonic on $\bar{\mathbb{C}} \setminus K_f$ and such that $u(z) \to 0$ if $z \to K_f$, $z \in \bar{\mathbb{C}} \setminus \Omega^i_Q(r)$ and $u(z) \to 1$ if $z \to K_f$, $z \in \Omega^i_Q(r)$. We define a sequence of functions $(u_n)_n$ by the formula:

$$u_n(z) = \sum_{y \in f^{-n} z} \frac{1}{d^n} u(y),$$

where d is the degree of the polynomial f.

It is easy to see that u_n are harmonic in $\bar{\mathbb{C}} \setminus K_f$. Furthermore if $z \neq \infty$, $u_n(z) \to 0$. On the other hand, since ∞ is superattracting fixed point we have $u_n(\infty) = u(\infty) > 0$.

This observation leads to a contradiction proving the assertion.

Lemma 4.4. *If f is semihyperbolic then*

$$\frac{\operatorname{dist}(z_n(\alpha), J)}{r_n(\alpha)} \sim 1,$$

where constants depend only on f.

This is another standard application of [CJY, Theorem 2.1] and Lemma C of [BV1], [BV2] (see also [HR]).

Let us now consider (f, V, U), $\Omega = \bar{\mathbb{C}} \setminus K_f$. Our main assumption is that f is separated semihyperbolic (see the introduction for definitions).

Then Theorem C claims that Ω is uniformly John. We are going to apply Corollary 3.3 to

$$u = G \circ f, \qquad v = G,$$

where G is Green's function of Ω with pole at ∞. Immediately we obtain

Theorem 4.5. *In the setting above for each $\alpha \in \Sigma$ there exists the limit*

$$\varphi(\alpha) = \lim_{n\to\infty} -\log \frac{(G\circ f)(z_n(\alpha))}{G(z_n(\alpha))} \tag{4.4}$$

and moreover φ is a Hölder continuous function on (Σ, ρ_Σ).

Let h_ν denote the entropy of the invariant measure ν on Σ (see *e.g.* [Wa]). Applying Sinai-Bowen-Ruelle thermodynamical formalism we construct the measure μ as in

Theorem D. *Let φ be a Hölder continuous function on (Σ, σ). Then there exists the limit (independent of $\beta \in \Sigma$)*

$$P = P(\varphi, \sigma) = \lim \frac{1}{n} \log\Big(\sum_{\alpha:\sigma^n\alpha=\beta} E_n(\alpha)\Big), \tag{4.5}$$

where $E_n(\alpha) = \exp(\varphi(\alpha) + \varphi(\sigma\alpha) + \cdots + \varphi(\sigma^{n-1}\alpha))$. Furthermore, there exists an invariant, ergodic probability measure μ on Σ such that for any $\beta \in \Sigma$

$$C_1\, e^{-nP} \le \frac{\mu\{\alpha:\ \alpha_1 = \beta_1, \dots, \alpha_n = \beta_n\}}{\exp(\varphi(\beta) + \cdots + \varphi(\sigma^{n-1}\beta))} \le C_2\, e^{-nP}. \tag{4.6}$$

The measure μ is unique with this property and it is equilibrium measure for φ which means that μ maximizes the functional $h_\nu + \int \varphi\, d\nu$. Moreover,we have:

$$h_\mu + \int \varphi\, d\mu = P.$$

Definition. *Measure μ is called Gibbs measure with potential φ.*

Theorem E. *If φ, ψ are two Hölder continuous function on $\sum$ then their Gibbs measures are either singular or the homologous equation*

$$\gamma\circ\sigma - \gamma = \varphi - \psi - P_\varphi + P_\psi$$

has a solution γ among Hölder continuous functions.

See [Bo] for Theorems D, E. In what follows φ is the function constructed in Theorem 4.5.

Lemma 4.6. $P(\varphi, \sigma) \leq 0$.

PROOF. Boundary Harnack principle of Corollary 3.3 implies more than the existence of the limit in (4.5). Actually

$$\left|\varphi(\alpha) - \log \frac{G(z_n(\alpha))}{G(f(z_n(\alpha)))}\right| \leq C q^n, \tag{4.8}$$

uniformly in α and n. In particular

$$\frac{\sum_{\sigma^n \alpha = \beta} E_n(\alpha)}{\sum_{\sigma^n \alpha = \beta} G(z_n(\alpha))} \asymp 1.$$

Now we apply Lemma 1.3 (even with B_ρ replaced by B):

$$G(z_n(\alpha)) \leq M(K, \alpha)\, \omega(B(z_n(\alpha), 4\, r_n(\alpha))).$$

Denote $B_{n,\alpha} = B(z_n(\alpha), 4\, r_n(\alpha))$. Discs $\{B_{n,\alpha}\}_{\alpha \in \sum}$ form the covering of J. Lemma 4.2 implies readily that this covering has a finite multiplicity independent of n.

Combining these facts we get

$$\sum E_n(\alpha) \leq C(K, \alpha)$$

and the lemma follows.

REMARK. One can prove that $P(\varphi, \sigma) = 0$, see [B]. But this requires more careful estimates. Moreover it is proved in [B] that $\pi^* \mu$ is mutually absolutely continuous with respect to harmonic measure ω.

Definition. *We call $\pi^* \mu$ the shift model measure for ω. It is clearly f-invariant.*

5. $\omega \approx m \Leftrightarrow$ conformal maximality.

We recall the definition from [BPV]. The system (f, V, U) is called *conformally maximal* if it is conformally equivalent to (g, V_g, U_g) for which harmonic measure ω_g of $\mathbb{C} \setminus K_g$ equals measure of maximal entropy of g on J_g. Measure of maximal entropy will be denoted by m (or m_g, m_f, m_σ to highlight the dynamical system if we need this).

Lemma 5.1. $\pi^* m_\sigma = m_f$.

PROOF. This is clear from [LW] and Theorem 4.3.

From Lemma 4.2 it is easy to deduce the following natural assertion.

Lemma 5.2. *Let F be a Borel subset of Σ such that $m_\sigma(F) = 0$. Then $m_\sigma(\pi^{-1}\pi F) = 0$.*

Now we are going to consider two cases:

First case: $\mu \perp m_\sigma$. Then Theorem D and Lemma 4.6 give

$$\log d + \int_\Sigma \varphi \, dm_\sigma < P(\varphi) \leq 0 .$$

Second case: $\mu = m_\sigma$. Then Theorem E gives

$$\varphi + \log d = \gamma \circ \sigma - \gamma + P(\varphi) .$$

Anyway, either we have (for a certain positive ε)

$$\log d + \int_\Sigma \varphi \, dm_\sigma = -2\varepsilon < 0 \tag{5.1}$$

or

$$\varphi + \log d = \gamma \circ \sigma - \gamma, \qquad \gamma \in \text{Höld}(\Sigma) . \tag{5.2}$$

The last possibility occurs only if $\mu = m_\sigma$ and $P(\varphi) = 0$.

Lemma 5.3. *If* (5.1) *happens then* $\omega \perp m$.

PROOF. By Birkhoff's ergodic theorem:

$$\frac{1}{n}(\varphi(\alpha) + \cdots + \varphi(\sigma^{n-1}\alpha)) \to \int_\Sigma \varphi \, dm_\sigma ,$$

for almost every α with respect to m_σ. Combining this with (5.1) we see that

$$\frac{1}{n}(\varphi(\alpha) + \cdots + \varphi(\sigma^{n-1}\alpha)) \leq -\varepsilon + \log \frac{1}{d} , \qquad n \geq n(\alpha) ,$$

for almost every α with respect to m_σ. Combining this with (4.7) we get the estimate of Green's function

$$G(z_n(\alpha)) \le C\, e^{-\varepsilon n}\, d^{-n}, \qquad n \ge n(\alpha), \tag{5.3}$$

for almost every α with respect to m_σ. Let E be the set of α for which (5.3) holds.

By Lemma 5.2 we may assume that

$$E = \pi^{-1} e, \qquad e \subset J, \quad m(J \setminus e) = 0. \tag{5.4}$$

We are going to show that

$$\omega(e) = 0. \tag{5.5}$$

Clearly (5.4), (5.5) finishes the proof. To prove (5.5) we need Lemma 2.4. Put $n(Q) = \max_{\alpha \in \pi^{-1}(Q)} n(\alpha)$, where $n(\alpha)$ is taken from (5.3). Let N be fixed and put

$$r(Q) = \frac{1}{100} \min_{\alpha \in \pi^{-1}(Q)} |z_{\max(N, n(Q))}(\alpha) - Q|.$$

"Discs" $\{\bar{B}_\rho(Q, r(Q))\}_{Q \in e}$ cover e and let B_ρ^i, $B_\rho^i := \bar{B}_\rho(Q_i, r(Q_i))$, be a disjoint family such that $e \subset \bigcup_{i \ge 1} 5\, B_\rho^i$. This family exists by Vitali's lemma. Lemma 2.4 claims that

$$\omega(5\, B_\rho^i) \le M_1 \sum_\ell G(A_{Q_i}^\ell (10\, r(Q_i))). \tag{5.6}$$

Remind that $b_k(\alpha) = \cup_{i \ge k} \gamma^j(\alpha)$. Let m_i be the first index greater than $\max(N, n(Q_i))$ such that $\alpha \in \pi^{-1}(Q_i)$ implies

$$z_{m_i}(\alpha) \quad \text{and} \quad b_{m_i}(\alpha) \quad \text{lie inside} \quad \bigcup_\ell \Omega_{Q_i}^\ell \left(\frac{r(Q_i)}{M_2} \right). \tag{5.7}$$

By the Remark after Theorem 4.3 we see that in each local domain there is at least one point $z_{m_i}(\alpha)$.

We can apply now Harnack's inequality:

$$\sum_\ell G(A_{Q_i}^\ell (10\, r(Q_i))) \le M_3 \sum_{\alpha \in \pi^{-1}(Q_i)} G(z_{m_i}(\alpha)), \tag{5.8}$$

where M_3 depends only on M_2 and John constant of John arcs $b(\alpha)$ (see Lemma 4.4 which claims that $b(\alpha)$ are John arcs with uniform John constant). Now if M is the constant from (3) of the introduction and $M_2 \geq M$ then

$$\Omega^{\ell}_{Q_i}\Big(\frac{r(Q_i)}{M_2}\Big) \subset B_\rho(Q_i, r(Q_i)) = B^i_\rho \,.$$

In particular from (5.7) we see that

$$\text{(5.9)} \quad \alpha \in \pi^{-1}(Q_i) \quad \text{implies} \quad z_{m_i}(\alpha) \quad \text{and} \quad b_{m_i}(\alpha) \ \text{lie inside} \ B^i_\rho \,.$$

Combine (5.8), (5.6) and (5.3) to write

$$\text{(5.10)} \qquad \begin{aligned} \sum_{i\geq 1} \omega(5B^i_\rho) &\leq C\, M_1\, M_3\, e^{-\varepsilon N} \sum_{i\geq 1} \#\pi^{-1}(Q_i)\, d^{-m_i} \\ &\leq M_4\, e^{-\varepsilon N} \sum_{i \geq 1} d^{-m_i} \,. \end{aligned}$$

On the other hand we have that if $Q_i \neq Q_j$ then $z_{m_i}(\alpha) \notin b_{m_j}(\beta)$ for any pair α, β, $\alpha \in \pi^{-1}(Q_i)$, $\beta \in \pi^{-1}(Q_j)$. This is just (5.9) combined with the fact that B^i_ρ and B^j_ρ are disjoint. If we denote $C_i := \{x \in \Sigma : x_k = \alpha_k$, for some $\alpha \in \pi^{-1}(Q_i)$, and all k, $1 \leq k \leq m_i\}$ we conclude that

$$\text{(5.11)} \qquad Q_i \neq Q_j \quad \text{implies} \quad C_i \cap C_j = \varnothing.$$

On the other hand $m_\sigma(C_i) \geq d^{-m_i}$. This, (5.10) and (5.11) now imply

$$\omega(e) \leq M_4\, e^{-\varepsilon N}\, m_\sigma(\cup C_i) \leq M_4\, e^{-\varepsilon N} \,.$$

We are done because N is arbitrary.

In what follows an important part is played by the so called automorphic harmonic function. Given a generalized polynomial-like system (f, V, U) (see the introduction) we say that the function τ on U is an automorphic harmonic function (for f) if

1) τ is nonnegative subharmonic function on U,

2) τ is positive and harmonic in $U \setminus K_f$,

3) $\tau|_{K_f} \equiv 0$,

and

$$(Aut) \qquad \tau(fz) = d\,\tau(z).$$

Lemma 5.4. *Assume that* (5.2) *holds. Then there exists an automorphic harmonic function for* f.

PROOF. To avoid unimportant technicalities we are going to present the proof only in the case of totally disconnected J. This gives the advantage of only one local domain for given Q, r. The general case follows along the same lines. The procedure in the "boundedly many local domains" case differs only technically from the "only one local domain" case.

Applying Corollary 3.3 to $u = G \circ f$, $v = G$ and taking into account (5.2) we write

$$\left| \log \frac{G(fz)}{G(z)} - \log d - (\gamma(\sigma\alpha) - \gamma(\alpha)) \right| \leq C\,\rho(z, \pi\alpha)^{\varepsilon}. \tag{5.12}$$

Let $\alpha_i = i, \dots, i, \dots, i = 1, \dots, d$, and let $q_i = \pi(\alpha_i)$.

Let $B = B_{q_i}$ to be such a small disc centered at q_i that all components W_n of $f^{-n}(B)$ containing q_i are free from critical points of f.

Thus a univalent branch $g = g_{q_i}$ of f^{-1} is defined in B and $g^n : B \to W_n$ is univalent. Putting $\alpha = \alpha_i$ into (5.12) we define a positive finite limit

$$U_{q_i}(z) = \lim_{n \to \infty} d^n\, G(g^n z), \qquad z \in B,$$

and $U_{q_i}(z) \leq C\,G(z)$, $z \in B$. Let $V_{q_i} = U_{q_i}\, e^{\gamma(q_i)}$. Thus V_{q_i} is positive harmonic on $B \backslash K_f$, nonnegative subharmonic on B and by construction

$$\lim_{z \to \pi\alpha} \frac{G(z)}{U_{q_i}(z)} = e^{\gamma(\alpha) - \gamma(q_i)},$$

that is

$$\lim_{z \to \pi\alpha} \frac{G(z)}{V_{q_i}(z)} = e^{\gamma(\alpha)}. \tag{5.13}$$

Now let I be the set of periodic points of Σ, $q_\chi = \pi(\chi)$, $\chi \in I$. We can repeat the construction above and obtain $U_\chi, V_\chi = U_\chi\, e^{\gamma(\chi)}$

defined in B_{q_χ}. Let $\mathrm{L} = \cup_{n\geq 0}\sigma^{-n}I$. Fix $l \in \mathrm{L}$ and define V_l in a small neighborhood of B_l of $\pi(l)$ by the relation

$$V_l(z) = d^{-n}V_\chi(f^n z)$$

if $\chi = \sigma^n l$. We clearly can use (5.12) to extend (5.13) on V_l literally. Let us try to prove that

$$V_{l_1} = V_{l_2} \ \text{ on } \ B_{l_1} \cap B_{l_2}\,. \tag{5.14}$$

Actually we are going to prove that either (5.14) holds, or a certain symmetrized version (see (5.16) below) holds.

We use (5.13) for V_{l_1}, V_{l_2} to conclude that

$$|V_{l_1}(z) - V_{l_2}(z)| = o(G(z))\,, \qquad z \in B_{l_1} \cap B_{l_2}\,. \tag{5.15}$$

Let us remind

Lemma F (A.F. Grishin [Gr]). *Let w_1, w_2 be two nonnegative subharmonic functions on an open set $\mathcal{O}$ and $w_1 \geq w_2$. Let*

$$J = \left\{z \in \mathcal{O}:\ (w_1 - w_2)(z) = \liminf_{r\to 0} \int_0^{2\pi} (w_1 - w_2)(z + re^{i\theta})\, d\theta = 0\right\}.$$

Then

$$\Delta w_1 \geq \Delta w_2 \ \text{ on } \ J\,.$$

Lemma F and (5.15) imply that

$$\Delta V_{l_1} = \Delta V_{l_2} \ \text{ on } \ B_{l_1} \cap B_{l_2}\,.$$

Thus the difference $V_{l_1} - V_{l_2}$ is harmonic in $B_{l_1} \cap B_{l_2}$. Remind that it vanishes on $K_f \cap B_{l_1} \cap B_{l_2}$.

We conclude that either (5.14) holds or $K_f \cap B_{l_1} \cap B_{l_2}$ is covered by finitely many real analytic curves. In the latter case $K_f = J_f$ is covered by finitely many real analytic curves. As in [LyV] we conclude that these curves are disjoint. Now let $*$ be a holomorphic symmetry with respect to these curves.

Consider

$$\tilde{V}_l(z) := \frac{V_l(z) + V_l(z^*)}{2}\,.$$

Now $\tilde{V}_{l_1} - \tilde{V}_{l_2}$ vanishes on these curves together with its normal derivative. That means $\tilde{V}_{l_1} = \tilde{V}_{l_2}$ in $B_{l_1} \cap B_{l_2}$. In this case we have a family of functions $\tilde{V}_l$ and small neighborhoods B_l of l, $l \in \mathrm{L}$ with properties

$$\tilde{V}_{l_1} = \tilde{V}_{l_2} \text{ on } B_{l_1} \cap B_{l_2}, \qquad \tilde{V}_{\sigma l}(fz) = d\tilde{V}_l(z). \tag{5.16}$$

Let $\tilde{V}_l$ denote V_l if (5.14) holds and denote $\tilde{V}_l$ if (5.16) holds. The occurance of these alternatives depends on whether J is contained in a finite union of analytic curves or not.

Put $\tau|_{B_l} = \tilde{V}_l|_{B_l}$. So we choose a neighborhood $\mathcal{O}$ of L, such that $f^{-1}(\mathcal{O}) \subset \mathcal{O}$ and a subharmonic nonnegative function τ on $\mathcal{O}$ such that

$$\tau(z) = d\,\tau(f^{-1}z), \qquad z \in \mathcal{O}. \tag{5.17}$$

The function τ is positive and harmonic in $\mathcal{O} \setminus K_f$. It may happen that $\mathcal{O}$ does not contain the whole J_f inside. So we are going to extend τ as follows: Let ε be a number such that diameters of components of $f^{-n}(B(x,\varepsilon))$ are at most $C\,\theta^n$ (see [CJY, Theorem 2.1]). We are going to extend τ to $B(q_\chi, \varepsilon)$, $\chi \in I$ ($\equiv$ periodic points). Let N be so large that the component W_N of $f^{-N}(B(q_\chi,\varepsilon))$ which contains q_χ is contained in $\mathcal{O}$. Thus $f^N : W_N \to B(q_\chi,\varepsilon)$ is a branched covering. Let $z \in B(q_\chi,\varepsilon)$ be a critical value of this covering and let Γ be a curve not meeting the critical values of f^N and lying in $B(q_\chi,\varepsilon)$. We choose Γ to connect z with a certain $\xi \in W_N$.

Let x, y be any two f^N-preimages of z lying in W_N. Let γ_x, γ_y be two liftings of Γ by f^N into W_N, starting at x, y respectively. In a small neighborhood U of Γ we can define τ^x and τ^y by

$$\tau^x(t) = d^N\,\tau(f_x^{-N}t), \qquad \tau^y(t) = d^N\,\tau(f_y^{-N}t).$$

Here f_x^{-N}, f_y^{-N} are branches of f^{-N} on U mapping U into neighborhoods of γ_x, γ_y respectively. Note that τ^x, τ^y coincide near ξ because of (5.17). So as harmonic function they should coincide in the whole U.

Clearly we have for this extension

$$\lim_{\substack{z \to \pi\alpha \\ z \in B(q_\chi,\varepsilon)}} \frac{G(z)}{\tau(z)} = e^{\gamma(\alpha)}.$$

As before we can see that on $\Omega = \cup_{\chi \in I} B(q_\chi,\varepsilon)$ a subharmonic function τ is defined such that

$$\tau(fz) = d\,\tau(z), \qquad \text{if } z,\ f(z) \in \Omega.$$

The advantage of Ω in comparison to $\mathcal{O}$ is that $\Omega \supset J_f$. Let m be so large that $f^{-m}(U) \setminus K_f \subset \Omega$. Define

$$\tau(z) = d^m \tau(f^{-m} z), \qquad z \in U \setminus K_f . \tag{5.18}$$

Repeating the lifting argument above we see that (5.18) gives a single valued function. Clearly τ is an automorphic harmonic function we were looking for and Lemma 5.4 is proved.

Now let us use the following

Theorem G ([BPV]). *Let (f, V, U) be a generalized polynomial-like system. Then it is conformally maximal if and only if there exists a harmonic automorphic function for f.*

Uniting Lemmas 5.3, 5.4 with Theorem G we obtain the following criterion of *conformal maximality.*

Theorem 5.5. *Let (f, V, U) be a system with separated semihyperbolic f. Then it is conformally maximal if and only if harmonic measure on J_f is not singular with respect to measure of maximal entropy on J_f.*

Aknowledgements: We would like to thank Juha Heinonen for many interesting conversations. We also thank to a referee for many helpful suggestions in improving the exposition of this paper.

References.

[A1] Ancona, A., Principe de Harnack à la frontière et théorème de Fatou pour un opérateur elliptique dans un domaine lipschitzien. *Ann. Inst. Fourier* **28** (1978), 169-213.

[A2] Ancona, A., *Theorie du potential pour les graphes et les varietes.* Lecture Notes in Math. **1427**, 1990.

[B] Balogh, Z., Invariant harmonic measure for separated semihyperbolic dynamics. Preprint, 1994.

[BB] Bass, R. F., Burdzy, K., A boundary Hanack principle in twisted Hölder domains. *Ann. of Math.* **34** (1991), 253-276.

[BV1] Balogh, Z., Volberg, A., Geometric localization in John domains. Preprint, Uppsala Univ., 1994.

[BV2] Balogh, Z., Volberg, A., Geometric localization, uniformly John property and separated semihyperbolic repellers. To appear in *Arkiv Mat.*

[BV3] Balogh, Z., Volberg, A., Principe de Harnack à la frontière pour les repulseur holomorphes non-recurrents. *C. R. Acad. Sci. Paris* **319** (1994), 351-355.

[BPV] Balogh, Z., Popovici, I., Volberg, A., Conformally maximal polynomial-like dynamics and invariant harmonic measure. To apear in *Ergodic Theory & Dynamical Systems.*

[Bo] Bowen, R., *Equilibrium states and the ergodic theory of Anosov diffeomorphisms.* Lecture Notes in Math. **470**, 1975.

[CJY] Carleson, L., Jones, P. W., Yoccoz, C.-J., Julia and John. *Bol. Soc. Bras. Mat.* **25** (1194), 1-30.

[C] Carleson, L., On the existence of boundary values of harmonic functions in several variables. *Arkiv Mat.* **4** (1962), 393-399.

[E] Eremenko, A., Julia sets are uniformly perfect, Preprint, Purdue Univ., 1992.

[GHM] Gehring, F., Hag, K., Martio, O., Quasihyperbolic geodesic in John domains. *Math. Scand.* **65** (1989), 75-92.

[Gr] Grishin, A. F., On sets of regular growth of entire functions. *Teor. Funct. Func. Anal. i ich prilozh* **40** (1983), (in Russian).

[HR] Heinonen, J., Rohde, S., Koenigs functions, quasicircles and BMO. Preprint, Univ. of Michigan, 1994.

[H] Hinkkanen, A., Julia sets of rational functions are uniformly perfect. *Math. Proc. Cambridge Phil. Soc.* **113** (1993), 543-559.

[JK] Jerison, D., Kenig, C., Boundary behavior of harmonic functions in non-tangentially accessible domains. *Adv. in Math.* **46** (1982), 80-147.

[J] Jones, P., A geometric localization theorem. *Adv. in Math.* **46** (1982), 71-79.

[LW] Ledrappier, F., Walters, P., A relativised variational principle for continuous transformations. *J. London Math. Soc.* **2** (1977), 568-576.

[Lo] Lopes, A., Equilibrium measure for rational functions. *Ergodic Theory & Dynamical Systems* **6** (1986), 393-399.

[LyV] Lyubich, M., Volberg, A., A comparison of harmonic and balanced measures on Cantor repellers. To appear in *Asterisque.*

[Ma] Mañé, R., On a lemma of Fatou, *Bol. Soc. Bras. Mat.* **24** (1993), 1-12.

[MR] Mañé, R., da Rocha, L. F., Julia sets are uniformly perfect. *Proc. Amer. Math. Soc.* **116** (1992), 251-257.

[NV] Näkki, R., Väisälä, J., John disks. *Exposit. Math.* **9** (1991), 3-43.

[Po] Pommerenke, Ch., *Boundary behaviour of conformal maps.* Graduate Texts in Math. ???, Springer, 1992.

[P] Przytycki, F., Accessibility of typical points for inariant measure of positive Lyapunov exponent for iterations of holomorphic maps. *Inst. Math. Sci., SUNY Stony Brook* **3** (1993), 1-19.

[PS] Przytycki, F., Skrzypczak, J., Convergence and pre-images of limit points for coding trees for iterations of holomorphic maps. *Math. Ann.* **90** (1991), 425-440.

[Wa] Walters, P., *An Introduction to Ergodic Theory.* Graduate Texts in Math. **79**, Springer, 1982.

[W] Wu, J.-M., Comparison of kernel functions, boundary Harnack principle and relative Fatou theorem on Lipschitz domains. *Ann. Inst. Fourier* **28** (1978), 147-167.

Recibido: 28 de octubre de 1.994
Revisado: 20 de octubre de 1.995

Z. Balogh
Mathematisches Institute
Universitaet Bern
Sidlestrasse 5, 3012 Bern
SWITZERLAND
zoltan@math-stat.unibe.ch

A. Volberg*
Department of Mathematics
Michigan State University
East Lansing, MI 48824
U.S.A.
volberg@math.msu.edu

* Partially supported by NSF grant DMS 9302728

Revista Matemática Iberoamericana
Vol. 12, N.° 2, 1996

On the nonexistence of bilipschitz parameterizations and geometric problems about A_∞-weights

Stephen Semmes

Abstract. How can one recognize when a metric space is bilipschitz equivalent to a Euclidean space? One should not take the abstraction of metric spaces too seriously here; subsets of $\mathbb{R}^n$ are already quite interesting. It is easy to generate geometric conditions which are necessary for bilipschitz equivalence, but it is not clear that such conditions should ever be sufficient. The main point of this paper is that the optimistic conjectures about the existence of bilipschitz parameterizations are wrong. In other words, there are spaces whose geometry is very similar to but still distinct from Euclidean geometry. Related questions of bilipschitz equivalence and embeddings are addressed for metric spaces obtained by deforming the Euclidean metric on $\mathbb{R}^n$ using an A_∞ weight.

1. Introduction.

How can one recognize when a metric space is bilipschitz equivalent to a Euclidean space? Recall that a map $f : M_1 \to M_2$ between two metric spaces M_1, M_2 is *bilipschitz* if there is a constant K such that $K^{-1}\, d_1(x,y) \le d_2(f(x), f(y)) \le K\, d_1(x,y)$ for all $x, y \in M_1$, where d_1

and d_2 are the metrics on M_1 and M_2. We shall sometimes say K-*bilipschitz* to make the constant explicit. Two metric spaces are said to be bilipschitz equivalent if there is a bilipschitz mapping from one onto the other. If two metrics spaces are bilipschitz equivalent then they should have approximately the same behavior in terms of lengths, Hausdorff measure, topology, etc., and the question is whether some nice combination of such conditions can detect bilipschitz equivalence with Euclidean spaces.

Let $(M, d(\cdot,\cdot))$ be a metric space, and suppose that it is bilipschitz equivalent to $\mathbb{R}^d$ with the Euclidean metric. What are some of the conditions that M must satisfy? For each pair of points $x, y \in M$ there must be a curve that joins them whose length is at most $Cd(x,y)$ for some constant C (which depends only on M). There must be a measure μ on M with the property that the μ-mass of a ball of radius $r > 0$ is approximately r^d, *i.e.*, the μ-mass must be bounded from above and below by constants times r^d. In fact this must be true with μ equal to d-dimensional Hausdorff measure on M. There must be another constant C such that any metric ball B in M of radius r is contained in a d-dimensional topological ball U which is itself contained in a metric ball of radius $C\,r$, and one could impose further restrictions on U. Suitable formulations of the Sobolev and Poincaré inequalities on M must also hold.

On the other hand, plenty of the familiar properties of $\mathbb{R}^d$ do not have to be satisfied, even approximately, by a bilipschitz-equivalent metric space M. Bilipschitz mappings need not be smooth, or even C^1, and so a bilipschitz-equivalent space could have a lot of corners. For instance, if $A : \mathbb{R}^d \to \mathbb{R}$ is any Lipschitz function (so that $|A(x) - A(y)| \le C\,|x - y|$ for some C and all $x, y \in \mathbb{R}^d$), then the graph of A in $\mathbb{R}^{d+1}$ equipped with the ambient Euclidean metric is bilipschitz equivalent to $\mathbb{R}^d$.

The main purpose of this paper is to provide examples which show that there is no hope for finding simple general conditions of the type just described which ensure the existence of a bilipschitz parameterization. The examples below will all be subsets of some Euclidean space (with the inherited metric) or "conformal" deformations of $\mathbb{R}^d$. Let us begin with the former, starting with a definition.

Definition 1.1. *A subset E of $\mathbb{R}^n$ is said to be (Ahlfors) regular of dimension d if it is closed and if there is a constant $C_0 > 0$ such that*

$$C_0^{-1}\, r^d \le H^d(E \cap B(x,r)) \le C_0\, r^d\,, \tag{1.2}$$

for all $x \in E$ *and* $r > 0$. *Here (and forevermore)* H^d *denotes d-dimensional Hausdorff measure and* $B(x, r)$ *denotes the open ball with center* x *and radius* r.

This condition is equivalent to the apparently more general version in which one merely asks that there exist a measure μ supported on E which satisfies (1.2). In other words, if such a measure exists, it has to be comparable in size to the restriction of d-dimensional Hausdorff measure to E.

Roughly speaking, a set is regular if it behaves measure-theoretically like $\mathbb{R}^d$, even though it may be very different geometrically. There are examples of regular sets which are self-similar Cantor sets, or snowflake curves, or tree-like objects. Regular sets can have non-integer dimension. Of course any set which is bilipschitz equivalent to $\mathbb{R}^d$ is regular.

Theorem 1.3. *There is a 3-dimensional regular set* E *in* $\mathbb{R}^4$ *which is the image of a hyperplane under a global quasiconformal mapping from* $\mathbb{R}^4$ *onto itself but which is not bilipschitz equivalent to* $\mathbb{R}^3$. *This quasiconformal mapping can also be taken to be Lipschitz continuous. The set* E *enjoys the additional property that there is a constant* $L_0 > 0$ *so that every pair of distinct points* $x, y \in E$ *is contained in a closed subset* W *of* E *which is* L_0*-bilipschitz equivalent to a closed Euclidean 3-ball. In particular,* x *and* y *can be connected by a curve in* E *of length at most* $L_0^2\,|x - y|$.

A *quasiconformal mapping* is one which does not distort *relative* distances by more than a bounded factor. There are many equivalent characterizations, one of which is that $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ is quasiconformal if there is a constant C so that for each $x \in \mathbb{R}^n$ and $r > 0$ there is an $s > 0$ so that $B(\Phi(x), s) \subseteq \Phi(B(x, r)) \subseteq B(\Phi(x), C\,s)$. In other words, Φ maps a ball to a set which is trapped between two balls of comparable radii. It is important that the image radius s is allowed to be very different from r -otherwise this condition would reduce to bilipschitzness- but that the constant C not be allowed to depend on x or r. Recall that Φ is *Lipschitz continuous* if there is a constant C so that $|\Phi(x) - \Phi(y)| \leq C\,|x - y|$ for all x, y.

Although the image of a 3-plane under a global quasiconformal mapping (or under a quasisymmetric mapping) can be fractal, an image like this which is also regular of dimension 3 automatically has many of the properties that it would have if it were bilipschitz equivalent to $\mathbb{R}^3$. Such a set must be uniformly rectifiable (in the sense of [DS4]), every pair of points in the set can be connected by a path which is not too long compared to the distance between them, and suitable versions of Sobolev and Poincaré inequalities must hold on the set. This uses a method of Gehring [Ge] which allows one to control well the extent to which the quasiconformal mapping distorts distances in this situation. (See also [DS1] and [Se4].) Theorem 1.3 resolves (negatively) the obvious question of whether a quasiconformal image of a plane which is regular of the same dimension must actually be bilipschitz equivalent to a plane.

The set E in Theorem 1.3 is far from being bizarre or pathological. It will be smooth off a self-similar Cantor set, around which it does some spiralling. It can be approximated nicely by smooth submanifolds, and it agrees with a hyperplane outside of a compact set. The obstruction to the existence of a bilipschitz parameterization will be present locally in the sense that even small neighborhoods of the singular points will not admit bilipschitz parameterizations. By adjusting the parameters we can make this set E especially nice, so that the inverse of the quasiconformal parameterization is Hölder continuous of order as close to 1 as we like, while the mapping itself remains Lipschitz. Analogous statements for Sobolev spaces are also true. On the other hand, by choosing the parameters differently we can build E in such a way that there is no homeomorphism from E onto $\mathbb{R}^3$ which is locally Hölder continuous of any order $\alpha > 0$ given in advance. See Theorem 5.27 below.

The construction of these examples will be based on "Antoine's Necklaces" [Mo] and the following lemma.

Lemma 1.4. *If U is an open set in $\mathbb{R}^d$ and K is a closed subset of $\mathbb{R}^d$, and if K has Hausdorff dimension less than $d-2$, then any loop in $U \setminus K$ can be contracted in $U \setminus K$ if it can be contracted in U. In particular, $\mathbb{R}^d \setminus K$ is simply connected.*

In other words, a set is invisible in terms of the properties of π_1 of its complement if it is thin enough. This is given in [MRV, Lemma 3.3, p. 9]. Actually, that result dealt only with the simple-connectivity of the relevant domains, but the same proof applies to this formulation. See also [LV] and [SS, p. 506].

This lemma provides us with a necessary condition for a set to be bilipschitz equivalent to $\mathbb{R}^d$, since any such metric space would have to enjoy a similar property, and it is this necessary condition that the examples in Theorem 1.3 will violate. That is, E will be constructed in such a way that it contains a compact set with Hausdorff dimension less than 1 whose complement is not simply connected.

Let us consider briefly the conceptual ramifications of the existence of sets like E in Theorem 1.3. One could say that the above list of geometric necessary conditions for bilipschitz equivalence with a Euclidean space is simply too small, and that it should be enlarged to include the property in Lemma 1.4 (and others). I am inclined toward a different view. I am not optimistic that there are nice geometric characterizations of sets which are bilipschitz equivalent to a Euclidean space (although I would not be surprised if there were nice results in more narrowly focussed situations, in which π_1 conditions as in Lemma 1.4 would appear naturally). I think that the existence of sets like E in Theorem 1.3 means that one should view bilipschitz parameterizations as luxuries which are desirable but not reasonable to expect in general and also not crucial. In other words, these examples are sufficiently nice that they should be accommodated rather than excluded. Instead of looking for more stringent criteria to ensure the existence of a bilipschitz parameterization one can look for simpler conditions which imply a lot of good structure (of the type that these examples enjoy) if not an actual bilipschitz parameterization. In this connection the notion of uniform rectifiability (as in [DS4]) is very natural, because it incorporates many (but by no means all) of the nice features of bilipschitz equivalence with a Euclidean space while being much more flexible and easier to detect. See also [DS2], [DS3], [DS5], and [Se5].

Let us now consider analogous issues for some general "conformal" deformations of Euclidean geometry.

Definition 1.5. *A continuous weight ω on $\mathbb{R}^d$ is a nonnegative continuous function whose zero set has Lebesgue measure zero. If A is a measurable subset of $\mathbb{R}^d$ then $\omega(A)$ will be used to denote $\int_A \omega$. A continuous weight ω on $\mathbb{R}^d$ is doubling (or satisfies a doubling condition) if there is a constant C so that $\omega(2B) \leq C\,\omega(B)$ for all balls B in $\mathbb{R}^d$, where $2B$ denotes the ball with the same center as B but twice the radius. We shall view ω as defining a measure on $\mathbb{R}^d$, and also a conformal deformation of Euclidean geometry. To this end we associate to ω the (possibly degenerate) distance function $D_\omega(x,y)$, which is the infimum of the ω-length of all rectifiable paths in $\mathbb{R}^d$ which join x to y. (The ω-length of a path γ is defined to be $\int_\gamma \omega^{1/d}\,ds$, where ds denotes arclength measure.) We say that ω is a strong A_∞ continuous weight if it is doubling and if there is a $C > 0$ so that $C^{-1}\,\omega(B_{x,y})^{1/d} \leq D_\omega(x,y) \leq C\,\omega(B_{x,y})^{1/d}$*

for all $x, y \in \mathbb{R}^d$, where $B_{x,y}$ is the smallest closed Euclidean ball which contains x and y.

Constant functions provide trivial examples of strong A_∞ continuous weights. Less trivial examples are given by $\omega(x) = |x|^a$, $a \geq 0$. On the other hand the strong-A_∞ condition prevents ω from vanishing on a nontrivial line segment, or on any rectifiable curve for that matter, since $D_\omega(x, y)$ would then vanish for some $x \neq y$. However, strong A_∞ weights can vanish on Cantor sets of large Hausdorff dimension, as in Proposition 4.4.

The "strong-A_∞" condition was originally defined in [DS1]. The strange-looking name was motivated by an observation that is recalled below. The main point is that the strong-A_∞ condition by itself implies that the metric D_ω has many nice properties, without any smoothness assumptions on ω or anything like that. To understand this condition better it is helpful to consider $\delta_\omega(x, y) = \omega(B_{x,y})^{1/d}$ as some kind of distance function in its own right. Specifically, it is a quasimetric, which means that it satisfies all the conditions normally required of a metric except that the triangle inequality should be weakened to $\delta_\omega(x, z) \leq C\,(\delta_\omega(x, y) + \delta_\omega(y, z))$ for some C and all x, y, z. This condition is easy to verify using the doubling property of ω. Moreover, δ_ω is "quasiconformally" equivalent to the Euclidean metric, in the sense that its balls are approximately the same as Euclidean balls. More precisely, if $x \in \mathbb{R}^d$ and $r > 0$ are given, and if $R > 0$ is chosen so that $\omega(B(x, R)) = R^d$, then

$$B(x, C^{-1}R) \subseteq \{y \in \mathbb{R}^d : \ \delta_\omega(x, y) < r\} \subseteq B(x, C\,R). \tag{1.6}$$

Here C depends on the doubling constant of ω but not on x, r or R. On the other hand, R and r can be wildly different from each other, and either can be larger than the other. This fact (1.6) is easy to derive from the definition of δ_ω and the doubling condition on ω, and one can verify also that the ω-diameters of $B(x, C^{-1}R)$ and $B(x, CR)$ are both approximately equal to r, *i.e.*, they are bounded from above by $C'r$ and from below by $C'^{-1}r$ for some constant C'. The strong-A_∞ condition says that the geodesic distance D_ω is comparable in size to δ_ω, so that D_ω has these features too.

If ω is a strong A_∞ continuous weight, then the metric space $(\mathbb{R}^d, D_\omega)$ enjoys many of the same properties as ordinary Euclidean space. For instance, if β is a D_ω-ball, then there is a Euclidean (and hence topological) ball B containing β whose D_ω-diameter is bounded by a constant times that of β. This can be derived from (1.6). Also, there is a constant $C > 0$ so that

$$C^{-1} r^d \le \omega(\{y \in \mathbb{R}^d : D_\omega(x,y) < r\}) \le C r^d , \tag{1.7}$$

because of (1.6) and the doubling condition. This is analogous to the regularity condition 1.2.

It was observed in [DS1] that the strong-A_∞ condition implies the (much older) A_∞ condition that there exist constants $p > 1$ and $C > 0$ such that

$$\left(\frac{1}{|B|}\int_B \omega^p\right)^{1/p} \le C \frac{1}{|B|}\int_B \omega , \tag{1.8}$$

for all balls B in $\mathbb{R}^d$, where $|B|$ denotes the Lebesgue measure of B. In other words, the strong-A_∞ condition implies that $\omega \in L^p_{\mathrm{loc}}$ with $p > 1$, and with a uniform and scale-invariant bound. This is basically a reformulation of a result of Gehring [Ge], and there is also a bound on averages of (small) negative powers of ω. (See [Ga] and [Jr] for more information about A_∞ weights.) These bounds on ω imply that D_ω-geometry is closer to Euclidean geometry than one might think. For instance, they imply the "uniform rectifiability" condition that every D_ω-ball in $\mathbb{R}^d$ has a definite proportion (with respect to the ω-volume) which is uniformly bilipschitz equivalent to a subset of $\mathbb{R}^d$ with the Euclidean metric. Alternatively one can use (1.8) and its cousins to obtain Sobolev space estimates on the identity mapping as a map from $(\mathbb{R}^d, |x - y|)$ to $(\mathbb{R}^d, D_\omega)$ and vice-versa. (Hölder continuity can be derived directly from the doubling condition, as in Proposition 4.22.)

The main result of [DS1] states that the analogues of the usual Sobolev and Poincaré inequalities on Euclidean spaces are also true for $(\mathbb{R}^d, D_\omega)$ when ω is a strong A_∞ weight. In view of all these common features between D_ω and the Euclidean metric when ω satisfies the strong-A_∞ condition it is natural to ask the following.

QUESTION 1.9. ([DS1]) *If ω is a strong A_∞ continuous weight on $\mathbb{R}^d$, then is $\mathbb{R}^d$ equipped with the metric D_ω bilipschitz equivalent to $\mathbb{R}^d$ equipped with the Euclidean metric?*

This is equivalent to asking whether there is a quasiconformal mapping $h : \mathbb{R}^d \to \mathbb{R}^d$ whose jacobian J satisfies $C^{-1}\omega \le J \le C\omega$ for some constant C. It is not hard to show that the doubling and strong-A_∞

conditions on ω are necessary in order for such a quasiconformal mapping to exist. Continuity of ω is not necessary, though, but it is also not required in the "true" definition of a strong A_∞ weight in [DS1]. The restriction to continuous weights here allows us to avoid some technical complications which are not necessary for the purposes of this paper, but discontinuous weights are indispensable for the complete story. The general case of discontinuous weights is also discussed in [Se4].

The answer to Question 1.9 is no.

Theorem 1.10. *There is a strong A_∞ continuous weight ω on $\mathbb{R}^3$ for such that $(\mathbb{R}^3, D_\omega)$ is not bilipschitz equivalent to $\mathbb{R}^3$ equipped with the Euclidean metric. We can also take ω to be bounded. Moreover, there is a constant L_1 so that for every pair of distinct points $x, y \in \mathbb{R}^3$ there is a closed subset W of $\mathbb{R}^3$ which contains x and y and such that (W, D_ω) is L_1-bilipschitz equivalent to a closed Euclidean 3-ball equipped with the standard Euclidean metric.*

The counterexamples for Theorem 1.10 will be based on the violation of the property in Lemma 1.4, and they can be taken to be powers of the distance to a set in $\mathbb{R}^3$ like an Antoine's necklace. The resulting spaces $(\mathbb{R}^3, D_\omega)$ will then be embedded into $\mathbb{R}^4$ to produce the counterexamples for Theorem 1.3. In fact Theorem 1.3 has to include Theorem 1.10 (modulo continuity of the weight ω), because one can show that any set E as in Theorem 1.3 must be bilipschitz equivalent to $(\mathbb{R}^3, D_\omega)$ for some (not necessarily continuous) strong A_∞ weight ω. In our case this fact will come from the construction. (See Lemma 5.25.)

An amusing consequence of Theorem 1.10 is that the Sobolev-Poincaré inequalities obtained in [DS1] are nontrivial in the sense that they cannot be reduced to the Euclidean case by a change of variables. (However, for these particular examples there are much more direct arguments than the ones in [DS1] for the general case.)

By adjusting the parameters of the construction one can obtain various improvements of Theorem 1.10, just as for Theorem 1.3. See Theorem 4.20 and the remarks that follow it.

Theorem 1.10 raises the same kind of conceptual issues as Theorem 1.3 does. If one really wants to classify the weights that arise as jacobians of quasiconformal mappings (or even give nice criteria for this to happen), then one has to strengthen the strong-A_∞ condition.

Alternatively, one can take the view that a Euclidean space with its geometry deformed by a strong A_∞ weight as above is a rather nice object, even if it is not quite as nice as the standard Euclidean space, and that it should be accommodated rather than excluded. Again, I am inclined toward this "big tent" philosophy.

Let us consider now the following weaker version of Question 1.9.

QUESTION 1.11. *If ω is a strong A_∞ continuous weight on $\mathbb{R}^d$, then is $(\mathbb{R}^d, D_\omega)$ bilipschitz equivalent to a subset of some $\mathbb{R}^n$ equipped with the Euclidean metric?*

It is known that the answer to Question 1.11 is yes for many strong A_∞ weights. The precise statement is complicated, but the main result in [Se4] says that if a strong A_∞ weight ω has the property that it can be made smaller to a nontrivial extent and remain a strong A_∞ weight, then the answer to Question 1.11 is yes for ω. This criterion is not satisfied by all strong A_∞ weights, but it is satisfied by the important subclass of A_1 weights (see Definition 2.8), and any strong A_∞ weight can be approximated by weights which satisfy this stronger condition. Roughly speaking, the strong A_∞ weights which do not satisfy this stronger condition are sitting at the boundary of the space of strong A_∞ weights.

Nonetheless, the answer to Question 1.11 is no.

Theorem 1.12. *There is a strong A_∞ weight on some $\mathbb{R}^d$ such that $(\mathbb{R}^d, D_\omega)$ is not bilipschitz equivalent to any subset of any $\mathbb{R}^n$ (equipped with the Euclidean metric).*

Of course the statement of Theorem 1.12 is much stronger than that of Theorem 1.10, except for the knowledge of the dimension d. However, the example used to prove Theorem 1.10 will have the property that one can embed $(\mathbb{R}^d, D_\omega)$ bilipschitzly in $\mathbb{R}^4$, so that we actually know that there are counterexamples to Question 1.9 which are not pathological for Question 1.11. In fact the examples for Theorem 1.10 satisfy the stronger version of the strong-A_∞ condition in [Se4] mentioned above, and they are nicer, simpler, and more flexible than the examples for Theorem 1.12. The examples for Theorem 1.12 are not based on Lemma 1.4 or Antoine's necklaces or anything like that, and unfortunately they are not very explicit.

Strong A_∞ weights are related to an abstract version of Question

1.11 in an interesting way. To explain this we need another definition.

Definition 1.13. *A metric space $(M, d(\cdot,\cdot))$ satisfies a doubling condition if there is a constant C so that any ball in M can be covered by at most C balls of half the radius.*

This condition shows up in [A1], [A2], [A3], [CW1], [CW2], and [Gr], but with different names. It provides a kind of boundedness on the geometry of M. It is satisfied by Euclidean spaces, and hence their subspaces, and it is not hard to show that $(\mathbb{R}^d, D_\omega)$ satisfies this doubling condition when ω is a strong A_∞ continuous weight, because of the doubling condition on ω in Definition 1.5.

The abstract version of Question 1.11 asks whether any metric space which satisfies a doubling condition must be bilipschitz equivalent to a subset of some $\mathbb{R}^n$. The answer is known to be no, and this point will be discussed more in Section 7. There is very nice positive result due to Assouad ([A3, Proposition 2.6, p. 436], see also [A1], [A2]), however.

Theorem 1.14 (Assouad). *If $(M, d(\cdot,\cdot))$ is a metric space which satisfies a doubling condition, then for each $\alpha \in (0,1)$ the metric space $(M, d(\cdot,\cdot)^\alpha)$ is bilipschitz equivalent to a subset of some $\mathbb{R}^n$.*

Thus, given a metric space which satisfies a doubling condition, one can perturb the metric in order to get a space which embeds into a Euclidean space bilipschitzly. Although this perturbation is small in some ways, it does have the unfortunate feature that it enlarges the class of Lipschitz functions on the space enormously, in such a way as to destroy any sort of differentiability almost everywhere theorem as one has on Euclidean space. In fact the counterexamples to the $\alpha = 1$ case of Theorem 1.14 show that such destruction is necessary. (See Section 7.)

Note that the doubling condition is necessary for the embedding in Theorem 1.14 to exist (for any $\alpha > 0$). Also, Theorem 1.14 implies that every metric space which satisfies a doubling condition admits a quasisymmetric embedding into some Euclidean space. (Basically a quasisymmetric embedding is one that distorts relative distances by a bounded amount, like quasiconformal mappings on $\mathbb{R}^d$, but we shall not need the precise definition here.)

There is a converse to the fact that the deformations of Euclidean spaces associated to strong A_∞ continuous weights satisfy the doubling condition in Definition 1.13.

Theorem 1.15. *If $(M, d(\cdot,\cdot))$ is a metric space which satisfies a doubling condition, then there is a positive integer n and a strong A_∞ continuous weight ω on $\mathbb{R}^n$ such that $(M, d(\cdot,\cdot))$ is bilipschitz equivalent to a subset of $(\mathbb{R}^n, D_\omega(\cdot,\cdot))$.*

Thus Question 1.11 is equivalent to its abstract version for metric spaces which satisfy doubling conditions, and so the negative answer to the abstract version (which was known to Assouad) implies the negative answer to Question 1.11 itself (Theorem 1.12).

The next section provides more information about related results and open problems but is not essential to the rest of the paper. Antoine's necklaces are reviewed in Section 3, and they are used to prove Theorems 1.10 and 1.3 in Sections 4 and 5, respectively. Theorem 1.15 is proved in Section 6, and Theorem 1.12 is derived from it in Section 7. Section 8 is devoted to regular mappings, which are more flexible cousins of bilipschitz mappings.

2. Some related results and problems.

One of the most interesting examples of a set which is not bilipschitz equivalent to the expected standard model is provided by the following.

Theorem 2.1 (Edwards). *There is a finite polyhedron P (in a Euclidean space of modest dimension) which is homeomorphic to the 5-sphere $\mathbf{S}^5$ and which contains a polygonal arc Γ such that $P \setminus \Gamma$ is not simply connected. In fact, for no open set $U \subseteq P$ which intersects Γ is it true that $U \setminus \Gamma$ is simply connected.*

Corollary 2.2. *If P and Γ are as above, then P is not bilipschitz equivalent to $\mathbf{S}^5$, and no open subset U of P which intersects Γ is bilipschitz equivalent to an open subset of $\mathbb{R}^5$.*

The corollary follows from Edwards' theorem and Lemma 1.4 (as was explained in [SS,, Remark (b), p. 504]). Lemma 1.4 implies that a closed set in $\mathbf{S}^5$ with Hausdorff dimension less than 3 must always have simply-connected complement, so that any homeomorphism from P onto $\mathbf{S}^5$ takes Γ to a set with Hausdorff dimension at least 3. Thus no such homeomorphism can be Lipschitz or even Hölder continuous of order larger than 1/3. Similar arguments apply to open sets U as in the corollary.

Edwards' example is of the following form. Let H be a compact smooth 3-dimension manifold which is not simply connected but which does have the same (integer) homology as $\mathbf{S}^3$. Many examples of such manifolds are known to exist, but note that if H were also simply connected then the (as yet unproven) Poincaré conjecture would imply that H is diffeomorphic to $\mathbf{S}^3$. Also, recall that a homology sphere of dimension 1 or 2 has to be a standard sphere. Now take P to be the "join" of H with a copy of $\mathbf{S}^1$. That is, P contains a copy of H and of $\mathbf{S}^1$, and for each point in H and each point in $\mathbf{S}^1$, P contains a line segment which joins the two points, these line segments are disjoint except possibly for necessary intersections at their endpoints in H and $\mathbf{S}^1$, and P does not contain any other points. We can do this in such a way that P is a polyhedron, by applying this to polyhedral copies of H and $\mathbf{S}^1$ instead of smooth copies. The curve Γ mentioned above is simply the copy of $\mathbf{S}^1$ inside P that we are using. It is easy to see that $P \setminus \Gamma$ is homotopy-equivalent to the polyhedral copy of H inside P, and hence is not simply connected. A little more thought gives the local version of this lack of simple-connectivity mentioned above.

The deep part of Edwards' theorem is that H can be chosen so that P is homeomorphic to $\mathbf{S}^5$. To understand this better we need to recall the notion of a suspension of a space. The *suspension* of H is just like the join of H with $\mathbf{S}^1$, except that we use $\mathbf{S}^0$ instead. That is, we take two points off of H, and then we build a new space by joining each of these points with a line segment to each point in H. For example, if we take the suspension of a sphere, then we get a sphere of one higher dimension. One can check that P is the same as the suspension of the suspension (the double suspension) of H. Thus if we used $\mathbf{S}^3$ instead of H to build P, then it would be immediately clear that P would be homeomorphic to $\mathbf{S}^5$, even piecewise-linearly. In the case of a non-simply connected homology sphere H this is less clear. The suspension of H is not a manifold, because the two cones points are not manifold points. The idea is that it is almost a manifold, and in fact almost a sphere, so that the second suspension makes it into a topological manifold. This is not at all obvious. Edwards proved that this worked for some homology spheres, and Cannon then proved that all homology spheres work. Edwards then proved a very general theorem for recognizing when a topological space is a topological manifold (which is the key point here). See [C1], [C2], [Da], and [E].

Let us think a little about the geometric properties of the space P in Theorem 2.1. It is a finite polyhedron, *i.e.*, a finite union of (5-dimensional) simplices, it is homeomorphic to the 5-sphere, but it is not bilipschitz equivalent to the 5-sphere, let alone piecewise-linearly equivalent to $\mathbf{S}^5$. Nonetheless P has many of the same geometric properties as the 5-sphere. It is a 5-dimensional regular set in the sense of Definition 1.1, modulo the necessary restriction to $r < 1$ (say) in (1.2), since P is compact. It is easy to see that the usual Sobolev and Poincaré inequalities hold on P, and that other aspects of analysis work as well. It is not hard to prove that P also enjoys the following property: there is a $C > 0$ so that if $x \in P$ and $0 < r < C^{-1}$, then there is an open subset V of P such that $P \cap B(x,r) \subseteq V \subseteq P \cap B(x, Cr)$ and V is homeomorphic to a 5-ball. (This is true for any finite polyhedron which is a topological manifold -see [Se6, Section 11]- but in the case of P it is a little easier to see using its special form.) One can strengthen this statement to include uniform scale-invariant bounds on the moduli of continuity of the homeomorphisms between these sets V and the unit ball in $\mathbb{R}^5$ and their inverses. (Roughly speaking, there are really only two different V's that one has to worry about, modulo some simple operations like dilating.)

Thus P has many of the same nice metric properties as a smooth manifold, even though it is not bilipschitz equivalent to one. It is in very much the same spirit as the examples mentioned in the introduction that will be constructed in the next sections, but it is much more impressive, since it is even a finite polyhedron. On the other hand, the examples constructed below have the advantages that they are easier to verify, they work in lower dimensions (3 instead of 5), and their parameterizability properties are better controlled (in terms of the existence or nonexistence of Hölder continuous coordinates, for instance; see Theorems 4.20 and 5.27).

Let us now consider more fully the question of the dimensions in which these various types of examples exist.

PROBLEM 2.3. *Are there analogues of Theorems* 1.3 *and* 1.10 *in dimension* 2*? For which dimensions* d *does Theorem* 1.12 *hold?*

The Edwards' examples work in all dimensions greater or equal than 5, but the construction does not make sense in lower dimensions (because homology spheres are true spheres in dimensions 1 and 2). The examples given here for Theorems 1.3 and 1.10 could be adapted to higher dimensions, but the method breaks down completely in dimension 2. Of course many things are better in two dimensions, and there could even be positive results in that case. (For the record, these questions all degenerate into triviality when $d = 1$, because one can use arclength parameterizations to build the required mappings.)

As for Theorem 1.12, the proof will give a value of d which is computable in principle but whose smallness is not clear. See Section 7, and see Section 8 for some related questions.

There are some positive results which are special to dimension 2 but which address a slightly different question. That is, there are some reasonable geometric criteria for a 2-dimensional metric space to admit a quasisymmetric parameterization (which need not be bilipschitz). See [Se2, Section 5], [DS3, Section 6], and [HK]. The reason that dimension 2 is special here is that one has the uniformization theorem which can provide a conformal parameterization right from the start. (This is analogous to the special role of the arclength parameterization in dimension 1.) One still has the problem of passing from the infinitesimal conformality condition to distortion estimates at large scales, but this can be managed, and in fact the results of Heinonen and Koskela [HK] deal effectively with this problem under very general circumstances. Nonetheless, the difficulty posed by the absence of the uniformization theorem in higher dimensions remains, and in fact there are examples [S6] which show that the analogues of the 2-dimensional results fail in higher dimensions.

There is another special case of these questions which is not addressed by the known examples.

Definition 2.4. *Let M be a hypersurface in $\mathbb{R}^{d+1}$, and assume a priori that M is smooth and nice at ∞. Given $\varepsilon \geq 0$, we say that M is ε-flat if*

$$D_M(x,y) \leq (1+\varepsilon)\,|x-y|\,, \qquad \textit{for all } x,y \in M\,, \tag{2.5}$$

where $D_M(x,y)$ denotes the geodesic distance on M (the infimum of the lengths of all paths on M which join x to y), and if

$$(1-\varepsilon)\,\nu_d\, r^d \leq H^d(M \cap B(x,r)) \leq (1+\varepsilon)\,\nu_d\, r^d\,, \tag{2.6}$$

for all $x \in M$ *and* $r > 0$*, where* ν_d *denotes the volume of the unit ball in* $\mathbb{R}^d$*.*

It is easy to see that M must be a hyperplane if it is ε-flat with $\varepsilon = 0$.

PROBLEM 2.7. *If* $M \subseteq \mathbb{R}^{d+1}$ *is* ε*-flat, and if* ε *is small enough, then must it be true that* M *is bilipschitz equivalent to* $\mathbb{R}^d$ *with a bilipschitz constant that depends only on* d*? Is it bilipschitz equivalent to* $\mathbb{R}^d$ *with a constant which tends to* 1 *uniformly as* $\varepsilon \to 0$*?*

If M admits a K-bilipschitz parameterization by $\mathbb{R}^d$, then (2.5) and (2.6) must hold with $\varepsilon \to 0$ as $K \to 1$, as one can easily check. The point of Problem 2.7 is to know whether the converse holds, with estimates which do not depend quantitatively on the a priori assumptions.

It turns out that the ε-flatness condition implies many other flatness conditions that are necessary for the existence of such a bilipschitz parameterization. For instance, it implies the existence of local coordinates with good estimates in terms of Hölder spaces and Sobolev spaces, as long as one stays away from the Lipschitz class (which is the question of Problem 2.7), and when $d = 2$ there is a positive result for quasisymmetric parameterizations, based on the uniformization theorem. (See [Se2].) Also, there are some sufficient conditions for the existence of bilipschitz parameterizations (with constant close to 1) in terms of conditions which are stronger than ε-flatness but which are roughly of the same order of magnitude [T2]. When $d = 2$ there is a nice sufficient condition in terms of the L^2 norm of the curvature being small [T1]. This condition is stronger than ε-flatness, and has approximately the same relationship with ε-flatness that the Sobolev space $W^{1,2}(\mathbb{R}^2)$ has with BMO($\mathbb{R}^2$) via the Sobolev embedding. When $d > 2$ it is not known whether the L^d norm of the principal curvatures being small is sufficient to ensure the existence of a bilipschitz parameterization. This curvature condition is the natural one, in that it scales correctly and implies ε-flatness.

There are also some nice equivalent characterizations of ε-flatness with small ε, in terms of the Gauss map having small BMO norm (and hence small oscillation in a certain sense) [Se3] and in terms of singular integral operators and Clifford analysis [Se1].

Despite all these good properties of ε-flat surfaces, I am pessimistic about Problem 2.7. However, I do not know any counterexamples, and the type of examples given in this paper will not work for this.

One could also consider small constant versions of Question 1.9. When $d = 2$ one can reduce Problem 2.7 to such a question using [Se2].

There is another variant of Question 1.9 which has some hope. For this it is much less reasonable to deal only with continuous weights, and so we use the correct general definition.

Definition 2.8. *A nonnegative measurable function ω on $\mathbb{R}^d$ which is positive almost everywhere is called an A_1 weight if there is a constant C such that*

$$\frac{1}{|B|}\int_B \omega(y)\,dy \le C \operatorname*{essinf}_{y\in B} \omega(y)\,, \tag{2.9}$$

for all balls B in $\mathbb{R}^d$.

The A_1 condition implies the strong-A_∞ condition, although the definition of $D_\omega(\cdot,\cdot)$ needs to be modified since we are not assuming continuity. (This issue is treated thoroughly in [Se4].) The A_1 condition is much stronger; it forbids any vanishing, while the strong-A_∞ condition forbids only certain kinds of vanishing. A simple example of an A_1 weight on $\mathbb{R}^d$ is $\omega(x) = |x|^{-s}$ for $0 \le s < d$. Similarly, an A_1 weight can blow up along a submanifold, but not as rapidly.

PROBLEM 2.10. *Does Question* 1.9 *have an affirmative answer for A_1 weights?*

Question 1.11 does have an affirmative answer in the case of A_1 weights [Se4]. The situation for Question 1.9 is unclear, but the method for producing an example as in Theorem 1.10 definitely does not work in the case of A_1 weights. (The whole point will be to make the weight vanish on a certain set, which an A_1 weight cannot do.)

See [Ga] and [Jr] for more information about A_1 weights.

There is a notion of "regular mappings" which is weaker (and more flexible) than bilipschitzness and for which there are some interesting results and problems related to Questions 1.9 and 1.11 and Problems 2.3 and 2.10. See Section 8.

For related questions and examples pertaining to strong-A_∞ weights, see [Se4], especially Sections 4 and 5.

3. Antoine's necklaces.

This section will be devoted to the necklaces of Antoine, which are Cantor sets in $\mathbb{R}^3$ whose complements are not simply connected. These sets will be used heavily in the next two sections, to construct the examples promised in the introduction. The basic reference for this section is [Mo, Chapter 18]. (See also [B] for higher-dimensional versions of Antoine's necklaces.)

Let k be a reasonably large positive integer. This is a parameter which is at our disposal; it needs to be reasonably large for the construction to work nicely ($k \geq 10^7$ is fine), but there is no reason for us to try to choose k as small as possible anyway. It will be important later for us to have the option of taking k to be arbitrarily large. The construction is a little nicer when k is even.

Fix a circle Γ_0 in $\mathbb{R}^3$ with radius 1. Let P_0 be a collection of k equally spaced points on Γ_0. For each $p \in P_0$ choose a circle $\gamma_0(p)$ in $\mathbb{R}^3$ centered at p in such a way that all the $\gamma_0(p)$'s have the same radius $\rho(k)$ and the following properties are satisfied:

$$k^{-1} \leq \rho(k) \leq 2\pi\, k^{-1}\,, \tag{3.1}$$

$$\operatorname{dist}(\{p\} \cup \gamma_0(p), \{q\} \cup \gamma_0(q)) \geq (100\, k)^{-1}\,, \quad \text{when } p \neq q\,, \tag{3.2}$$

(3.3) $\gamma_0(p)$ and $\gamma_0(q)$ are linked (as circles in $\mathbb{R}^3$) if and only if p and q are adjacent to each other (as elements of P_0).

It is not hard to check that these circles actually exist. (It is helpful to remember that by taking k large we have that Γ_0 is very flat at the scale of k^{-1}. If we did not want adjacent circles to link, but instead to touch at a single point, then we would want to take $\rho(k) \approx \pi\, k^{-1}$. As it is, we need to take them a little larger, but (3.1) gives us enough room. We could also impose additional symmetry requirements on the $\gamma_0(p)$'s, but we shall not bother.) The union of the $\gamma_0(p)$'s, $p \in P_0$, link together in a necklace near Γ_0, and we denote their union by $N(\Gamma_0)$. See [Mo, Figures 18.1 and 18.2, p. 127-8] for excellent pictures.

To build Antoine's necklace we shall replace each $\gamma_0(p)$ with a smaller copy of $N(\Gamma_0)$, and then replace each component of the resulting set with a smaller copy of $N(\Gamma_0)$, and so forth. To do this carefully we need to "mark" our circles.

Recall that a similarity on $\mathbb{R}^n$ is an affine mapping which is a combination of a translation, (nonzero) dilation, and orthogonal transformation.

Definition 3.4. *A marked circle in $\mathbb{R}^3$ is a circle Γ together with an orientation-preserving similarity ϕ on $\mathbb{R}^3$ (called the marking) such that $\phi(\Gamma_0) = \Gamma$. We shall often let Γ denote both a marked circle (with ϕ not mentioned explicitly) and the circle as a set.*

Of course we shall always take Γ_0 to be marked in the obvious way (with $\phi =$ the identity).

Choose now markings ϕ_p for the circles $\gamma_0(p)$'s, $p \in P_0$. The choice of the markings is insignificant, but they need to be fixed once and for all. With the selection of these markings we can view $N(\Gamma_0)$ as a union of marked circles.

If Γ is a marked circle, with marking ϕ, then we set

$$N(\Gamma) = \phi(N(\Gamma_0)). \tag{3.5}$$

We shall consider $N(\Gamma)$ to be a union of marked circles, with the markings induced by ϕ in the obvious way. These circles are also naturally labelled by P_0.

If Γ and Γ' are marked circles and if $\psi : \mathbb{R}^3 \to \mathbb{R}^3$ is an orientation-preserving similarity which takes Γ to Γ' in a way which is compatible with the markings, then

$$\psi(N(\Gamma)) = N(\Gamma'), \tag{3.6}$$

and the markings of the constituent circles of $N(\Gamma_1)$ and $N(\Gamma_2)$ correspond under ψ in the obvious way.

If $E \subseteq \mathbb{R}^3$ is a finite union of marked circles, then we take $N(E)$ to be the union of $N(\alpha)$, where α runs through the consitituent circles of E. Thus if $\mathcal{C}$ denotes the set of subsets of $\mathbb{R}^3$ which are finite unions of marked circles, then N defines a mapping from $\mathcal{C}$ to itself. Note that if $E, E' \in \mathcal{C}$, then $E \cup E' \in \mathcal{C}$ and $N(E \cup E') = N(E) \cup N(E')$. Also, orientation-preserving similarities act on $\mathcal{C}$ in the obvious way, and this action commutes with N, by (3.6).

Set $A_0 = \Gamma_0$, $A_1 = N(\Gamma_0)$, and define A_l in general by $A_l = N^l(\Gamma_0)$, where N^l denotes the l^{th} power of N, viewed as a mapping on $\mathcal{C}$. That is, we view Γ_0 and the A_l's as elements of $\mathcal{C}$, so that we can iterate N in this way. Each A_l is a union of k^l marked circles of radius

$\rho(k)^l$, and A_{l+1} is the union of the $N(\alpha)$'s, where α runs through the circles which make up A_l. The Hausdorff limit of the A_l's will give a necklace of Antoine. Before we deal with the Hausdorff limit we should record some simple preliminary facts.

Lemma 3.7. *For each j, l with $0 \le j \le l$ we have that $A_l = \cup_\alpha N^{l-j}(\alpha)$, where the union is taken over all the consitituent (marked) circles α in A_j, and the constituent circles of these two collections are marked in the same way.*

In other words, A_l and $\cup_\alpha N^{l-j}(\alpha)$ are the same as elements of $\mathcal{C}$. This follows easily from the definitions.

Lemma 3.8. *If $k \ge 10^7$, then every marked circle Γ in $\mathbb{R}^3$ satisfies*

$$\sup_{x \in N(\Gamma)} \operatorname{dist}(x, \Gamma) \le \rho(k)\, \text{radius}(\Gamma) < 10^{-6}\, \text{radius}(\Gamma)\,.$$

To prove this one reduces to the case where $\Gamma = \Gamma_0$ and uses (3.1) and the definitions.

From now on we assume that k is at least 10^7. Given a circle Γ in $\mathbb{R}^3$, let $\tau(\Gamma)$ be the small solid torus containing Γ given by

$$\tau(\Gamma) = \{x \in \mathbb{R}^3 : \operatorname{dist}(x, \Gamma) \le 10^{-5}\, \text{radius}\, \Gamma\}\,. \tag{3.9}$$

If Γ and Γ' are two circles and ψ is a similarity that maps one onto the other, then $\psi(\tau(\Gamma)) = \tau(\Gamma')$. If E is a finite union of circles in $\mathbb{R}^3$, then we define $\tau(E)$ to be the union of the sets obtained by applying τ to the constituent circles of E.

Lemma 3.10. *$\tau(N^l(\Gamma)) \subseteq \tau(\Gamma)$ for all marked cirles Γ and all $l \ge 0$.*

This follows from Lemma 3.8 and the definitions by estimating the relevant geometric series.

Lemma 3.11. *$\tau(A_j) \supseteq \tau(A_l)$ when $0 \le j \le l$.*

This is an immediate consequence of Lemmas 3.10 and 3.7.

We now define our Antoine's necklace A by

$$A = \bigcap \tau(A_l)\,. \tag{3.12}$$

This is a compact subset of $\mathbb{R}^3$. It is the same as the Hausdorff limit of the A_l's, but this definition is a little easier to use.

This set A is totally disconnected. To quantify this disconnectedness define $\tau_a(\Gamma)$ for $a > 0$ and Γ a circle in $\mathbb{R}^3$ to be the solid torus containing Γ given by

$$\tau_a(\Gamma) = \{x \in \mathbb{R}^3 : \ \mathrm{dist}\,(x, \Gamma) \leq a\, 10^{-5}\,\mathrm{radius}\,\Gamma\}\,. \tag{3.13}$$

Thus if $a \geq 1$ then this is somewhat larger than $\tau(\Gamma)$.

Lemma 3.14. *Let α and α' be two distinct circles among those which make up A_l, $l \geq 1$. Then $\tau_{50}(\alpha)$ and $\tau_{50}(\alpha')$ are disjoint.*

To see this notice first that

$$\mathrm{dist}\,(\gamma_0(p), \gamma_0(q)) > 700^{-1}\,\mathrm{radius}\,\gamma_0(p) \tag{3.15}$$

when $p, q \in P_0$, $p \neq q$, because of (3.2) and (3.1). This implies that $\tau_{50}(\gamma_0(p))$ and $\tau_{50}(\gamma_0(q))$ are disjoint when $p \neq q$, which is the same as the $l = 1$ case of Lemma 3.14. The case where $l > 1$ and α and α' have the same "parent" in A_{l-1} reduces to the $l = 1$ case by a similarity transformation. Suppose now that $l > 1$ and α and α' have different parents in A_{l-1}. Let δ be the parent of α, so that α is one of the circles in $N(\delta)$. Then

$$\tau_{50}(\alpha) \subseteq \tau_{50}(\delta) \tag{3.16}$$

because of Lemma 3.8. Thus the disjointness of $\tau_{50}(\alpha)$ and $\tau_{50}(\alpha')$ would follow if we could establish the disjointness of $\tau_{50}(\delta)$ and $\tau_{50}(\delta')$, where δ' is the parent of α'. By iterating this procedure we can reduce to the previous case where the two circles have the same parent. This proves the lemma.

For the record:

Theorem 3.17. *$\mathbb{R}^3 \setminus A$ is not simply connected.*

This is Theorem 4 of [Mo, p. 141].

Corollary 3.18. *No homeomorphism from $\mathbb{R}^3$ to itself can map A to a set with Hausdorff dimension less than* 1.

This follows from Theorem 3.17 and Lemma 1.4. It is very amusing, since A itself is homeomorphic to a standard Cantor set. Thus we cannot find a global homeomorphism on $\mathbb{R}^3$ which sends A to a standard Cantor set. See also [Mo, Theorem 5, p. 131].

Notice that when $k \to \infty$ A tends to the circle Γ_0 in the Hausdorff topology. The Hausdorff dimension A is always larger than 1, and it tends to 1 as $k \to \infty$.

There is also a "local" version of Corollary 3.18.

Corollary 3.19. *If U is an open set in $\mathbb{R}^3$ which intersects A, then any homeomorphism from U onto another open set in $\mathbb{R}^3$ must take $U \cap A$ to a set with Hausdorff dimension at least* 1.

Indeed, because of the obvious self-similarity of A, there must be a small copy $\widetilde{A}$ of A inside U. Theorem 3.17 implies that there is a small loop in $U \setminus \widetilde{A}$ which is contractible in U but not in $U \setminus \widetilde{A}$, and so Lemma 1.4 implies the corollary.

For more information about this kind of wildness phenomena see [R].

4. Strong-A_∞ weights.

This section will be devoted to Theorem 1.10 and some variants of it. We shall consider continuous weights of the form

$$\omega(x) = \min\left(1, \operatorname{dist}(x, A)^{3s}\right) \tag{4.1}$$

on $\mathbb{R}^3$, where A is a compact set and $s > 0$ is at our disposal. In fact we shall take A to be the set constructed in the preceding section, with the parameter k also at our disposal. We shall see that this weight satisfies the *strong-A_∞ condition* and has other interesting properties. The typographically simpler weights of the form $\operatorname{dist}(x, A)^{3s}$ would be practically as good, but it is nice to force ω to be constant off a compact set. There is nothing magical about the 3 in the exponent, it merely reflects the fact that we shall use ω as a density on $\mathbb{R}^3$, and it simplifies some of the later formulae.

Our first task is to show that ω is a strong A_∞ weight. This point is clearer in a more general setting.

Definition 4.2. *A closed set E in $\mathbb{R}^n$ is said to be uniformly disconnected if there is a constant $C_0 > 0$ so that for each $x \in E$ and $r > 0$ there is a set F such that $E \cap B(x,r) \subseteq F \subseteq E \cap B(x, C_0 r)$ and* $\operatorname{dist}(F, E) \geq C_0^{-1} r$. *(The latter will be considered to hold vacuously when $E = F$.)*

In other words, $E \cap B(x,r)$ should be contained in a little island in E which is not too much larger and which is at a definite distance from the rest of E. The standard Cantor set has this property, as well as practically anything constructed in a similar manner, like the set A in the preceding section.

Proposition 4.3. *The set A constructed in Section 3 is uniformly disconnected (for all k, which we require to be $\geq 10^7$, as in Section 3).*

Proposition 4.4. *If $E \subseteq \mathbb{R}^n$ is closed and uniformly disconnected, then the weight $\Omega(x) = \min(1, \operatorname{dist}(x, E)^s)$ is a strong A_∞ continuous weight for all $s > 0$.*

Corollary 4.5. *If ω is defined on $\mathbb{R}^3$ as in (4.1), then ω is a strong A_∞ continuous weight for all $s > 0$ and $k \geq 10^7$.*

Of course the corollary is an immediate consequence of the propositions.

Let us prove Proposition 4.3. We shall use freely the notations and results from the previous section. All constants C which appear in the argument below will be permitted to depend on k. Indeed, since A approximates a circle in the Hausdorff topology when k gets big, the uniform disconnectedness constant for A has to blow up as $k \to \infty$.

Let $x \in A$ and $r > 0$ be given, as in Definition 4.2. We may as well assume that $r < 10^{-10}\rho(k)$; if $r \geq 10^{-10}\rho(k)$, then we can simply take $F = A$, and the requirements of Definition 4.2 will be satisfied.

Choose $l \geq 1$ as large as possible so that $r < 10^{-10}\rho(k)^l$, and let α be the constituent circle in A_l such that $x \in \tau(\alpha)$. Let δ be the parent of α in A_{l-1}, and set $F = A \cap \tau(\delta)$. Then $B(x,r) \subseteq \tau(\delta)$, because $\operatorname{dist}(x, \delta) < 10^{-6}\operatorname{radius}(\delta) < 10^{-5}\operatorname{radius}(\delta) - r$ by Lemma 3.8, our choice of l, and the fact that the radius of δ is $\rho(k)^{l-1}$. Thus $F \supseteq B(x,r) \cap A$. On the other hand, $\operatorname{dist}(F, A \setminus F) \geq 10^{-5}\operatorname{radius}(\delta)$ by Lemma 3.14 (applied to δ and its cousins, and with l replaced by $l-1$) when $l-1 \geq 1$. When $l-1 = 0$ we have that $\delta = \Gamma_0$ and $F = A$.

Since $r \geq 10^{-10}\rho(k)^{l+1} \geq C^{-1}\,\text{radius}\,(\delta)$ we have that $F \subseteq B(x, Cr)$. Altogether we conclude that F has the properties required in Definition 4.2, and Proposition 4.3 follows.

Now let us prove Proposition 4.4.

Lemma 4.6. *If $E \subseteq \mathbb{R}^n$ is closed and uniformly disconnected, then there is a $C > 0$ so that for each ball B in $\mathbb{R}^n$ there is a subball $B' \subseteq B$ such that $B' \subseteq \mathbb{R}^n \setminus E$ and the radius of B' is at least C^{-1} times the radius of B.*

To prove this we may as well assume that B is centered on E. Indeed, either $(1/2)B$ is disjoint from E, in which case the conclusion of the lemma holds, or it is not, in which case we can find a ball contained in B with exactly half the radius and whose center is an element of E.

With B centered on E let us apply Definition 4.2 with x, r chosen so that $B(x, 2C_0 r) = B$. This gives us a set $F \subseteq (1/2)B$ such that $\text{dist}\,(F, E \setminus F) \geq C_0^{-1} r$. Using this it is easy to check that B has a subball with the required properties.

One consequence of Lemma 4.6 is that E has Lebesgue measure zero. The vulgar reason for this is that the conclusion of Lemma 4.6 implies that E cannot have any points of density (in the sense of Lebesgue). A better reason is that E must even have Minkowski dimension less than n. At any rate we deduce that Ω as in Proposition 4.4 is at least a continuous weight.

Lemma 4.7. *If $E \subseteq \mathbb{R}^n$ is closed and uniformly disconnected, and if Ω is as in Proposition 4.4, then there is a $C > 0$ so that*

$$\sup_{2B} \Omega \leq C \, \frac{1}{|B|} \int_B \Omega \,, \tag{4.8}$$

for all balls B in $\mathbb{R}^n$.

By Lemma 4.6, every ball with radius 1 contains a ball disjoint from E whose radius is larger than some fixed positive number. Since Ω is bounded from below on at least one-half this ball we get that the right side of (4.8) is bounded from below when B has radius at least 1 (by an easy argument). This implies that (4.8) holds (with a suitable constant) when the radius of B is at least 1, since $\Omega \leq 1$ by definition.

Suppose that B has radius less than 1. If $3B$ is disjoint from E, then $\sup_{2B} \Omega \leq C \inf_B \Omega$ for a suitable constant C by the definition of Ω and simple geometric considerations, and so (4.8) is satisfied. If $3B$ intersects E, then $\sup_{2B} \Omega \le (5\,\mathrm{radius}\,(B))^s$, while the right side of (4.8) is bounded from below by $C^{-1}\mathrm{radius}(B)^s$ for some constant C, because of Lemma 4.6 and the definition of Ω. In this case also (4.8) holds, and Lemma 4.7 follows.

Lemma 4.7 implies that Ω is doubling and that

$$D_\Omega(x,y) \le C\,\Omega(B_{x,y})^{1/n}.$$

(See Definition 1.5.)

Lemma 4.9. *Suppose that $E \subseteq \mathbb{R}^n$ is closed and uniformly disconnected, and let $x, y \in \mathbb{R}^n$ and a curve γ which connects them be given. Then there is a point p on γ such that $|p - x| \le |y - x|$ and $\mathrm{dist}\,(p, E) \ge C^{-1}|y - x|$ for some C which does not depend on x, y, or p.*

This is easy to verify using the uniform disconnectedness condition.

To finish the proof of Proposition 4.4 it remains to show that if E and Ω are as above and $x, y \in \mathbb{R}^n$ and a curve γ which connects them are given, then $\int_\gamma \Omega^{1/n}\,ds \ge C^{-1}\Omega(B_{x,y})^{1/n}$, where $B_{x,y}$ and $\Omega(B_{x,y})$ are as in Definition 1.5. Apply Lemma 4.9 to get a point p as above. This means that there is a ball B centered at p and with radius equal to $C_1^{-1}|y - x|$ such that $2B$ is disjoint from E for some constant C_1. This condition implies in turn that $\sup_B \Omega \le C_2 \inf_B \Omega$ for some constant C_2. Hence

$$\begin{aligned}\int_\gamma \Omega^{1/n}\,ds &\ge \int_{\gamma\cap B} \Omega^{1/n}\,ds \\ &\ge C_1^{-1}\,(\inf_B \Omega^{1/n})\,|y - x| \ge C^{-1}\,\Omega(B)^{1/n}.\end{aligned} \tag{4.10}$$

On the other hand, $\Omega(B)^{1/n} \ge C^{-1}\Omega(B_{x,y})^{1/n}$, since Ω is doubling, $|p - x| \le |y - x|$, and the radius of B is not too small compared to $|y - x|$, and hence $\int_\gamma \Omega^{1/n}\,ds \ge C^{-1}\Omega(B_{x,y})^{1/n}$, as desired. Thus Ω satisfies the strong-A_∞ condition, and the proof of Proposition 4.4 is complete.

Notice that for the weight Ω as in Proposition 4.4 the condition (1.8) is much more obvious than for generic strong A_∞ weights, because of Lemma 4.7.

Next, let A be as in Section 3 and ω be as in (4.1), and let us estimate the Hausdorff dimension of A with respect to D_ω in terms of the parameters s and k (from (4.1) and Section 3). For simplicity we shall call this the ω-Hausdorff dimension of A. The main point is to know when this dimension is less than 1 (or even small) so that we can apply Corollary 3.18 to get restrictions on Lipschitz or Hölder continuous maps from $(\mathbb{R}^3, D_\omega)$ to $\mathbb{R}^3$ equipped with the Euclidean metric.

To estimate the ω-Hausdorff dimension of A we need to cover A with little blobs, estimate the ω-diameter of the little blobs, and then bound the usual series. The computation is simplified by the homogeneity and self-similarity properties of A and ω. For each l let $\mathcal{A}_l$ denote the collection of constituent circles in A_l. Then $\tau(A_l) = \cup_{\alpha \in \mathcal{A}_l} \tau(\alpha)$, by definition, and so

$$A \subseteq \bigcup_{\alpha \in \mathcal{A}_l} \tau(\alpha) \tag{4.11}$$

by (3.12). Thus we can use the $\tau(\alpha)$, $\alpha \in \mathcal{A}_l$, as the little blobs.

We need to estimate the ω-diameters of the $\tau(\alpha)$'s, where "ω-diameter" means the diameter with respect to D_ω. Observe first that $A \cap \tau(\alpha) \neq \varnothing$. (Lemma 3.10 is helpful in this regard.) Thus

$$\sup_{\tau(\alpha)} \omega \leq (2\,\text{radius}\,(\alpha))^{3s} \tag{4.12}$$

by definition of ω. Hence

$$\omega\text{-diameter}\,(\tau(\alpha)) \leq C\,\text{radius}\,(\alpha)^{1+s} \tag{4.13}$$

by the definition of D_ω. By construction all the $\alpha \in \mathcal{A}_l$ have radius $\rho(k)^l$, and so we get

$$\omega\text{-diameter}\,(\tau(\alpha)) \leq C\,\rho(k)^{l(1+s)}\,. \tag{4.14}$$

Remember also that there are k^l elements of $\mathcal{A}_l$. Using this fact, (4.14), and the definition of Hausdorff measure we get the following.

Lemma 4.15. *The ω-Hausdorff dimension of A is less or equal than a if* $\limsup_{l\to\infty} k^l \rho(k)^{l(1+s)a} < \infty$.

It is not hard to see that the estimate on the ω-Hausdorff dimension provided by this lemma is sharp, but we shall not need that fact. (The main point is to use the natural probability measure on A to get lower bounds on ω-Hausdorff measure.)

Although we could go back now and analyze $\rho(k)$ carefully to get as much out of Lemma 4.15 as possible, a much cruder analysis will be sufficient for our purposes.

Lemma 4.16. *The ω-Hausdorff dimension of A is at most* $3(1+s)^{-1}$.

To see this we observe that

$$\sup_l k^l \rho(k)^{3l} < \infty \,. \tag{4.17}$$

This bound follows from the observation that $\rho(k)^{3l}$ is a constant multiple of the Lebesgue measure of each $\tau(\alpha)$, $\alpha \in \mathcal{A}_l$, so that the left side of (4.17) is dominated by the Lebesgue measure of a compact set in $\mathbb{R}^3$. Lemma 4.16 follows immediately from (4.17) and Lemma 4.15. Of course Lemma 4.16 is not at all sharp, but it enjoys the simplicity of providing a bund which does not depend on k.

Lemma 4.18. *For each fixed $s > 0$ the ω-Hausdorff dimension of A is less or equal than $(1+s/2)^{-1}$ for k sufficiently large.*

Recall from (3.1) that $\rho(k) \leq 2\pi k^{-1}$. Thus Lemma 4.15 implies that the ω-Hausdorff dimension of A is less or equal than a if $\limsup_{l\to\infty} k^{l-l(1+s)a}(2\pi)^{l(1+s)a} < \infty$. If we take $a = (1+s/2)^{-1}$, then $1-(1+s)a < 0$, and $k^{l-l(1+s)a}(2\pi)^{l(1+s)a} \to 0$ as $l \to \infty$ when k is sufficiently large. Lemma 4.18 follows.

We are now ready to Prove Theorem 1.10 and some variants of it. Let us first set some terminology. Let h be a map from an open subset U of $\mathbb{R}^3$ in to $\mathbb{R}^3$, viewed as a map from U equipped with the metric D_ω into $\mathbb{R}^3$ equipped with the Euclidean metric. We shall write this more succinctly as $h : (U, D_\omega(x,y)) \to (\mathbb{R}^3, |x-y|)$. We say that h is *locally Hölder continuous of order δ* if for each compact set $K \subseteq U$ there is a constant $C = C(K)$ such that

$$|h(x) - h(y)| \leq C\, D_\omega(x,y)^\delta \,, \tag{4.19}$$

for all $x, y \in K$. When this holds with $\delta = 1$ we say that h is *locally Lipschitz.*

Theorem 4.20. *Let A and k be as in Section 3, and let ω and $s > 0$ be as in (4.1). Let U be any open subset of $\mathbb{R}^3$ which intersects A, and let V be any other open subset of $\mathbb{R}^3$. (For instance take $U = V = \mathbb{R}^3$.)*

a) *If $s > 2$ then there does not exist a homeomorphism $h : (U, D_\omega(x,y)) \to (V, |x-y|)$ which is locally Lipschitz.*

b) *For any $s > 0$ there does not exist a homeomorphism $h : (U, D_\omega(x,y)) \to (V, |x-y|)$ which is locally Lipschitz if k is large enough.*

c) *For any $s > 0$ there does not exist a homeomorphism $h : (U, D_\omega(x,y)) \to (V, |x-y|)$ which is locally Hölder continuous of order greater than $3(1+s)^{-1}$.*

Like the lemmas that came before it, the bounds in Theorem 4.20 are not sharp.

For the proof of Theorem 4.20 we we shall need the following.

Lemma 4.21. *If $U \subseteq \mathbb{R}^3$ is open, $h : (U, D_\omega(x,y)) \to (\mathbb{R}^3, |x-y|)$ is locally Hölder continuous of order δ, and $K \subseteq U$ is a compact set with ω-Hausdorff dimension less or equal than a, then the Euclidean Hausdorff dimension of $h(K)$ is less or equal than $\delta\, a$.*

This is a well-known and straightforward consequence of the definitions.

Recall now that Corollary 3.19 says that no homeomorphism $h : U \to V$ can send $U \cap A$ to a set of Euclidean Hausdorff dimension less than 1. Using this and Lemma 4.21, parts a) and c) of Theorem 4.20 follow from Lemma 4.16, and part b) follows from Lemma 4.18.

Notice that part c) contains a) as a special case, but the point of c) is more the fact that the Hölder exponent of any homeomorphism $h : (\mathbb{R}^3, D_\omega(x,y)) \to (\mathbb{R}^3, |x-y|)$ has to go to 0 as s gets large. By contrast we have the following simple and well-known fact.

Proposition 4.22. *If ω is any strong A_∞ weight on $\mathbb{R}^3$ (or $\mathbb{R}^n$, for that matter), then the identity mapping on $\mathbb{R}^3$ is locally Hölder continuous as a map from $(\mathbb{R}^3, D_\omega(x,y))$ to $(\mathbb{R}^3, |x-y|)$ of some positive order.*

Because of Definition 1.5, Proposition 4.22 reduces itself to the following.

Lemma 4.23. *If Ω is a continuous weight on $\mathbb{R}^n$ which is doubling, then for each compact set $K \subseteq \mathbb{R}^n$ there exist positive constants C and N such that $\Omega(B(x,r)) \geq C^{-1}r^N$ for all $x \in K$ and $0 < r < 1$. Here C depends on K but N does not.*

Indeed, the doubling property implies that there is a constant C_0 such that $\Omega(B(y,t)) \geq C_0^{-1}\Omega(B(y,2t))$ for all $y \in \mathbb{R}^n$ and $t > 0$. Let K, x, and r be given as in the lemma. If j is a positive integer which is sufficiently large so that $B(x,2^j r) \supseteq K$, then $\Omega(B(x,r)) \geq C_0^{-j}\Omega(B(x,2^j r)) \geq C_0^{-j}\Omega(K)$. The lemma follows from this estimate, since we can choose j to be $-\log_2 r + C$.

Proposition 4.22 can be improved in various ways. A similar argument can be used to show that the identity is locally Hölder continuous as a map in the other direction, *i.e.*, from $(\mathbb{R}^3, |x-y|)$ to $(\mathbb{R}^3, D_\omega(x,y))$. The identity mapping is in truth quasisymmetric, essentially by definition of D_ω and the strong-A_∞ condition. This implies a scale-invariant version of Proposition 4.22 in particular. Estimates like the reverse Hölder inequality (1.8) can be used to show that the identity mapping as a map from $(\mathbb{R}^3, D_\omega(x,y))$ to $(\mathbb{R}^3, |x-y|)$ (or the other way around) satisfies local Sobolev space estimates. (These Sobolev spaces have to be formulated carefully, since we are dealing with metric spaces.)

One can think of part c) of Theorem 4.20 as saying that $(\mathbb{R}^3, D_\omega(x,y))$ can be as far as possible from being bilipschitz equivalent to $(\mathbb{R}^3, |x-y|)$ even though ω is a strong A_∞ weight. Part b) is a complement to this. It says that we can make the singularity of the weight ω as small as we want while still having $(\mathbb{R}^3, D_\omega(x,y))$ and $(\mathbb{R}^3, |x-y|)$ be bilipschitz inequivalent. In terms of the metrics this means that the identity mapping as a map from $(\mathbb{R}^3, D_\omega(x,y))$ to $(\mathbb{R}^3, |x-y|)$ is locally Hölder continuous of order δ where $\delta \to 1$ as $s \to 0$. On the other hand, the identity is trivially Lipschitz as a map from $(\mathbb{R}^3, |x-y|)$ to $(\mathbb{R}^3, D_\omega(x,y))$, since $\omega \leq 1$ by definition. As before, there are various improvements of the statement about Hölder continuity, in terms of Sobolev space estimates, for instance. The bottom line is that we can choose ω so that it is a strong A_∞ weight and so that $(\mathbb{R}^3, D_\omega(x,y))$ is as close as we want to being bilipschitz equivalent to $(\mathbb{R}^3, |x-y|)$ without actually being bilipschitz equivalent.

Note however that the strong-A_∞ constant of ω blows up as s gets small; thus part b) does not really provide counterexamples to suitable small-constant versions of Question 1.9 (as mentioned briefly in Section 2, after the discussion of Problem 2.7).

REMARK 4.24. If one took a non-wild version of the Cantor set A from the preceding section (*e.g.*, by replacing (3.3) by the requirement that the $\gamma_0(p)$'s be unlinked), and defined ω as in (4.1), then $(\mathbb{R}^3, D_\omega(x,y))$ would be bilipschitz equivalent to $(\mathbb{R}^3, |x-y|)$, no matter how large s is. (There is also nothing special about dimension 3 here.) This is not very hard to prove. One could use the same construction as in the next section. (See Remark 5.28.)

To finish the proof of Theorem 1.10 it remains to establish the last part, about connecting an arbitrary pair of points by a set which is bilipschitz equivalent to a standard Euclidean ball.

Proposition 4.25. *Let A be as in Section* 3 (*with some choice of k*) *and let ω be as in* (4.1) (*with some choice of s*). *There is a constant L_2* (*depending on k and s*) *with the property that for every pair of distinct points $p, q \in \mathbb{R}^3$ there is a closed subset W of $\mathbb{R}^3$ containing p and q such that (W, D_ω) is L_2-bilipschitz equivalent to a closed Euclidean 3-ball with the standard Euclidean metric.*

The rest of this section will be devoted to the proof of Proposition 4.25. The proof is basically trivial, but one should be a little careful. The basic idea is to connect p and q by a curve which stays away from A as much as possible, and which is as smooth as it can be subject to this constraint (and has no self-intersections), and then to take W to be a fattened-up version of this curve. On this set W the "strangeness" (deviation from Euclidean geometry) of the metric D_ω will be essentially like the strangeness of D_ω on a curve, and we shall be able to get rid of it easily.

Let $p, q \in \mathbb{R}^3$ be given, $p \neq q$. Given a pair of points y, z in $\mathbb{R}^3$ and an ε in $(0,1)$, let $S(y,z)$ denote the segment which joins y to z, and let $S_\varepsilon(y,z)$ be the set of points x in $\mathbb{R}^3$ such that $\operatorname{dist}(x, S(y,z)) \leq \varepsilon \operatorname{dist}(x, \{y,z\})$. Thus $S_\varepsilon(y,z)$ is the union of two truncated cones, one with vertex y, the other with vertex z. It is also bilipschitz equivalent to a Euclidean ball, with a bilipschitz constant which depends only on ε. In order to produce a set W as in the proposition it is better to think of W as being bilipschitz equivalent to some $S_\varepsilon(y,z)$ rather than a round ball. Typically W will look like a twisted version of $S_\varepsilon(y,z)$, with some spiralling at the ends. The segment $S(y,z)$ will correspond to the curve mentioned in the previous paragraph.

Let us begin with a preliminary fact.

Lemma 4.26. *Suppose that X is a nonempty subset of $\mathbb{R}^3 \setminus A$ which satisfies $\operatorname{dist}(X,A) \geq \mu \min\{\operatorname{diam} X, 1\}$ for some $\mu > 0$. Then there is a constant $C = C(\mu)$ such that*

$$\sup_X \omega \leq C \inf_X \omega \tag{4.27}$$

and

$$C^{-1}\omega(x)^{1/3}|x-y| \leq D_\omega(x,y) \leq C\,\omega(x)^{1/3}|x-y|\,, \tag{4.28}$$

for all $x, y \in X$.

The first part (4.27) follows immediately from the definition (4.1) of ω. To establish the second part let us observe that

$$C^{-1}\omega(z)\,|B| \leq \omega(B) \leq C\,\omega(z)\,|B|\,, \tag{4.29}$$

whenever B is a Euclidean ball which contains some $z \in X$ and which has diameter at most the diameter of X. Here $|B|$ denotes the Euclidean volume of B. These inequalities follow from our assumptions on X and the definition of ω. (It is helpful to distinguish between the cases where $\operatorname{dist}(X,A) \leq 1$ and $\operatorname{dist}(X,A) \geq 1$.) To prove (4.28) we use the fact that ω is a strong-A_∞ weight (Corollary 4.5 and Definition 1.5) to reduce to (4.29). This proves the lemma.

We shall use heavily the notation and definitions from Section 3 in the rest of the proof of Proposition 4.25.

Lemma 4.30. *The conclusion of Proposition 4.25 holds when there is a circle α in some A_l such that $p, q \in \tau(\alpha) \setminus \tau(N^2(\alpha))$. The same is true when $p, q \in \mathbb{R}^3 \setminus \tau(N(\Gamma_0))$.*

To prove this let us first observe that we can find a mapping g from a closed Euclidean ball into $X = \tau(\alpha) \setminus \tau(N^2(\alpha))$ or $X = \mathbb{R}^3 \setminus \tau(N(\Gamma_0))$ (as appropriate) such that the image of g contains p, q and g

is bilipschitz with respect to the Euclidean metrics (and not D_ω) with a uniformly bounded constant. In the first case, where $X = \tau(\alpha) \setminus \tau(N^2(\alpha))$, this is true because X is a nice smooth (connected) domain, a torus with a few tori removed. These domains can all be realized as images of each other under similarities, which makes transparent the existence of estimates which do not depend on α. The second case requires a tiny bit of extra care but is standard. (Think about $\mathbb{R}^3 \setminus B(0,1)$ first.) Once we have such a mapping g we can precompose with a dilation if necessary to get a uniformly bilipschitz map from a Euclidean ball into $(\mathbb{R}^3, D_\omega)$ whose image contains p and q (because of Lemma 4.26). This proves Lemma 4.30.

To deal with the remaining cases of Proposition 4.25 we cannot simply "localize" in this manner, but instead we have to connect p and q with chains of sets which each satisfy separately the hypotheses of Lemma 4.26. The next lemma covers the most interesting case, and its proof will take a while.

Lemma 4.31. *If $p, q \in A$, then the conclusions of Proposition 4.25 hold.*

Choose δ in some A_m such that $p, q \in \tau(\delta)$ and m is as large as possible. Note that $m < \infty$. Let α_l, β_l be the (unique) circles in A_l such that $p \in \tau(\alpha_l)$, $q \in \tau(\beta_l)$, respectively, so that $\alpha_{l+1} \in N(\alpha_l)$ for each l, etc. Choose (arbitrarily) points p_l and q_l in the boundaries of $\tau(\alpha_l)$ and $\tau(\beta_l)$ for each $l > m$. Of course $p_l \to p$ and $q_l \to q$ as $l \to \infty$. Our bilipschitz ball W will be obtained by combining a family of smooth tubes which connect the successive p_l's and q_l's.

We should record some bounds on distances and diameters. Let us write $X \approx Y$ when the two quantities X and Y are each bounded by a constant times the other, where the constant is allowed to depend on our parameters k and s but nothing else. Thus

$$(4.32) \qquad \begin{aligned} &|p_l - p_{l+1}| \approx \operatorname{diam} \alpha_l , \\ &D_\omega(p_l, p_{l+1}) \approx D_\omega\text{-diam}\, \alpha_l \approx (\operatorname{diam} \alpha_l)^{1+s} , \end{aligned}$$

and similarly for the q_l's. Here "diam" refers to the diameter with respect to the Euclidean metric, while "D_ω-diam" refers to the diameter with respect to D_ω. Remember that

$$(4.33) \qquad \operatorname{diam} \gamma = \rho(k)^l \operatorname{diam} \Gamma_0 \qquad \text{when } \gamma \in A_l ,$$

because of the construction in Section 3. This implies that

$$D_\omega(p_l, p_{l+1}) \approx D_\omega(p_{l+1}, p_{l+2}), \tag{4.34}$$

for all l, and similarly for the q_l's, and also

$$D_\omega(p_{m+1}, p_{m+2}) \approx D_\omega(p_{m+1}, q_{m+1}) \approx D_\omega(q_{m+1}, q_{m+2}). \tag{4.35}$$

Notice that $\alpha_l \neq \beta_l$ when $l > m$, because of the maximality of m. This implies that

$$|p_{m+1} - q_{m+1}| \geq C^{-1} \operatorname{diam} \delta \tag{4.36}$$

for some constant C.

Fix any line J in $\mathbb{R}^3$, and choose points z_l and w_l in J for $l > m$ in the following manner. These points are supposed to correspond to the p_l's and q_l's, and this will be made more precise soon. Choose z_{m+1} and w_{m+1} first, in such a way that $|z_{m+1} - w_{m+1}| = D_\omega(p_{m+1}, q_{m+1})$. Except for this constraint the specific choices do not matter. When $l > m+1$ let z_l and w_l be the (unique) points such that $|z_l - z_{l-1}| = D_\omega(p_l, p_{l-1})$, $|w_l - w_{l-1}| = D_\omega(q_l, q_{l-1})$ for all $l > m+1$ and such that the z_l's and w_l's are ordered correctly. This means that z_l is always on the opposite side of z_{l-1} from z_{l-2}, and similarly for the w_l's, and that z_{m+2} lies on the opposite side of z_{m+1} from w_{m+1}, and that w_{m+2} lies on the opposite side of w_{m+1} from z_{m+1}. Let $z \in J$ be the limit of the z_l's, and let $w \in J$ be the limit of the w_l's. These points will correspond to our original p and q. Note that all the z_l's and w_l's lie on $S(z,w)$, because of our ordering, and that the sequences $\{|z_l - z_{l-1}|\}_l$ and $\{|w_l - w_{l-1}|\}_l$ are approximately geometric sequences, because of (4.32) and (4.33).

In order to prove Lemma 4.31 it suffices to find $\varepsilon > 0$ and a bilipschitz mapping f from $(S_\varepsilon(z,w), |x-y|)$ into $(\mathbb{R}^3, D_\omega)$ (with uniform choices of ε and the bilipschitz constant) such that $f(z) = p$ and $f(w) = q$. We shall define f in stages. To understand how f is constructed it is helpful to visualize the region $f(S_\varepsilon(z,w))$ that we shall have to construct. It will be a union of little tubes, where the tubes connect the successive p_l's and q_l's. These tubes will be diffeomorphic to rectangles and they will be neither too thin nor too close to A. To build these tubes we shall first choose some smooth Jordan arcs which connect the successive p_l's and q_l's, and the tubes will be little tubular neighborhoods of these arcs. Before we do all these things let us define f initially on the z_l's and w_l's in the obvious way.

Sublemma 4.37. *Let $\mathcal{E}$ be the set consisting of z, w, and the z_l's and w_l's for $l > m$, and define $f : \mathcal{E} \to \mathbb{R}^3$ by $f(z) = p$, $f(w) = q$, $f(z_l) = p_l$, and $f(w_l) = q_l$. Then f is bilipschitz as a map from $(\mathcal{E}, |x-y|)$ into $(\mathbb{R}^3, D_\omega)$ with uniformly bounded bilipschitz constant.*

We defined f so that it satisfies the bilipschitz condition for consecutive points in $\mathcal{E}$. In order to check the bilipschitz condition for pairs of points which are further apart it is helpful to make some additional observations. If $j > l+1$, then $\tau(\alpha_j) \subseteq \tau(N^2(\alpha_l))$, because of Lemma 3.10. This implies that $\operatorname{dist}(\tau(\alpha_l) \setminus \tau(N(\alpha_l)), \tau(\alpha_j)) \approx \operatorname{diam} \alpha_l$, and similarly for the β's. The constants implicit in this statement do not depend on l or j, as one can verify most easily using the self-similarity of the constuction in Section 3. We also get that $D_\omega\text{-}\operatorname{dist}(\tau(\alpha_l) \setminus \tau(N(\alpha_l)), \tau(\alpha_j)) \approx D_\omega\text{-}\operatorname{diam} \alpha_l$, and similarly for the β's, with a uniform choice of the constant. This is a variant of (4.28) which can also be derived from (4.29). (Note that $X = \tau(\alpha_l) \setminus \tau(N(\alpha_l))$ satisfies the hypotheses of Lemma 4.26.) Thus we can control the interaction between the various α's, and between the β's. We also have that the α's and the β's do not interact with each other. Specifically, $\tau(\alpha_j) \subseteq \tau(\alpha_{m+1})$ when $j > m$, because of Lemma 3.10, and similarly for the β's. The maximality of m implies that $\alpha_{m+1} \neq \beta_{m+1}$, and so Lemma 3.14 yields $\operatorname{dist}(\tau(\alpha_{m+1}), \tau(\beta_{m+1})) \approx \operatorname{diam} \delta$. We can convert this into $D_\omega\text{-}\operatorname{dist}(\tau(\alpha_{m+1}), \tau(\beta_{m+1})) \approx D_\omega\text{-}\operatorname{diam} \delta$ using the definition of ω and D_ω. In other words we can control the interaction between all the α's and all the β's. It is easy to verify Sublemma 4.37 using these estimates.

Next we want to define f on $S(z, w)$. Let us first record a simple observation about curves which will provide the building blocks for this extension of f.

Sublemma 4.38. *Given any (marked) circle γ in $\mathbb{R}^3$ and any pair of points a, b in different components of the boundary of $\tau(\gamma) \setminus \tau(N(\gamma))$, we can find an arc σ in the closure of $\tau(\gamma) \setminus \tau(N(\gamma))$ which connects a to b and has the following properties: if u and v are two points on σ, then the length of the arc in σ which connects u to v is bounded by $C|u-v|$; inside $B(a, C^{-1}\operatorname{diam}\gamma)$ the curve σ agrees with the line segment emanating from a which is orthogonal to the boundary of $\tau(\gamma) \setminus \tau(N(\gamma))$ at a and goes inside $\tau(\gamma) \setminus \tau(N(\gamma))$, and similarly for b; if $u \in \sigma$, then* $\operatorname{dist}(u, \mathbb{R}^3 \setminus \{\tau(\gamma) \setminus \tau(N(\gamma))\}) \geq C^{-1} \operatorname{dist}(u, \{a, b\})$ *(so that*

σ does not get close to the boundary except near the endpoints); for each positive integer i the Euclidean norm of the i^{th} derivative of the arclength parameterization of σ is bounded by $C(i)(\operatorname{diam}\gamma)^{1-i}$. (This is the "scale-invariant" estimate on the higher derivatives.) Here the constants C and $C(i)$ depend only on the parameter k from Section 3.

This is an easy exercise. Note that $\tau(\gamma) \setminus \tau(N(\gamma))$ is connected, and that we can reduce to the case where $\gamma = \Gamma_0$ by using a similarity.

Sublemma 4.39. *There is a map $f : S(z,w) \to \mathbb{R}^3$ which is smooth away from the endpoints and satisfies the following properties: f is defined on $\mathcal{E}$ as in Sublemma 4.37; f is bilipschitz as a map from $(S(z,w), |x-y|)$ into $(\mathbb{R}^3, D_\omega)$ with uniformly bounded bilipschitz constant; $f(S(z_{m+1}, w_{m+1}))$ is contained in the closure of $\tau(\delta) \setminus \tau(N(\delta))$, $f(S(z_l, z_{l+1}))$ is contained in the closure of $\tau(\alpha_l) \setminus \tau(N(\alpha_l))$ when $l > m$, and $f(S(w_l, w_{l+1}))$ is contained in the closure of $\tau(\beta_l) \setminus \tau(N(\beta_l))$ when $l > m$. In particular,*

$$\text{(4.40)} \qquad \operatorname{dist}(f(t), A) \approx \operatorname{diam}\alpha_l\,, \qquad \textit{when } t \in S(z_l, z_{l+1})\,,\ l > m\,,$$

and similarly for $S(z_{m+1}, w_{m+1})$ and the $S(w_l, w_{l+1})$'s, and

$$\text{(4.41)} \qquad \operatorname{dist}(f(t), A)^{1+s} \approx \operatorname{dist}(t, \{z,w\})\,, \qquad \textit{for all } t \in S(z,w)\,.$$

Moreover, if $f^{(i)}$ denotes the i^{th} order derivative of f on $S(z,w)$, then

$$\text{(4.42)} \qquad |f^{(i)}(t)| \le C(i) \operatorname{dist}(f(t), A) \operatorname{dist}(t, \{z,w\})^{-i}\,,$$

for all $t \in S(z,w) \setminus \{z,w\}$ and $i \ge 1$, where $C(i)$ depends on i and the parameters k and s but nothing else and $|\cdot|$ denotes the ordinary Euclidean norm.

Before explaining how to build f -which comes down to connecting arcs as in Sublemma 4.38 and parameterizing them at the correct "speed"- let us consider the slightly odd-looking bound for $|f^{(i)}|$ in (4.42). The first point to notice is that we could write (4.42) in many different ways using (4.41). When $i = 1$, for instance, (4.42) reduces to saying that the first derivative of f at t is bounded by $C \operatorname{dist}(f(t), A)^{-s}$, which is compatible with the bilipschitz condition on f. (In fact the bilipschitzness requires that

$$\text{(4.43)} \qquad |f'(t)| \approx \operatorname{dist}(f(t), A)^{-s} \approx \operatorname{dist}(f(t), A) \operatorname{dist}(t, \{z,w\})^{-1}$$

and this estimate will also be clear from the proof of Sublemma 4.39.) The bounds on the higher derivatives of f in (4.42) basically mean that f is chosen to be as smooth as it can be on each of the segments $S(z_{m+1}, w_{m+1})$, $S(z_l, z_{l+1})$, and $S(w_l, w_{l+1})$, subject to the speed limit (4.43) and the fact that f will probably have to do some nontrivial turning on each of these segments.

To prove Sublemma 4.39 we begin by defining f on $\mathcal{E}$ as in Sublemma 4.37. To define f on the segments $S(z_{m+1}, w_{m+1})$, $S(z_l, z_{l+1})$, and $S(w_l, w_{l+1})$ we connect the p_l's and q_l's by arcs as in Sublemma 4.38 and we take f to be certain parameterizations of these arcs. We cannot use arclenth parameterizations, but instead we parameterize these arcs at roughly constant speed. "Roughly constant" means that the maximum speed is bounded by a constant times the minimal speed. On an $S(z_l, z_{l+1})$, for instance, this approximate speed is comparable to $|p_l - p_{l+1}|/|z_l - z_{l+1}|$, and this is in turn comparable to $(\operatorname{diam} \alpha_l)^{-s}$, because of our choices of the p_l's and z_l's. Note that this average speed on $S(z_l, z_{l+1})$ is comparable to that of the adjacent intervals, as is also the case for $S(z_{m+1}, w_{m+1})$ and the $S(w_l, w_{l+1})$'s. We cannot use parameterizations which have truly constant speed, since that would lead to discontinuities of the derivative of f at the z_l's and w_l's. Instead we require that the speeds be roughly constant on these intervals while also making a gentle transition from one interval to the next. If we take some care to make the "gentle transitions" of the parameterizations of the adjacent arcs approximately as gentle as they can be, then the estimates for the higher derivatives of f in (4.42) will follow from the corresponding estimates in Sublemma 4.38 and the normalized behavior of the arcs in Sublemma 4.38 at their endpoints. It is not hard to see that the mapping f that we produce in this way is bilipschitz on the union of any two adjacent intervals among $S(z_{m+1}, w_{m+1})$, the $S(z_l, z_{l+1})$'s, and the $S(w_l, w_{l+1})$'s. This uses the properties of the curves in Sublemma 4.38 (especially the chord-arc property), the fact that D_Ω is approximately a constant multiple of the Euclidean metric on regions like $\tau(\alpha_l) \setminus \tau(N^2(\alpha_l))$ (by Lemma 4.26), and the fact that we chose the z_l's and the w_l's so that the constant multiples work out correctly. The bilipschitzness on all of $S(z, w)$ follows from an argument like the one used to prove Sublemma 4.37, using also the fact that $f(S(z_l, z_{l+1}))$ is contained in the closure of $\tau(\alpha_l) \setminus \tau(N(\alpha_l))$ when $l > m$, etc. The estimate (4.40) is an immediate consequence of the definition of f (and the constructions in Section 3), and (4.41) follows from (4.40) and our choices of the z_l's and w_l's. This proves Sublemma 4.39.

From now on we assume that f is defined on $S(z,w)$ as in Sublemma 4.39. We want to extend f to some $S_\epsilon(z,w)$. In the following we set $v_0 = (w-z)/|w-z|$ and we let $f'(t)$ denote the derivative of f in the direction v_0. This is defined for all $t \in S(z,w) \setminus \{z,w\}$, and it is a vector in $\mathbb{R}^3$.

Sublemma 4.44. *There is a smooth map F from $S(z,w)\setminus\{z,w\}$ into linear mappings on $\mathbb{R}^3$ such that each $F(t)$ is an orientation-preserving similarity (a combination of a rotation and a dilation) which satisfies $F(t)v_0 = f'(t)$ and*

$$(4.45) \qquad |F^{(i)}(t)| \leq C(i)\,\mathrm{dist}\,(f(t),A)\,\mathrm{dist}\,(t,\{z,w\})^{-i-1}\,,$$

for all $t \in S(z,w)\setminus\{z,w\}$ and $i \geq 0$, where $C(i)$ depends on i and the parameters k and s but nothing else. (In other words, F satisfies the same bounds on its derivatives as f' does.)

Notice first that the dilation factor for $F(t)$ must simply be $|f'(t)|$, which is in turn controlled by (4.43). In particular $|f'(t)|$ never vanishes.

Sublemma 4.44 is basically trivial but we should be a little careful. Let F' denote the derivative of F in the direction of v_0. Instead of trying to choose F directly let us write $F'(t) = \Phi(t)\,F(t)$ and choose $\Phi(t)$ instead. We should choose Φ so that $f''(t) = \Phi(t)\,f'(t)$ and $\Phi(t) = (\log|f'(t)|)'\,I + \phi(t)$, where $\phi(t)$ is antisymmetric (with respect to the usual Euclidean inner product). If we set $v(t) = |f'(t)|^{-1}f'(t)$, then we can reformulate the constraint that $f''(t) = \Phi(t)\,f'(t)$ in terms of ϕ as $v'(t) = \phi(t)\,v(t)$. We can produce such a ϕ algebraically; we take $\phi(t)$ to be the of the rank 1 map which sends $v(t)$ to $v'(t)$ minus its transpose. This choice of $\phi(t)$ satisfies $v'(t) = \phi(t)\,v(t)$ because $v'(t)$ is orthogonal to $v(t)$ (since $|v(t)| = 1$ for all t).

We are now ready to define F by solving the differential equation. Choose $F(z_{m+1})$ to be any orientation-preserving similarity on $\mathbb{R}^3$ which satisfies $F(z_{m+1})v_0 = f'(z_{m+1})$, and extend F to all of $S(z,w)\setminus\{z,w\}$ by solving $F'(t) = \Phi(t)\,F(t)$. Our selection of $\Phi(t)$ ensures that each $F(t)$ is an orientation-preserving similarity on $\mathbb{R}^3$ which satisfies $F(t)v_0 = f'(t)$ for all t. It is easy to see that the derivatives of F satisfy the same estimates as the derivatives of f' do, because $\Phi(t)$ was chosen so that it satisfies the same estimates as $(\log|f'|)'$ does, *i.e.*,

$|\Phi^{(i)}(t)| \le C(i)\,\mathrm{dist}\,(t,\{z,w\})^{-i-1}$ for all $t \in S(z,w)\setminus\{z,w\}$ and $i \ge 0$, and because F itself has the same size as f'. This proves Sublemma 4.44.

Let us now use F to extend f in the directions orthogonal to J. Let π denote the orthogonal projection of $\mathbb{R}^3$ onto J. Given $x \in \mathbb{R}^3$ such that $\pi(x) \in S(z,w)\setminus\{z,w\}$, set

$$f(x) = f(\pi(x)) + F(\pi(x))\,(x - \pi(x)). \tag{4.46}$$

Before we analyze this extension, let us record some simple observations about $S_\varepsilon(z,w)$. If $x \in S_\varepsilon(z,w)$, then

$$|x - \pi(x)| \le \varepsilon\,\mathrm{dist}\,(x,\{z,w\}), \tag{4.47}$$

$$\begin{aligned}\mathrm{dist}\,(\pi(x),\{z,w\}) &\le \mathrm{dist}\,(x,\{z,w\})\\ &\le 2\,\mathrm{dist}\,(\pi(x),\{z,w\}),\end{aligned} \tag{4.48}$$

and

$$\pi(x) \in S(z,w) \tag{4.49}$$

when $0 < \varepsilon < 1/2$. These are all easy consequences of the definition of $S_\varepsilon(z,w)$. Actually, (4.49) holds as soon as $\varepsilon < 1$, and it implies that $S_\varepsilon(z,w)$ is contained in the domain of our extension of f.

Sublemma 4.50. *We can choose $\varepsilon > 0$ small enough so that f defines a bilipschitz map from $(S_\varepsilon(z,w), |x-y|)$ into $(\mathbb{R}^3, D_\omega)$, with ε and the bilipschitz constant depending only on the parameters k and s.*

Using the definition of f and (4.45) (with $i = 0$) we get that

$$\begin{aligned}|f(x) - f(\pi(x))| \le C\,|x - \pi(x)|\,&\mathrm{dist}\,(f(\pi(x)),A)\\ &\cdot \mathrm{dist}\,(\pi(x),\{z,w\})^{-1}.\end{aligned} \tag{4.51}$$

If $x \in S_\varepsilon(z,w)$ and $\varepsilon < 1/2$ then

$$|f(x) - f(\pi(x))| \le C\,\varepsilon\,\mathrm{dist}(f(\pi(x)),A), \tag{4.52}$$

because of (4.47) and (4.48). If ε is small enough then we get that

$$\frac{1}{2}\,\mathrm{dist}\,(f(\pi(x)),A) \le \mathrm{dist}(f(x),A) \le 2\,\mathrm{dist}\,(f(\pi(x)),A). \tag{4.53}$$

Thus $f(x)$ will not stupidly fall into A too soon. From now on we assume that $\varepsilon < 1/2$ and that ε is small enough so that (4.53) holds.

Next we observe that

$$|\nabla^i f(x)| \leq C(i)\,\mathrm{dist}\,(f(x), A)\,\mathrm{dist}\,(x, \{z, w\})^{-i}\,, \tag{4.54}$$

for all $x \in S_\varepsilon(z,w) \setminus \{z,w\}$ and $i \geq 1$, where $C(i)$ depends on i and the parameters k and s but nothing else. This is easy to check, using (4.42) and (4.45) to control the derivatives of this extension (4.46) of f, and then (4.48), and (4.53) to get the estimates in the form of (4.54) (*i.e.*, to replace $\pi(x)$ with x when necessary). Let us check that

$$|\nabla f(x)| \leq C(i)\,\mathrm{dist}\,(f(x), A)^{-s}\,, \tag{4.55}$$

for all $x \in S_\varepsilon(z,w) \setminus \{z,w\}$. Using (4.41), (4.53), and (4.48) we get that

$$\begin{aligned} \mathrm{dist}\,(f(x), A)^{1+s} &\approx \mathrm{dist}\,(f(\pi(x)), A)^{1+s} \\ &\approx \mathrm{dist}\,(\pi(x), \{z, w\}) \\ &\approx \mathrm{dist}\,(x, \{z, w\})\,. \end{aligned} \tag{4.56}$$

This and (4.54) imply (4.55). From (4.55) we obtain that f is Lipschitz as a map from $(S_\varepsilon(z,w), |x-y|)$ into $(\mathbb{R}^3, D_\omega)$ (modulo some small additional attention at the points z and w).

To show that f is bilipschitz let us consider first a special case. Let $\eta > 0$ be small, to be chosen soon, and let $t \in S(z,w)$ be given. Set $r = \mathrm{dist}\,(t, \{z,w\})$, and consider the ball $B = B(t) = B(t, \eta\, r)$. Let us show that if η is small enough, then the restriction of f to B is bilipschitz (as a map into $(\mathbb{R}^3, D_\omega)$), with a uniformly bounded bilipschitz constant. Notice first that

$$\mathrm{dist}\,(x, \{z, w\}) \approx \mathrm{dist}\,(t, \{z, w\}) \tag{4.57.a}$$

and

$$\mathrm{dist}\,(f(x), A) \approx \mathrm{dist}\,(f(t), A)\,, \qquad \text{when } x \in B\,, \tag{4.57.b}$$

by the triangle inequality, (4.51), (4.41), and the requirement that η be small. Also,

$$\mathrm{diam}\, f(B) \leq C\,\mathrm{dist}\,(f(t), A) \tag{4.58}$$

by (4.54) (with $i = 1$) and (4.57). Thus the hypotheses of Lemma 4.26 are satisfied with $X = f(B)$, and we conclude that

$$D_\omega(\xi, \zeta) \approx \operatorname{dist}(f(t), A)^s \, |\xi - \zeta| \,, \quad \text{when } \xi, \zeta \in f(B) \,. \tag{4.59}$$

To prove that f is bilipschitz on B as a map into $(\mathbb{R}^3, D_\omega)$ we should show that

$$\operatorname{dist}(f(t), A)^s \, |f(x) - f(y)| \approx |x - y| \,, \quad \text{when } x, y \in B \,. \tag{4.60}$$

To do this we shall approximate f on B using Taylor's theorem.

Define $\psi(x)$ by $\psi(x) = f(t) + F(t)\,(x - t)$. This is the linear Taylor approximation to f at t. (See (4.46) and Sublemma 4.44.) Because $F(t)$ is a similarity with dilation factor $|f'(t)|$ we have that

$$|\psi(x) - \psi(y)| = |f'(t)| \, |x - y| \approx \operatorname{dist}(f(t), A)^{-s} \, |x - y| \,, \tag{4.61}$$

because of (4.43). We can control $f - \psi$ using (4.54) and Taylor's theorem. In fact we are really interested in $\nabla(f - \psi)$, and we have that

$$\begin{aligned} \sup_B |\nabla f - \nabla \psi| &\leq C \,(\sup_B |\nabla^2 f|) \operatorname{diam} B \\ &\leq C \operatorname{dist}(f(t), A) \operatorname{dist}(t, \{z, w\})^{-2} \eta \, r \,, \end{aligned} \tag{4.62}$$

because of (4.54), (4.57), and the definition of B. We can simplify this further using (4.41) and the definition of r to get

$$\sup_B |\nabla f - \nabla \psi| \leq C \, \eta \operatorname{dist}(f(t), A)^{-s} \,. \tag{4.63}$$

This means that

$$|(f - \psi)(x) - (f - \psi)(y)| \leq C \, \eta \operatorname{dist}(f(t), A)^{-s} \, |x - y| \tag{4.64}$$

when $x, y \in B$. If η is small enough then we get (4.60) from (4.61) and (4.64). Choose η small enough so that this is true (*i.e.*, the bilipschitzness of f on the ball $B = B(t)$), and let it be fixed from now on.

Suppose now that we are given any $x, y \in S_\varepsilon(z, w)$, and let us estimate $D_\omega(f(x), f(y))$ from below. If x and y both lie in a ball of the form $B(t)$ as above then we already have the estimate that we need from (4.60), and so we may assume that this is not the case. Note that $u \in B(\pi(u))$ for all $u \in S_\varepsilon(z, w)$ if ε is small enough (compared to η). If

$|x-y|$ is small compared to $\eta \operatorname{dist}(x,\{z,w\})$ then $y \in B(\pi(x))$ (if also ε is small enough), and we are assuming that this is not true. We are making the same assumption with the roles of x and y reversed, and so we obtain that

$$(4.65) \qquad |x-y| \geq C^{-1}\eta \max\{\operatorname{dist}(x,\{z,w\}), \operatorname{dist}(y,\{z,w\})),$$

for some constant C. (The η in (4.65) does not matter, since it is fixed now anyway.)

We need to estimate $D_\omega(f(x), f(y))$ from below, and we shall do this in terms of $D_\omega(f(\pi(x)), f(\pi(y)))$. We already know that f is bilipschitz on $S(z,w)$, and so

$$(4.66) \qquad D_\omega(f(\pi(x)), f(\pi(y))) \geq C^{-1}|\pi(x)-\pi(y)|.$$

This implies that

$$(4.67) \qquad D_\omega(f(\pi(x)), f(\pi(y))) \geq C^{-1}|x-y|$$

when ε is small enough, because (4.65) and (4.47) yield $|x-\pi(x)| + |y-\pi(y)| \leq C\eta^{-1}\varepsilon|x-y|$. To get from here to $D_\omega(f(x), f(y))$ we have to control some error terms. Since f is Lipschitz as a map from $(S_\varepsilon(z,w), |x-y|)$ into $(\mathbb{R}^3, D_\omega)$ we have that

$$(4.68) \qquad D_\omega(f(x), f(\pi(x))) \leq C|x-\pi(x)| \leq C\varepsilon \operatorname{dist}(x,\{z,w\}).$$

The same estimate is true with x replaced by y. These estimates combined with (4.65) and (4.67) give the desired

$$(4.69) \qquad D_\omega(f(x), f(y)) \geq C^{-1}|x-y|$$

when ε is small enough.

This completes the proof of Sublemma 4.50, and Lemma 4.31 follows. The remaining cases involve similar constructions of connecting curves and their tubular neighborhoods, and we shall treat them in less detail.

Lemma 4.70. *If $p, q \in \tau(\Gamma_0)$, then the conclusion of Proposition* 4.25 *holds.*

Again choose δ in some A_m so that $p, q \in \tau(\delta)$ and m is as large as possible. We may as well assume that one of p and q lies in $\tau(N^2(\delta))$, since otherwise we can apply Lemma 4.30. This implies that $|p - q| \geq C^{-1}\text{diam}\,\delta$ for some constant C (which depends only on the parameter k from Section 3): if $|p - q|$ were small compared to $\text{diam}\,\delta$, then we could use the fact that one of p and q lies in $\tau(N^2(\delta))$ to conclude that $p, q \in \tau(\gamma)$ for some child $\gamma \in A_{m+1}$ of δ, in contradiction to the maximality of m.

Under these conditions we can apply the same basic construction as in the proof of Lemma 4.31. The difference is that now one or both of p and q may not lie in A, so that the sequences of α_l's and β_l's might stop in a finite number of steps. In fact, we could have that one of p or q lies in $\tau(\delta) \setminus \tau(N(\delta))$, so that there would be no α_l's, or no β_l's. Thus it may be necessary to make some adjustments to the construction at one or both of the "ends", but the estimates and underlying principles remain the same. (We use Sublemma 4.38 to find nice curves, we connect them as in Sublemma 4.39, we extend out to little neighborhoods of the curves as in (4.46), and we conclude as in Sublemma 4.50.) The details are left to the reader.

Lemmas 4.30, 4.31, and 4.70 cover all the possible locations of p and q except for $p \in \tau(N(\Gamma_0))$ and $q \in \mathbb{R}^3 \setminus \tau(\Gamma_0)$ (or the other way around). In this case we have that $|p-q|$ is bounded below by some fixed constant. This situation also lends itself to the same basic construction in Lemma 4.31. That is, we set $m = 0$, $\delta = \Gamma_0$, and we define α_l's and p_l's as before, except that these sequences will stop after finitely many steps if $p \notin A$. We can then connect p to the boundary of $\tau(\Gamma_0)$ by a sequence of smooth curves in the various $\tau(\alpha_l) \setminus \tau(N(\alpha_l))$'s. Since q now lies in $\mathbb{R}^3 \setminus \tau(\Gamma_0)$, we do not have to go through any contortions to connect it to the boundary of $\tau(\Gamma_0)$ in a nice way, we can simply do it. We can then combine these two curves and fatten them up as before. More precisely, this means that we can build a map f from a set S_ε like $S_\varepsilon(z, w)$ in the proof of Lemma 4.31 into $\mathbb{R}^3$ such that the restriction of f to one end of S_ε provides a connection from p to the boundary of $\tau(\Gamma_0)$ and the restriction of f to the rest of S_ε provides a connection from there to q. Note that the proportion of S_ε which corresponds to p will be much smaller than half of S_ε when $|p-q|$ is very large, in which case the part of S_ε that goes from the boundary of $\tau(\Gamma_0)$ to q will have to have a big bulge in the middle. This does not cause a problem, but one should be careful to map the bulge away from A. The remaining details are much like those in the proof of Lemma 4.31, and we leave them to the reader.

This completes the proof of Proposition 4.25.

5. The proof of Theorem 1.3.

In the previous section we saw how to build strong A_∞ continuous weights on $\mathbb{R}^3$ such that the associated metric on $\mathbb{R}^3$ is not bilipschitz equivalent to the Euclidean metric. In this section we want to build subsets of $\mathbb{R}^n$ which have many of the same nice properties as $\mathbb{R}^3$ (say) without being bilipschitz equivalent to $\mathbb{R}^3$ (with the Euclidean metric). One way to do this would be to show that we can embed $(\mathbb{R}^3, D_\omega(x,y))$ bilipschitzly into $\mathbb{R}^n$ with the Euclidean metric. There is a general result in [Se4] which could be applied to give such a bilipschitz embedding for the weights constructed in the previous section. In this special case, however, such embeddings can be produced with much less fuss than the general construction in [Se4]. Alternatively, one can go to some trouble and build embeddings with especially nice properties. That is what we shall do here.

The construction that we shall make in this section will proceed along the following lines. The first step will be to build a set F which is analogous to the set A from Section 3 except that at each stage we use slightly smaller circles. These circles will not be linked, unlike their predecessors. Then we shall build a sequence of diffeomorphisms on $\mathbb{R}^4$ which send the various approximations to A to the corresponding approximations to F. This would be impossible in $\mathbb{R}^3$, because of the linking properties, but none of the circles are linked in $\mathbb{R}^4$, and it will be easy to build these mappings. In the limit we shall obtain a quasiconformal map on $\mathbb{R}^4$ which sends A to F and maps $\mathbb{R}^3$ to a reasonably well-behaved surface. This surface will however be bilipschitz equivalent to $(\mathbb{R}^3, D_\omega(x,y))$ with ω as in (4.1) for a suitable choice of s, and so we shall be able to choose the parameters in such a way that it is not bilipschitz equivalent to $\mathbb{R}^3$ with the Euclidean metric.

Thus we need to begin by extending the construction in Section 3, and in the following we use the notation, assumptions, and results of Section 3 freely.

Let $\mu \in (0,1)$ be fixed but arbitrary. It will correspond eventually to the parameter s in (4.1). The parameter k from Section 3 should also be treated as fixed and will play the same role as before. However, for this section it will be convenient to assume that k is a little larger, and so we require that $k \geq 10^{10}$.

Our first task is to choose some circles $\beta_0(p)$, $p \in P_0$, which will be cousins to the γ_0's in Section 3. Each $\beta_0(p)$ should be a circle in $\mathbb{R}^3$

centered at p with radius $\mu\rho(k)$, and we require also that

$$\text{(5.1)} \qquad \operatorname{dist}(\{p\} \cup \beta_0(p), \{q\} \cup \beta_0(q)) \geq (100\,k)^{-1} \quad \text{when } p \neq q\,.$$

Fix any circles with these properties. When μ is small we can simply take $\beta_0(p)$ to be the circle centered at p which is obtained from $\gamma_0(p)$ by the obvious dilation, but when μ is closer to 1 this will not work (because the circles could touch) and so the circles should be tilted a little (or a lot, for that matter). To simplify the discussion let us simply require that the $\beta_0(p)$'s be unlinked, which in fact they must be when μ is small enough.

In short, the $\beta_0(p)$'s are a bunch of little circles centered on Γ_0 and placed at regular intervals. Think of [Mo, Figure 18.1, p. 127], but with the circles being smaller and unlinked. If we wanted we could impose some symmetry conditions, but we shall not bother.

Next, fix markings ψ_p for the $\beta_0(p)$'s. It does not matter how the markings are selected, but they need to be fixed forever.

Take $M(\Gamma_0)$ to be the union of the $\beta_0(p)$'s, $p \in P_0$, but viewed as a union of marked circles. We can define $M(\Gamma)$ for any marked circle Γ in $\mathbb{R}^3$ just as in (3.5), so that $M(\Gamma)$ is again a union of marked circles, and these circles are labelled by P_0 in the obvious way. As in Section 3 we define $M(E)$ when E is a finite union of marked circles, so that M defines a mapping on the space $\mathcal{C}$ of finite unions of marked circles in $\mathbb{R}^3$ with the same kinds of operational properties as N has (with respect to unions and the action of orientation-preserving similarities).

Define $F_l \in \mathcal{C}$, $l \geq 0$, in the same way that the A_l's were before, but using M now instead of N. That is, we set $F_0 = \Gamma_0$, $F_1 = M(\Gamma_0)$, and $F_l = M^l(\Gamma_0)$, where M^l denotes the l^{th} power of M, viewed as a mapping on $\mathcal{C}$. Each F_l is a union of k^l marked circles of radius $(\mu\rho(k))^l$, and F_{l+1} is the union of the $M(\alpha)$'s, where α runs through the circles which make up F_l. The Hausdorff limit of the F_l's will give a Cantor set which is not wild.

Most of the lemmas in Section 3 apply to F and M as well, and we summarize them in the following.

Scholium 5.2. a) *For each j, l with $0 \leq j \leq l$ we have that $F_l = \cup_\alpha M^{l-j}(\alpha)$, where the union is taken over all the constituent (marked) circles α in F_j, and the constituent circles of these two collections are marked in the same way.*

b) *For any marked circle Γ in $\mathbb{R}^3$ we have that* $\sup_{x \in M(\Gamma)} \operatorname{dist}(x, \Gamma) \leq \mu\,\rho(k)\operatorname{radius}(\Gamma) < 10^{-6}\operatorname{radius}(\Gamma)$.

c) $\tau(M_l(\Gamma)) \subseteq \tau(\Gamma)$ *for all marked cirles* Γ *and all* $l \geq 0$.

d) $\tau(F_j) \supseteq \tau(F_l)$ *when* $0 \leq j \leq l$.

e) *Let* α *and* α' *be two distinct circles among those which make up* F_l, $l \geq 1$. *Then* $\tau_{50}(\alpha)$ *and* $\tau_{50}(\alpha')$ *are disjoint.*

This is proved in exactly the same way as Lemmas 3.7, 3.8, 3.10, 3.11, and 3.14. The point is that the $\beta_0(p)$'s have the same properties as the $\gamma_0(p)$'s, except that they are smaller and unlinked, and these changes do not matter for these statements. For instance, the combinatorics in a) are exactly the same as in Lemma 3.7, while b) and c) are no more than simple applications of the triangle inequality.

Let us now proceed to $\mathbb{R}^4$. From now on we shall identify $\mathbb{R}^3$ with the $x_4 = 0$ hyperplane in $\mathbb{R}^4$, so that all of our constructions (A, F, etc.) can be viewed as living also in $\mathbb{R}^4$. Notice that every orientation-preserving similarity on $\mathbb{R}^3$ has a unique extension to an orientation-preserving similarity on $\mathbb{R}^4$, and so we can view all such transformations as acting on $\mathbb{R}^4$. In particular the similarities which provide markings for our circles will be viewed as acting on all of $\mathbb{R}^4$.

Given a circle Γ in $\mathbb{R}^4$ and $a > 0$ set

$$T(\Gamma) = \{x \in \mathbb{R}^4 : \operatorname{dist}(x, \Gamma) \leq 10^{-5} \operatorname{radius} \Gamma\}, \tag{5.3}$$

$$T_a(\Gamma) = \{x \in \mathbb{R}^4 : \operatorname{dist}(x, \Gamma) \leq a\, 10^{-5} \operatorname{radius} \Gamma\}. \tag{5.4}$$

These are the 4-dimensional versions of (3.9) and (3.13) in $\mathbb{R}^4$, and they enjoy properties analogous to those for $\tau(\Gamma)$ and $\tau_a(\Gamma)$. If E is a finite union of circles, then we define $T(E)$ and $T_a(E)$ to be the union of the sets obtained by applying T or T_a to the constituent circles.

Lemma 5.5. a) *If* Γ *is a marked circle in* $\mathbb{R}^3$, *then* $T(N_l(\Gamma)) \subseteq T(\Gamma)$ *and* $T(M_l(\Gamma)) \subseteq T(\Gamma)$ *for all* $l \geq 0$.

b) $T(A_j) \supseteq T(A_l)$ *and* $T(F_j) \supseteq T(F_l)$ *when* $0 \leq j \leq l$.

c) *Let* α *and* α' *be two distinct circles among those which make up* A_l *(or* F_l*),* $l \geq 1$. *Then* $T_{50}(\alpha)$ *and* $T_{50}(\alpha')$ *are disjoint.*

This is proved in the same way as for the analogous results for τ. (Nothing special about $\mathbb{R}^3$ was used; it all came down to the triangle inequality.)

Our next task is to build a homeomorphism from $\mathbb{R}^4$ to itself which sends A to F and which is otherwise as nice as possible. This mapping will have to shrink distances with some severity near A, but away from A it will be a diffeomorphism. We shall produce this mapping through an iterative process, and we first need to construct some building blocks.

Lemma 5.6. *There is a smooth diffeomorphism $\Phi_0 : \mathbb{R}^4 \to \mathbb{R}^4$ such that $\Phi_0(T(\Gamma_0)) = T(\Gamma_0)$, Φ_0 equals the identity on the complement of $T(\Gamma_0)$ and on a neighborhood of $\partial T(\Gamma_0)$, and the restriction of Φ_0 to a neighborhood of $T(\gamma_0(p))$ is, for each $p \in P_0$, the orientation-preserving similarity from $T(\gamma_0(p))$ onto $T(\beta_0(p))$ which is determined by their markings.*

For this lemma it is crucial that we are working in $\mathbb{R}^4$ instead of $\mathbb{R}^3$. In $\mathbb{R}^3$ the $\gamma_0(p)$'s are linked, while the $\beta_0(p)$'s are not; in $\mathbb{R}^4$, none of them are linked.

We shall obtain Φ_0 by composing a finite number of simpler pieces. It will be convenient to use also some auxiliary circles. For each $p \in P_0$ choose a circle $\delta_0(p)$ in $\mathbb{R}^3$ which is centered at p and which has radius $(10^4k)^{-1}$. These circles should also be given markings. The specific choices do not matter.

Sublemma 5.7. *For each $p \in P_0$ there is a diffeomorphism $f : \mathbb{R}^4 \to \mathbb{R}^4$ such that f equals the identity on $\mathbb{R}^4 \setminus T(\Gamma_0)$, on a neighborhood of $\partial T(\Gamma_0)$, and on neighborhoods of $T(\gamma_0(q))$ and $T(\delta_0(q))$ for each $q \in P_0 \setminus \{p\}$, and such that the restriction of f to a neighborhood of $T(\gamma_0(p))$ agrees with the orientation-preserving similarity which takes $\gamma_0(p)$ to $\delta_0(p)$ and which is determined by their markings. The analogous result holds for the β_0's instead of the γ_0's.*

Lemma 5.6 will follow once we have proved Sublemma 5.7. Indeed, if Sublemma 5.7 is true, then we can build a nice map on $\mathbb{R}^4$ which is the identity outside $T(\Gamma_0)$ and which sends $T(\gamma_0(p))$ to $T(\delta_0(p))$ for each $p \in P_0$ in the right way (in accordance with the markings), simply by composing the various pieces from Sublemma 5.7. We can then produce a similar map for the β_0's instead of the γ_0's, and Φ_0 is obtained by composing the previous map with the inverse of the second one.

Let us prove Sublemma 5.7. Let $p \in P_0$ be given. We want to pick up $T(\gamma_0(p))$ off the "floor" $\mathbb{R}^3$, shrink it and turn it around as necessary, and lay it down on $T(\delta_0(p))$ without disturbing any of the other $T(\gamma_0(q))$'s or $T(\delta_0(q))$'s or the complement of $T(\Gamma_0)$. This is very simple geometrically (or physically), and the following claim makes it precise.

Claim 5.8. *There is a smooth 1-parameter family of orientation-preserving similarities σ_t, $0 \leq t \leq 1$, on $\mathbb{R}^4$ such that σ_0 is the identity, σ_1 is the similarity which sends $\gamma_0(p)$ to $\delta_0(p)$ that is determined by the markings, and the compact set $\cup_{0\leq t\leq 1}\sigma_t(T(\gamma_0(p)))$ lies in the interior of $T(\Gamma_0) \setminus \cup_{q\in P_0\setminus\{p\}}\big(T(\gamma_0(q)) \cup T(\delta_0(q))\big)$.*

To see this, let H denote the $x_4 = 10^{-7}$ hyperplane. In the beginning (near $t = 0$) σ_t should simply make a translation in the x_4 direction by 10^{-7}, so that $\mathbb{R}^3$ is sent onto H. For the next period of time σ_t should preserve H while deforming the translate of $\gamma_0(p)$ in H into the translate of $\delta_0(p)$ in H. At the end σ_t should simply make a translation by -10^{-7} in the x_4 direction. If we do the deformation in H in the middle stage correctly then at the end we shall obtain the correct choice of σ_1. Because we are requiring that $k \geq 10^{10}$ in this section (so that $\text{radius}(\beta_0(p)) \leq \text{radius}(\gamma_0(p)) \leq 10^{-9}$, by (3.1)), we obtain easily that $\cup_{0\leq t\leq 1}\sigma_t(T(\gamma_0(p)))$ lies inside $T(\Gamma_0)$. It also remains disjoint from $\cup_{q\in P_0\setminus\{p\}}\big(T(\gamma_0(q)) \cup T(\beta_0(q))\big)$. Indeed, in the first and final stages of the motion σ_t (when we are simply translating in the x_4 direction) this follows easily from (3.2) and (5.1), while in the middle stage we are moving around in H, and H is too far from $\mathbb{R}^3 = \{x_4 = 0\}$ to cause any problems, since $k \geq 10^{10}$. (Keep in mind that $\gamma_0(q)$'s, $\beta_0(q)$'s, and $\delta_0(q)$'s lie in $\mathbb{R}^3$.) This proves the claim.

Next we want to extend the motion in Claim 5.8 to all of $\mathbb{R}^4$ with suitable properties. We use a standard argument for extending smooth isotopies, *i.e.*, we convert the problem to one of extending vector fields.

Claim 5.9. *There is a smooth $\mathbb{R}^4$-valued function $V(x,t)$ on $[0,1] \times \mathbb{R}^4$ such that $V(x,t) = 0$ for all t when $x \in \mathbb{R}^4 \setminus T(\Gamma_0)$ or x lies in a neighborhood of $\partial T(\Gamma_0)$ or of $\cup_{q\in P_0\setminus\{p\}}\big(T(\gamma_0(q)) \cup T(\delta_0(q))\big)$, and $V(\sigma_t(x),t) = (d\sigma_t/dt)(x)$ for all t when x lies in a neighborhood of $T(\gamma_0(p))$.*

This is an immediate consequence of Claim 5.8 and standard facts about smooth functions.

We are now almost finished with the proof of Sublemma 5.7. Let f_t be the flow associated to $V(x,t)$, so that f_0 is the identity and $(df_t/dt)(x) = V(f_t(x),t)$ for all x and t. Then $f_t(x) = x$ for all t when x lies in $\mathbb{R}^4 \setminus T(\Gamma_0)$ or a neighborhood of $\partial T(\Gamma_0)$ or a neighborhood of $\cup_{q \in P_0 \setminus \{p\}} \big(T(\gamma_0(q)) \cup T(\delta_0(q))\big)$, and $f_t(x) = \sigma_t(x)$ for all t when x lies in a neighborhood of $T(\gamma_0(p))$, by the uniqueness theorem for ordinary differential equations. Of course the f_t's are diffeomorphisms also, and so $f = f_1$ has all the required properties. This proves Sublemma 5.7.

Now that Lemma 5.6 is established we want to build a more complicated homeomorphism by piecing together many copies of Φ_0. This mapping will take A to F and we shall try to make it as nice as possible off of A. To do this we need some more notation and definitions.

Let $\mathcal{A}_l$ and $\mathcal{F}_l$ denote the collections of k^l marked circles which make up A_l and F_l, respectively. Let $\mathcal{S}_l$ denote the Cartesian product of l copies of P_0, so that elements of $\mathcal{S}_l$ are finite sequences of length l with entries in P_0. There are natural bijections from $\mathcal{S}_l$ onto each of $\mathcal{A}_l$ and $\mathcal{F}_l$ which code the history of the circles. If $\alpha \in \mathcal{A}_l$ and α' is its parent in $\mathcal{A}_{l-1}$, so that $\alpha \in N(\alpha')$, then the sequence in $\mathcal{S}_l$ associated to α is just the sequence in $\mathcal{S}_{l-1}$ associated to α' together with one additional element of P_0 at the end to specify the position of α relative to α'. (See (3.5).) These codings of $\mathcal{A}_l$ and $\mathcal{F}_l$ induce a natural (bijective) correspondence $\theta_l : \mathcal{A}_l \to \mathcal{F}_l$.

Let $\mathcal{S}$ denote the Cartesian product of countably many copies of P_0, so that the elements of $\mathcal{S}$ are sequences $\{q_j\}_{j \geq 1}$ which take values in P_0. There are natural bijections from $\mathcal{S}$ onto A and F which take a sequence of elements of P_0 and assign to it the point in A or F with that history. Alternatively, these points could be described as the limits of the circles in $\mathcal{A}_l$ and $\mathcal{F}_l$ corresponding to the initial l terms in the sequence in $\mathcal{S}$. These bijections induce a bijection $\theta : A \to F$. The homeomorphism on $\mathbb{R}^4$ that we are going to construct will be an extension of θ, and it will (in a certain sense) respect the codings provided by the θ_l's on the complement of A.

Given a marked circle Γ in $\mathbb{R}^3$, let us associate to it two compact sets $X(\Gamma)$ and $Y(\Gamma)$ by taking $X(\Gamma)$ to be the closure of $T(\Gamma) \setminus T(N(\Gamma))$ and $Y(\Gamma)$ to be the closure of $T(\Gamma) \setminus T(M(\Gamma))$. These sets are solid tori with k smaller solid tori removed. We are going to break up $\mathbb{R}^4$ into pieces using the $X(\Gamma)$'s and $Y(\Gamma)$'s and define our eventual homeomorphism initially on these various pieces before gluing them together. Notice that $\Phi_0(X(\Gamma_0)) = Y(\Gamma_0)$.

If β is a marked circle in $\mathbb{R}^3$, let ϕ_β denote the similarity which provides the marking. (This will not be confused with ϕ_p, which provides the marking of $\gamma_0(p)$, because p is not a circle.) Define Φ_β to be the composition $\phi_\beta \circ \Phi_0 \circ (\phi_\beta)^{-1}$. In other words this is a copy of Φ_0 which lives near β instead of Γ_0. Thus Φ_β has the following properties: it is a diffeomorphism on $\mathbb{R}^4$ which is the identity on the complement of $T(\beta)$ and on a neighborhood of $\partial T(\beta)$ and which takes $T(\beta)$ to itself; it sends $N(\beta)$ to $M(\beta)$, and it preserves the labellings of the circles in $N(\beta)$ to $M(\beta)$ by P_0; its restriction to a neighborhood of $T(\gamma)$ is an orientation-preserving similarity for each circle γ in $N(\beta)$, and this similarity is the one determined by the markings (and implicitly the labellings of the circles by P_0 also); and Φ_β sends $X(\beta)$ onto $Y(\beta)$. These properties are all easy consequences of the analogous statements for Φ_0 and Γ_0.

Given $\alpha \in \mathcal{A}_l$ let ξ_α denote the orientation-preserving similarity which takes α to $\beta = \theta_l(\alpha) \in \mathcal{F}_l$ and which is the one determined by the markings. Set $\Psi_\alpha = \Phi_\beta \circ \xi_\alpha$. The Ψ_α's will provide the building blocks for the homeomorphism that we want to build. Let us summarize some of their important properties in a lemma.

Lemma 5.10. *Suppose that $\alpha \in \mathcal{A}_l$ and $\beta = \theta_l(\alpha) \in \mathcal{F}_l$. Ψ_α has the following properties: it is a diffeomorphism on $\mathbb{R}^4$ which sends $T(\alpha)$ to $T(\beta)$; it agrees with ξ_α outside $T(\alpha)$ and on a neighborhood of $\partial T(\alpha)$; it maps $N(\alpha)$ to $M(\beta)$; if γ is one of the circles in $N(\alpha)$, then $\Psi_\alpha(\gamma)$ is the same as the circle $\delta = \theta_{l+1}(\gamma) \in \mathcal{F}_l$, and the restriction of Ψ_α to a neighborhood of $T(\gamma)$ agrees with the orientation-preserving similarity that takes γ to δ and is determined by the markings (i.e., ξ_γ); and $\Psi_\alpha(X(\alpha)) = Y(\beta)$.*

These properties are all easy to verify from the definitions and Lemma 5.6.

Let $\mathcal{A}$ and $\mathcal{F}$ denote the union of all the $\mathcal{A}_l$'s and $\mathcal{F}_l$'s for $l = 0, 1, \ldots$ The homeomorphism H that we really want is defined as follows:

$$H = \begin{cases} \text{the identity}, & \text{on } \mathbb{R}^4 \setminus T(\Gamma_0)\,, \\ \Psi_\alpha\,, & \text{on } X(\alpha) \text{ for } \alpha \in \mathcal{A}_l\,, \\ \theta\,, & \text{on } A\,. \end{cases} \tag{5.11}$$

We need to check that this is well defined, etc.

Lemma 5.12. *$\mathbb{R}^4$ is the union of $\mathbb{R}^4 \setminus T(\Gamma_0)$, the sets $X(\alpha)$ for $\alpha \in \mathcal{A}$, and A. $\mathbb{R}^4 \setminus T(\Gamma_0)$ is disjoint from all the $X(\alpha)$'s except $X(\Gamma_0)$, and $X(\alpha)$, $X(\alpha')$ intersect, $\alpha, \alpha' \in \mathcal{A}$, if and only if one of α and α' is the parent of the other, in which case $X(\alpha)$ and $X(\alpha')$ intersect only in a component of the boundary. Neither $\mathbb{R}^4 \setminus T(\Gamma_0)$ nor any of the $X(\alpha)$'s contain any elements of A. The analogous statements for $Y(\beta)$, $\beta \in \mathcal{F}$, and F are also true.*

This follows easily from Lemma 5.5, the definitions of A and F, etc. The main point is that if $\alpha, \alpha' \in \mathcal{A}$, then either one of α and α' is an ancestor of the other, say α' is an ancestor of α, in which case $T(\alpha) \subseteq T(\alpha')$, or $T(\alpha)$ and $T(\alpha')$ are disjoint (because they have distinct common ancestors in some $\mathcal{A}_l$).

Lemma 5.13. *H is well defined and smooth off A.*

For instance, we have taken H to be the identity on the complement of $T(\Gamma_0)$ and to be Ψ_{Γ_0} on $X(\Gamma_0)$, and the two share the torus $\partial T(\Gamma_0)$ as part of their boundaries. However, $\Psi_{\Gamma_0} = \Phi_0$ by definitions, and so we really have $H = \Phi_0$ on the union of $\mathbb{R}^4 \setminus T(\Gamma_0)$ and $X(\Gamma_0)$.

Similarly, if $\alpha \in \mathcal{A}$, $\alpha \neq \Gamma_0$, and $\alpha' \in \mathcal{A}$ is its parent, then $X(\alpha')$ and $X(\alpha)$ have a torus as their common boundary (namely, $\partial T(\alpha)$). However, on a neighborhood of $\partial T(\alpha)$ both Ψ_α and $\Psi_{\alpha'}$ agree with ξ_α, and so H is smooth across $\partial T(\alpha)$.

Lemma 5.14. *For any α in any $\mathcal{A}_l$ we have that $H(X(\alpha)) = Y(\theta_l(\alpha))$ and $H(T(\alpha)) = T(\theta_l(\alpha))$.*

The first part about $X(\alpha))$ is an immediate consequence of the definition (5.11). For the second part we begin by observing that $T(\alpha)$ is the union of $X(\gamma)$ over all $\gamma \in \mathcal{A}$ descended from α (including α itself) together with $A \cap T(\alpha)$, and that the analogous statement holds for a $\beta \in \mathcal{F}$, but with $Y(\cdot)$ instead of $X(\cdot)$. Next we observe that $H(A \cap T(\alpha)) = F \cap T(\theta_l(\alpha))$. Indeed, Lemma 5.5 implies that $A \cap T(\alpha)$ consists precisely of the points in A which are "descended" from α, and similarly $F \cap T(\theta_l(\alpha))$ consists of the points in F which are descended from $\theta_l(\alpha)$, and the two correspond under H because H is defined to be the same as θ on A. The second part of the lemma follows from the first part and these observations.

Lemma 5.15. *H is continuous on all of $\mathbb{R}^4$.*

The continuity of H away from A follows from Lemma 5.13, while the continuity on A follows from Lemma 5.14.

Lemma 5.16. *H is a homeomorphism on $\mathbb{R}^4$, and it is a diffeomorphism from $\mathbb{R}^4 \setminus A$ onto $\mathbb{R}^4 \setminus F$.*

This follows from Lemmas 5.14 and 5.12, the fact that the individual Ψ_α's are diffeomorphisms, etc.

Next let us estimate the differential of H, which we denote by dH, and which is well defined off A. More precisely, if $x \in \mathbb{R}^4 \setminus A$, then dH_x will be used to denote the differential of H at x as a linear transformation.

Lemma 5.17. *Choose $s > 0$ so that $\mu = \rho(k)^s$, set $t = s(1+s)^{-1}$, and set $\lambda(x) = \min\{1, \operatorname{dist}(x, A)^s\}$ and $\nu(y) = \max\{1, \operatorname{dist}(y, F)^{-t}\}$. Then there is a constant C so that*

$$C^{-1}\,\lambda(x)\,|v| \le |dH_x(v)| \le C\,\lambda(x)\,|v|$$

and

$$C^{-1}\,\nu(y)\,|v| \le |dH_y^{-1}(v)| \le C\,\nu(y)\,|v|,$$

for all $x \in \mathbb{R}^4 \setminus A$, $y \in \mathbb{R}^4 \setminus F$, and $v \in \mathbb{R}^4$, where $|v|$ denotes the Euclidean norm of the vector v.

To see this we need to first reexpress λ and ν in more useful forms.

Sublemma 5.18. *$\lambda(x) \approx 1$ on $\mathbb{R}^4 \setminus T(\Gamma_0)$, and $\lambda(x) \approx \mu^l$ on $X(\alpha)$ for any $\alpha \in \mathcal{A}_l$, $l \ge 0$. Here $a \approx b$ means that each of a, b is bounded by a constant times the other. Similarly, $\nu(y) \approx 1$ on $\mathbb{R}^4 \setminus T(\Gamma_0)$ and $\nu(y) \approx \mu^{-l}$ on $Y(\beta)$ for any $\beta \in \mathcal{F}_l$, $l \ge 0$.*

That $\lambda(x) \approx 1$ and $\nu(y) \approx 1$ on $\mathbb{R}^4 \setminus T(\Gamma_0)$ simply reflects the fact that A and F lie in the interior of $T(\Gamma_0)$ (which follows from $A \subseteq \tau(N(\Gamma_0))$ and $F \subseteq \tau(M(\Gamma_0))$, for instance). Using the definitions of s and t the remaining parts come down to

$$\begin{aligned} &\operatorname{dist}(x, A) \approx \rho(k)^l, && \text{on } X(\alpha), \\ &\operatorname{dist}(y, F) \approx (\mu\,\rho(k))^l, && \text{on } Y(\beta), \end{aligned} \tag{5.19}$$

for $\alpha \in \mathcal{A}_l$ and $\beta \in \mathcal{F}_l$. These estimates are easy to check, using $\varnothing \neq A \cap T(\alpha) \subseteq T(N^2(\alpha))$, $\text{radius}(\alpha) = \rho(k)^l$, $\varnothing \neq F \cap T(\beta) \subseteq T(M^2(\beta))$, and $\text{radius}(\beta) = (\mu\,\rho(k))^l$. (Do not forget Lemma 3.10 and Scholium 5.2.c).) This proves the sublemma.

As for Lemma 5.17, notice first that it is true on $\mathbb{R}^4 \setminus T(\Gamma_0)$, since H equals the identity there. Fix now an α in some $\mathcal{A}_l$, so that H agrees with Ψ_α on $X(\alpha)$ by the definition (5.11). Let us check that

$$|d(\Psi_\alpha)_x(v)| \approx \mu^l\,|v|\,, \tag{5.20}$$

for all x and v. Recall that, by definition, $\Psi_\alpha = \Phi_\beta \circ \xi_\alpha = \phi_\beta \circ \Phi_0 \circ (\phi_\beta)^{-1} \circ \xi_\alpha$, where $\beta = \theta_l(\alpha)$. Φ_0 is a single diffeomorphism which equals the identity outside $T(\Gamma_0)$, and so its differential distorts the Euclidean norm only by a bounded factor. Since ϕ_β is a similarity, it distorts distances by the same constant factor everywhere, and so the presence of ϕ_β and its inverse cancel each other out. Thus we are left with ξ_α, which is a similarity which maps α to β. This means that the dilation factor of ξ_α is simply the ratio of the diameters of β and α. By construction the diameter of the former is $(\mu\,\rho(k))^l$ while the diameter of the latter is $\rho(k)^l$, and so we get (5.20). The required estimates on dH and dH^{-1} follow easily. (Do not forget Lemma 5.14.)

Lemma 5.21. *H is Lipschitz continuous and continuously differentiable on all of $\mathbb{R}^4$.*

We already know that H is smooth off A, but it is not hard to see that the differential of H exists and vanishes at points in A. This can derived from Lemma 5.14, for instance, and the fact that the radius of $\theta_l(\alpha)$ is μ^l times the radius of α for all $\alpha \in \mathcal{A}_l$. Lemma 5.21 follows from Lemma 5.17 and the boundedness of λ.

REMARK 5.22. Although H is C^1 everywhere, it is not a C^1 diffeomorphism across A, because its differential vanishes there. However, there is a simple way to approximate H by C^1 diffeomorphisms. Define H_m by $H_m =$ the identity on $\mathbb{R}^4 \setminus T(\Gamma_0)$, $H_m = \Psi_\alpha$ on $X(\alpha)$ for $\alpha \in \mathcal{A}_l$, $l < m$, and $H_m = \xi_\alpha$ on $T(\alpha)$ when $\alpha \in \mathcal{A}_m$. This is approximately the same as the definition (5.11) of H, except at levels m and below, where it is flattened out. These mappings satisfy suitable versions of the preceding lemmas, and in particular the differentials of the H_m^{-1}'s satisfy the same sort of estimates as in Lemma 5.17, uniformly in m.

This observation can be useful in making it easier to derive properties of H from the bounds on dH^{-1} (*e.g.*, when verifying the precise Sobolev space properties of H^{-1}).

Lemma 5.23. *H is quasiconformal.*

According to the usual definition this follows from Lemmas 5.16 and 5.17. However it is not hard to verify directly the a priori stronger condition that there is a $C > 0$ so that for all $x \in \mathbb{R}^4$ and $r > 0$ there is an $R > 0$ (which may depend on x and r) such that $B(H(x), R) \subseteq H(B(x,r)) \subseteq B(H(x), CR)$. To do this one considers separately the cases where $r \geq 1$, $r < 1$ and r is small compared to $\operatorname{dist}(x, A)$, and $r < 1$ but r is not small compared to $\operatorname{dist}(x, A)$. These cases can be treated by reducing to facts about the Ψ_α's (which are clearly uniformly quasiconformal) and Lemma 5.14.

It is not too difficult to describe completely the manner in which H distorts distances, but this is slightly gory and best left as an exercise. We should at least make the connection with the preceding section by formulating Lemma 5.17 (and some consequences of it) in terms of the metrics associated to strong A_∞ weights.

Define $\Omega(x)$ on $\mathbb{R}^4$ by $\Omega(x) = \lambda(x)^4$, and define $\omega(x)$ on $\mathbb{R}^3$ by $\omega = \lambda(x)^3$. Thus ω is the same as in (4.1), with s chosen as in Lemma 5.17. These are both strong A_∞ weights on their respective domains, by Propositions 4.3 and 4.4. This uses also the simple fact that A is uniformly disconnected as subset of $\mathbb{R}^4$, and not just $\mathbb{R}^3$. (In fact uniform disconnectedness is an intrinsic property of a metric space.) Let $D_\Omega(\cdot,\cdot)$ and $D_\omega(\cdot,\cdot)$ be the associated metrics on $\mathbb{R}^4$ and $\mathbb{R}^3$, respectively.

Lemma 5.24. *$D_\omega(x, y) = D_\Omega(x, y)$ for all $x, y \in \mathbb{R}^3$.*

This is not an accident, but it is sort of pleasant that we get an actual equality and not just an equivalence in size. Let $x, y \in \mathbb{R}^3$ be given, and let γ be a rectifiable curve in $\mathbb{R}^4$ which connects x to y. Let γ' denote the projection of γ to $\mathbb{R}^3$. One can compute that the Ω-length of γ is at least as big as the ω-length of γ', and that the two are equal if $\gamma \subseteq \mathbb{R}^3$ to begin with. (The first part comes down to the fact that $\operatorname{dist}(z, A)$ is decreased by projecting z onto $\mathbb{R}^3$, while the second is just a question of unwinding definitions.) This implies the lemma.

Lemma 5.25. *H is bilipschitz as a map from $(\mathbb{R}^4, D_\Omega(x,y))$ to $(\mathbb{R}^4, |x-y|)$, and as a map from $(\mathbb{R}^3, D_\omega(x,y))$ to $(E, |x-y|)$, where $E = H(\mathbb{R}^3)$.*

The first part is basically a reformulation of Lemma 5.17. Strictly speaking, the fact that H is Lipschitz as a map from $(\mathbb{R}^4, D_\Omega(x,y))$ to $(\mathbb{R}^4, |x-y|)$ is an immediate consequence of Lemmas 5.17 and 5.21, but one should be a little more careful about the Lipschitzness in the reverse direction. (Note however that Lemma 5.14 can be very useful in providing control near the singular points, so that one can concentrate on the smooth parts, which are more amenable to calculus. Alternatively one can approximate H by diffeomorphisms as in Remark 5.22 in order to reduce the problem to calculus.) The second part follows from the first and Lemma 5.24.

Note that Lemma 5.25 implies that for any pair of points $x, y \in E$ there is a closed subset of E containing x and y which is bilipschitz equivalent to a closed Euclidean 3-ball, with a uniformly bounded bilipschitz constant, because of the corresponding property for $(\mathbb{R}^3, D_\omega(x,y))$ (Proposition 4.25).

Lemma 5.26. *E is a regular set of dimension 3 (as in Definition 1.1).*

This can be derived from the second part of Lemma 5.25 and the corresponding general fact for strong A_∞ weights (1.7). Alternatively, one could go back to the definitions of $E, H, \dots$, and simply compute directly, using the fact that H is basically a uniform contraction by a known quantity on each $X(\alpha)$, and Lemma 5.14, etc.

At this stage we can read off Theorem 1.3 and some variants of it from Lemma 5.25, Proposition 4.25, and Theorem 4.20.

Theorem 5.27. *With the notation as above, $H : \mathbb{R}^4 \to \mathbb{R}^4$ is a quasiconformal mapping which is also Lipschitz continuous and $E = H(\mathbb{R}^3)$ is a 3-dimensional regular set when $k \geq 10^{10}$ and $\mu \in (0,1)$. Every pair of points in E is contained in a closed subset of E which is bilipschitz equivalent to a closed Euclidean 3-ball, with a uniformly bounded bilipschitz constant (which depends on k and μ.) Define $s > 0$ by $\mu = \rho(k)^s$, and let U be a (relatively) open subset of E which intersects $F = H(A)$, and let V be any open subset of $\mathbb{R}^3$. (For instance take $U = E, V = \mathbb{R}^3$.)*

a) *If $s > 2$ then there does not exist a homeomorphism $h : (U, |x -$*

$y|) \to (V, |x-y|)$ which is locally Lipschitz.

b) *For any $s > 0$ there does not exist a homeomorphism $h : (U, |x-y|) \to (V, |x-y|)$ which is locally Lipschitz if k is large enough.*

c) *For any $s > 0$ there does not exist a homeomorphism $h : (U, |x-y|) \to (V, |x-y|)$ which is locally Hölder continuous of order greater than $3(1+s)^{-1}$.*

The Euclidean metric is made crudely explicit in a)-c) above to avoid confusion with the other metrics which have been used.

Notice that any value of $s > 0$ can occur by choosing $\mu \in (0,1)$ correctly.

The remarks that follow Theorem 4.20 have counterparts in this setting too. Quasiconformal mappings and their inverses are always Hölder continuous, but part c) of Theorem 5.27 shows that any homeomorphism from E to $\mathbb{R}^3$ must have small Hölder exponent if s is large enough. Thus E can be chosen so that it is very far from being bilipschitz equivalent to $\mathbb{R}^3$. On the other hand, we can take s to be as small as we like as long as we take k to be large enough, in such a way that we can make E be extremely close to being bilipschitz equivalent to $\mathbb{R}^3$ without actually being bilipschitz equivalent. For instance, H^{-1} will be locally Hölder continuous of order as close to 1 as we like if we take s small enough. There are similar statements in terms of Sobolev spaces (using Lemma 5.17).

REMARK 5.28. Suppose that we defined A in the same way as in Section 3 except that we replaced (3.3) with the requirement that the $\gamma_0(p)$'s be unlinked. Then $(\mathbb{R}^3, D_\omega(x,y))$ would be bilipschitz equivalent to $(\mathbb{R}^3, |x-y|)$, no matter how large s is. This follows from the same construction as above. The only place in the preceding construction where we really needed to be in $\mathbb{R}^4$ was in the proof of Lemma 5.6, but if the $\gamma_0(p)$'s are unlinked then Lemma 5.6 works in $\mathbb{R}^3$.

6. Another interesting class of strong-A_∞ weights.

In this section we shall give another construction of strong A_∞ weights which can be viewed as a more refined version of Proposition 4.4 and from which Theorem 1.15 will be derived. We begin with a definition.

Definition 6.1. *A closed set $E \subseteq \mathbb{R}^n$ will be called a snowflake of order s, $s > 1$, if there is a metric $\rho(x,y)$ on E and a constant $C > 0$ such that*

$$C^{-1}\,|x-y|^s \le \rho(x,y) \le C\,|x-y|^s\,, \tag{6.2}$$

for all $x, y \in E$.

In this definition a metric simply means a nonnegative symmetric function which satisfies the triangle inequality, etc.

The usual Von Koch snowflake curve in $\mathbf{R}^2$ is a snowflake of order s for a computable value of s. In general we can use Assouad's Theorem 1.14 to generate plenty of snowflakes in the sense of Definition 6.1, with the order s given by α^{-1}, where α is as in Theorem 1.14. One can show that uniformly disconnected sets (in the sense of Definition 4.2) are snowflakes of order s for all $s > 1$. On the other hand, no snowflake of order $s > 1$ can contain a nonconstant rectifiable curve, for if $\gamma(t)$, $0 \le t \le 1$, were such a curve, and if $\rho(x,y)$ is as in Definition 6.1, then $\rho(\gamma(0),\gamma(1))$ would have to be zero, because of the assumptions on ρ. (Think about $\sum \rho(\gamma(t_i),\gamma(t_{i+1}))$ for partitions $\{t_i\}$ of $[0,1]$ of small mesh.)

Theorem 6.3. *Let $E \subseteq \mathbb{R}^n$ be a snowflake of order s. Then $\omega(x) = \operatorname{dist}(x,E)^{n(s-1)}$ is a strong A_∞ continuous weight on $\mathbb{R}^n$ such that*

$$C^{-1}\,D_\omega(x,y) \le |x-y|^s \le C\,D_\omega(x,y)\,, \tag{6.4}$$

for some C and all $x, y \in E$, where D_ω is as in Definition 1.5. (*Actually,* (6.4) *holds as soon as one of x or y lies in E.*)

In other words, if ρ is as in Definition 6.1, then we can build our strong A_∞ weight ω in such a way that D_ω is comparable to ρ on E. This kind of extension principle works in much greater generality, *i.e.*, we take a distance function on a set E with certain properties, and we build a strong A_∞ weight on all of $\mathbb{R}^n$ whose associated distance function essentially gives this distance function back again on E. For the present purposes however this more modest result is adequate and it has the nice feature of being much simpler.

Theorem 1.15 will follow once we prove Theorem 6.3. Indeed, if $(M, d(x,y))$ is a metric space which satisfies a doubling condition, then Assouad's Theorem 1.14 implies that $(M, d(x,y))^\alpha$ is bilipschitz equivalent to some subset of some $\mathbb{R}^n$ for any given value of $\alpha \in (0,1)$. The closure of this subset is an s-snowflake with $s = \alpha^{-1}$ (with ρ coming from $d(\cdot,\cdot)$), and the continuous weight ω provided by Theorem 6.3 satisfies the requirements of Theorem 1.15.

The rest of this section will be devoted to the proof of Theorem 6.3. From now on let E and ω be as in Theorem 6.3, let $\rho(\cdot,\cdot)$ be as in Definition 6.1, and let $D_\omega(\cdot,\cdot)$, the ω-length, and $B_{x,y}$ be defined as in Definition 1.5. Set $\delta_\omega(x,y) = \omega(B_{x,y})^{1/n}$, as in Section 1. The following will be the main step.

Lemma 6.5. *There is a constant C so that $D_\omega(x,y) \geq C^{-1}|x-y|^s$ for all $x, y \in \mathbb{R}^n$.*

Let $x, y \in \mathbb{R}^n$ be given, and let Γ be a curve in $\mathbb{R}^n$ that joins them. More precisely, Γ should be a continuous map from $[0, L]$ into $\mathbb{R}^n$ for some L with $\Gamma(0) = x$ and $\Gamma(1) = y$, and we shall assume that Γ is parameterized by arclength, so that Γ is Lipschitz continuous with norm 1 and $|x - y| \leq L$ in particular. We want to show that the ω-length of Γ is bounded from below by $C^{-1}|x-y|^s$. The idea is that this is easy when Γ stays far away from E, where ω is large, and that we can use our snowflake condition to get estimates when Γ gets close to E. To make this precise we need to break up Γ into simpler pieces that stay away from E and then recombine them.

Let Δ denote the collection of closed subintervals of $[0, L]$ which are dyadic with respect to $[0, L]$. That is, if $L = 1$ these intervals are dyadic in the usual sense, but in general they are not quite the usual dyadic intervals because we take $[0, L]$ itself, its two halves, the two halves of those, etc. These dyadic intervals are all of the form $[k\,2^{-j}L, (k+1)\,2^{-j}L]$ for some nonnegative integers j and k.

Set $F = \{t \in [0, L] : \Gamma(t) \in E\}$ and $U = [0, L] \setminus F$. Let $\mathcal{M}$ denote the collection of maximal dyadic subintervals I of $[0, L]$ such that

$$10\,|I| \leq \inf_{t \in I} \operatorname{dist}(\Gamma(t), E)\,, \tag{6.6}$$

where $|I|$ denotes the length of the interval I. By standard reasoning U is the union of the elements of $\mathcal{M}$, and two distinct intervals in $\mathcal{M}$ are either disjoint or intersect in one of their common endpoints.

Sublemma 6.7. $|x-y|^s \leq C\,\omega\text{-length}(\Gamma)$ *when* $[0, L] \in \mathcal{M}$.

Indeed, in this case we have that $\operatorname{dist}(\Gamma(t), E) \geq 10\,L$ for all $t \in [0, L]$, and since $L \geq |x - y|$ (because we are using an arclength parameterization) we get the desired estimate from the definitions of ω and the ω-length.

From now on we assume that $[0, L] \notin \mathcal{M}$.

Sublemma 6.8. *If $I \in \mathcal{M}$ then* $\inf_{t\in I} \operatorname{dist}(\Gamma(t), E) \le 30\,|I|$.

Indeed, otherwise the (dyadic) parent of I would satisfy (6.6), in contradiction to the maximality of I.

Let Γ_J denote the restriction of Γ to any closed subinterval J of $[0, L]$. A finite sequence $I_1, I_2, \ldots, I_k$ of closed intervals will be called a chain if the the right endpoint of I_j is the same as the left endpoint of I_{j+1} for each $j < k$.

Sublemma 6.9. *There is a constant $C > 0$ so that if $J = [a, b]$ is the union of a chain of intervals $I_1, I_2, \ldots, I_k$ in $\mathcal{M}$, then $|a - b|^s \le C\,\omega$-length(Γ_J).*

Pick $v_j \in E$ such that $\operatorname{dist}(\Gamma_{I_j}, v_j) = \operatorname{dist}(\Gamma_{I_j}, E)$ for $j = 1, 2, \ldots$ Thus $\operatorname{dist}(\Gamma_{I_j}, v_j) \le 30\,|I_j|$ by Sublemma 6.8. Using $\rho(\cdot,\cdot)$ we get that

$$\begin{aligned} |v_1 - v_k|^s &\le C\rho(v_1, v_k) \\ &\le C\sum_{j=1}^{k-1} \rho(v_j, v_{j+1}) \le C\sum_{j=1}^{k-1} |v_j - v_{j+1}|^s . \end{aligned} \tag{6.10}$$

On the other hand

$$\begin{aligned} |v_j - v_{j+1}| &\le \operatorname{dist}(\Gamma_{I_j}, v_j) + \operatorname{dist}(\Gamma_{I_{j+1}}, v_{j+1}) \\ &\qquad + \operatorname{diam}(\Gamma_{I_j} \cup \Gamma_{I_{j+1}}) \\ &\le C(|I_j| + |I_{j+1}|) \end{aligned} \tag{6.11}$$

by our choice of the v_i's, Sublemma 6.8, and the fact that we chose Γ to be parameterized by arclength. Using (6.6) we get that the ω-length of each Γ_{I_j} is at least $C^{-1}|I_j|^s$, and so we conclude from (6.11) that $|v_j - v_{j+1}|^s \le C\,\omega\text{-length}(\Gamma_{I_j} \cup \Gamma_{I_{j+1}})$. Putting this back into (6.10) we obtain

$$|v_1 - v_k|^s \le C\sum_{j=1}^{k-1} \omega\text{-length}(\Gamma_{I_j} \cup \Gamma_{I_{j+1}}) \le C\ \omega\text{-length}(\Gamma_J). \tag{6.12}$$

In the same manner we can obtain that $|a - v_1| \le \operatorname{dist}(\Gamma_{I_1}, v_1) + \operatorname{diam}(\Gamma_{I_1}) \le C\,|I_1|$ and that $|I_1|^s \le C\,\omega\text{-length}(\Gamma_{I_1})$, whence $|a - v_1|^s \le$

$C\,\omega$-length(Γ_{I_1}). Similarly $|b - v_k|^s \le C\,\omega$-length$(\Gamma_{I_k})$. Combining these estimates with (6.12) we get the desired conclusion.

Sublemma 6.13. *Let J be a maximal subinterval of U, and let a, b be its endpoints. Then $|a - b|^s \le C\,\omega$-length(Γ_J).*

This follows from the previous lemma and a limiting argument.

Sublemma 6.14. *Let $\varepsilon > 0$ be given. There is a finite chain of points $0 = t_0 < t_1 < \cdots < t_m = L$, $t_j \in F \cup \{0, L\}$, such that for each j either $|t_j - t_{j+1}| < \varepsilon$ or t_j, t_{j+1} are the endpoints of an interval J as in Sublemma* 6.13.

This is an easy exercise.

We are now ready to finish the proof of Lemma 6.5. Let $\varepsilon > 0$ be given, and let $\{t_j\}$ be as in Sublemma 6.14. Then $\Gamma(t_j) \in E$ when $1 \le j < m$, and so using $\rho(\cdot,\cdot)$ we get that

$$\begin{aligned}
|\Gamma(t_1) - \Gamma(t_{m-1})|^s &\le C\,\rho(\Gamma(t_1), \Gamma(t_{m-1})) \\
&\le C \sum_{j=1}^{m-2} \rho(\Gamma(t_j), \Gamma(t_{j+1})) \\
&\le C \sum_{j=1}^{m-2} |\Gamma(t_j) - \Gamma(t_{j+1})|^s .
\end{aligned} \tag{6.15}$$

Hence $|\Gamma(t_0) - \Gamma(t_m)|^s \le C \sum_{j=0}^{m-1} |\Gamma(t_j) - \Gamma(t_{j+1})|^s$. The terms in this sum fall into two categories. The first are the terms for which t_j, t_{j+1} are the endpoints of an interval J as in Sublemma 6.13. The sum of these terms is at most $C\,\omega$-length(Γ), by Sublemma 6.13. For the remaining terms we have $|t_j - t_{j+1}| < \varepsilon$. Since we are assuming that Γ is parameterized by arclength, for such a j we have that $|\Gamma(t_j) - \Gamma(t_{j+1})|^s \le \varepsilon^{s-1} |\Gamma(t_j) - \Gamma(t_{j+1})|$, and so the sum of these terms is dominated by ε^{s-1} times the Euclidean length of Γ. Altogether we get that

$$|x - y|^s = |\Gamma(t_0) - \Gamma(t_m)|^s \le C\,\omega\text{-length}(\Gamma) + C\,\varepsilon^{s-1}\text{length}(\Gamma). \tag{6.16}$$

Sending ε to 0 we get that $|x - y|^s \le C\,\omega$-length(Γ), which proves Lemma 6.5.

In order to derive Theorem 6.3 from Lemma 6.5 we just need to make some simple observations.

Lemma 6.17. *If B is a ball in $\mathbb{R}^n$ such that $3B$ is disjoint from E, then* $\sup_{2B}\omega \le C \inf_{2B}\omega$ *for some constant C which does not depend on B.*

This follows from the definition of ω and simple geometric considerations.

Lemma 6.18. *There is a constant C so that $D_\omega(x,y) \ge C^{-1}\delta_\omega(x,y)$ for all $x,y \in \mathbb{R}^n$.*

If $3B_{x,y}$ touches E, then $\delta_\omega(x,y) \le C\,|x-y|^s$ by definition of δ_ω and ω, and the desired inequality follows from Lemma 6.5. If $3B_{x,y}$ is disjoint from E, then it is easy to derive the required inequality from Lemma 6.17 and the definition of $D_\omega(x,y)$.

Next we need the following estimate on the thinness (or porosity) of E.

Lemma 6.19. *There is a constant C so that for each ball $B(x,r)$ in $\mathbb{R}^n$ we can find a point $z \in B(x,r/2)$ such that* $\operatorname{dist}(z,E) \ge C^{-1}r$.

Let y be any point in $\partial B(x,r/2)$, and let γ denote the segment which joins x to y. If $2B_{x,y}$ is disjoint from E then there is nothing to prove. If not, then we can apply Lemma 6.5 to conclude that the ω-length of γ is at least $C^{-1}\,|x-y|^s$. This implies that $\sup_{z\in\gamma}\operatorname{dist}(z,E) \ge C^{-1}|x-y|$, because of the definitions of ω and the ω-length, and the lemma follows.

Lemma 6.20. *E has Lebesgue measure zero, and there is a constant $C > 0$ so that*

$$\sup_{2B}\omega \le C\,\frac{1}{|B|}\int_B \omega\,, \tag{6.21}$$

for all balls B in $\mathbb{R}^n$.

The first part is a consequence of Lemma 6.19, which implies that E can have no points of density. (See also the remarks after Lemma

4.6.) The second part follows from Lemma 6.17 when $3B$ is disjoint from E. If $3B$ intersects E, then $\sup_{2B}\omega \leq (5\,\text{radius}(B))^{n(s-1)}$, while the right side of (6.21) is bounded from below by $C^{-1}\text{radius}(B)^{n(s-1)}$ for some constant C, because of Lemma 6.19 and the definition of ω. This proves Lemma 6.20.

Lemma 6.20 implies that ω is a continuous weight which is also doubling. Another easy consequence of (the second part of) Lemma 6.20 is that $D_\omega(x,y) \leq C\delta_\omega(x,y)$. Combining this with Lemma 6.18 we obtain that ω is a strong A_∞ continuous weight.

To prove the first inequality in (6.4), notice that if at least one of x, y lies in E, then $\text{dist}(z, E) \leq |x-y|$ for all z on the line segment that joins x and y, and the ω-length of this line segment is less or equal than $C\,|x-y|^s$. The second inequality in (6.4) comes from Lemma 6.5 and is true for all x, y, and the proof of Theorem 6.3 is now complete.

7. Metric spaces which do not admit bilipschitz embeddings into Euclidean spaces.

Theorem 7.1. *There is a metric space M which satisfies a doubling condition (as in Definition* 1.13) *but which is not bilipschitz equivalent to a subset of any Euclidean space.*

This theorem was known to Assouad, and it is an easy consequence of [P], but it does not seem to have been stated explicitly anywhere.

To prove Theorem 7.1 we use the 3-dimensional Heisenberg group (or any other Carnot group, as in [P, Definition 1.2]) equipped with its Carnot metric. For the sake of simplicity we leave the precise definitions to [P] (see especially paragraph 1.1 on p. 3), but basically the Carnot metric on the Heisenberg group is a distance function that is defined by minimizing the length of the "horizontal" curves which connect a given pair of points, where a curve is said to be horizontal if at each point it is tangent to a certain (completely nonintegrable) distribution of planes. This distance function is invariant under group translations and scales a certain way under a natural family of dilations. The built-in degeneracy of the metric leads to a certain fractal quality of the resulting metric space. For instance, the 3-dimensional Heisenberg group has Hausdorff dimension 4 with respect to its Carnot metric.

The Heisenberg group with its Carnot metric certainly satisfies a doubling condition. The group of translations and dilations can be employed to reduce this propery to the (true) statement that the unit ball centered at the origin can be covered by a finite number of balls of radius 1/2.

One of the reasons that the Heisenberg group with its Carnot metric (and other Carnot groups) are so interesting geometrically is that real-valued Lipschitz functions on them are differentiable almost everywhere. This is a special case of [P, Theorem 2, p. 4]. The precise notion of "differentiability" is given in [P, Paragraph 1.3, p. 4], but it comes down to the usual idea: one "blows up" the given function at a given point and asks that a limiting object exist, and one uses the group of translations and dilations to realize the tangent map as a map on the original Heisenberg group (or Carnot group). The theorem in [P] states not only the existence of the differential almost everywhere, but also its realizability as a group homomorphism which is compatible with the respective groups of dilations.

Let us call the 3-dimensional Heisenberg group with its Carnot metric M. If M had a bilipschitz embedding f into some Euclidean space $\mathbb{R}^n$, then the aforementioned result would imply that f is differentiable almost everywhere in the sense of [P]. The blowing-up procedure used to define the differential scales in the natural way, so that the differential is bilipschitz since f itself is. This gives a contradiction, because any homomorphism from the 3-dimensional Heisenberg group into $\mathbb{R}^n$ must have a kernel which is at least 1-dimensional (all commutators in the Heisenberg group must be mapped to 0 by the homomorphism) and hence cannot be bilipschitz.

Theorem 2 in [P] on the differentiability almost everywhere of Lipschitz functions on the Heisenberg group (or other Carnot groups) actually allows the mapping to take values in another Carnot group and not just the real line. The special (linear) case of real-valued (and hence $\mathbb{R}^n$-valued) Lipschitz functions should be much older than [P], although I did not find a reference. It is certainly within the realm of the usual subelliptic analysis on Carnot groups, and I doubt that it would be very difficult to adapt the standard methods in Harmonic Analysis for proving the the differentiability a.e. of Lipschitz functions on Euclidean spaces (using maximal functions, etc.) to the case of the Heisenberg group and other Carnot groups using the standard tools for doing analysis on these groups (as in [Fo], [FS1], [FS2], [Je], [St2]).

Theorem 7.1 together with Theorem 1.15 imply Theorem 1.12, but unfortunately it is not at all clear how small we can take the dimension d

to be. By this method one would first have to know the best dimension n in Assouad's Theorem 1.14 when M is taken to be the Heisenberg group with the Carnot metric and α is allowed to be any positive number. In any case n is at least 5.

It would be very interesting to have some alternative construction for Theorem 7.1 which is more direct. Aside from understanding how small the dimension d in Theorem 1.12 can be, it would be good to have a simpler and more direct understanding of why examples as in Theorem 7.1 exist.

A related problem is to find other examples of metric spaces for which there is some kind of rigidity theorem along the lines of "real-valued Lipschitz functions are differentiable almost everywhere". I do not know of any examples which are not somehow based on Euclidean geometry or the geometry of Carnot groups, nor do I know whether any such examples should exist. This is related to the WALA and GWALA in [DS4]. (See [DS4, p. 45-6 and Chapter III.4]. The issue there is to decide whether certain uniform rigidity properties of Lipschitz functions on a set $E \subseteq \mathbb{R}^n$ should force E to be "uniformly rectifiable". This problem was not resolved satisfactorily in [DS4].) Notice that no such results are true for self-similar Cantor sets or snowflakes; Lipschitz functions on these types of sets are as flabby as Hölder continuous functions on $\mathbb{R}^n$ of order less than 1.

8. Regular mappings.

Definition 8.1. *Let M and N be metric spaces. A mapping $f : M \to N$ is said to be regular if it is Lipschitz continuous and if there is a constant $C > 0$ so that if B is a ball in N then $f^{-1}(B)$ can be covered by at most C balls in M of the same radius as B.*

Note that no requirements are being imposed on the position of these balls in M which cover $f^{-1}(B)$.

To my knowledge this kind of condition was first considered in the context of Euclidean spaces (with the metric perhaps deformed by a weight) in [D1], [D2]. The definition given in [D1], [D2] is slightly different but equivalent to this one in the case of Euclidean spaces.

In practice we shall only be considering metric spaces which satisfy a doubling condition, and so bilipschitz embeddings will automatically be regular. Notice that the bilipschitz condition can be reformulated as meaning that f is Lipschitz and $f^{-1}(B)$ is contained in a single ball whose radius is allowed to be larger than the radius of B by only a bounded factor. Roughly speaking, bilipschitzness is a uniform and scale-invariant version of injectivity, while regularity is a uniform and scale-invariant version of the requirement that a map have bounded multiplicity. (Note that regular maps have bounded multiplicity.)

A simple example of a regular mapping is $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = |x|$. It is easy to draw curves in the plane whose arclength parameterization is regular. They are allowed to cross themselves many times, but they are not allowed to have too much mass concentrate in a disk.

Regular mappings have many of the same nice properties as Lipschitz and bilipschitz mappings. For instance, the regularity condition scales the same way as the Lipschitz and bilipschitz conditions, and the composition of two regular mappings is regular. Regular mappings also share some of the important features of bilipschitz mappings, *e.g.*, they can increase Hausdorff measure of any dimension by only a bounded factor.

There is an analogue of Theorems 7.1 and 1.12 for regular mappings.

Theorem 8.2. *There is a metric space M which satisfies a doubling condition but which does not admit a regular mapping into any Euclidean space.*

Corollary 8.3. *There is a strong-A_∞ weight on some $\mathbb{R}^d$ such that $(\mathbb{R}^d, D_\omega)$ does not admit a regular mapping into any Euclidean space.*

Theorem 8.2 is proved in exactly the same way as Theorem 7.1. Take M to be the 3-dimensional Heisenberg group with its Carnot metric, and suppose that $f : M \to \mathbb{R}^n$ is regular. In particular it is Lipschitz, and so it is differentiable almost everywhere. Because of the natural scale-invariance of the regularity condition we have that the differential of f is also regular whenever it exists. Since the differential is almost always a group homomorphism, we get a contradiction as before, because the kernel of such a homomorphism has dimension at least one.

Corollary 8.3 is an immediate consequence of Theorem 8.2 and Theorem 1.15.

The questions posed at the end of the preceding section (concerning the possible small values of d in Corollary 8.3 and alternative constructions for Theorem 8.2 and Corollary 8.3) are also open in the case of regular mappings. Here is another one.

PROBLEM 8.4. *Let M be a metric which satisfies a doubling condition. If M admits a regular mapping into some Euclidean space, must M admit a bilipschitz embedding into one also?*

Of course the examples above were the same for Theorems 7.1 and 8.2.

Proposition 8.5. *The answer to Problem 8.4 is affirmative if and only if it is affirmative for the special case of metric spaces of the form $(\mathbb{R}^d, D_\omega(\cdot,\cdot))$, where ω is a strong A_∞ continuous weight.*

This is proved using Theorem 1.15, and the argument is given at the end of the section.

Regular mappings from Euclidean spaces into other (larger) Euclidean spaces are quite interesting. It turns out that such mappings have a considerable amount of bilipschitz behavior, and in particular that they have "large bilipschitz pieces". See [D1], [D2], [D3], [D4], and [Js], and see [DS3] for a related notion of "weakly bilipschitz". This good behavior is sufficient to ensure the L^p boundedness of singular integral operators on the image of the mapping, as in [D1], [D2]. Regular mappings are also flexible enough so that there are some general existence results. For instance, suppose that E is a d-dimensional regular subset of $\mathbb{R}^n$ which is "uniformly rectifiable" in the sense of [DS4]. This means that inside each ball centered on E there should be a substantial fraction of E which is bilipschitz equivalent to a subset of $\mathbb{R}^d$, with uniform bounds. Then part of the main result of [DS2] is that there is an A_1 weight ω on $\mathbb{R}^d$ and a regular mapping ϕ from the metric space $(\mathbb{R}^d, D_\omega(\cdot,\cdot))$ into $\mathbb{R}^{n+1}$ whose image contains E. (See Definition 2.8 for the definition of an A_1 weight, and note that for this result we need to allow discontinuous weights.) It is not known whether this last result is true with the weight ω simply taken to be constant. Because a regular image of $(\mathbb{R}^d, D_\omega(\cdot,\cdot))$ is necessarily uniformly rectifiable for any A_1 weight ω, this question comes down to the following.

PROBLEM 8.6. *Suppose that ϕ is a regular map from* $(\mathbb{R}^d, D_\omega(\cdot,\cdot))$

into some $\mathbb{R}^n$, where ω is an A_1 weight. Is there then a regular map $\psi : \mathbb{R}^d \to \mathbb{R}^m$ for some $m \geq n$ such that the image of ψ contains the image of ϕ (with $\mathbb{R}^n$ viewed as a subspace of $\mathbb{R}^m$)?

Remember that Problem 2.10 asks whether $(\mathbb{R}^d, D_\omega(\cdot,\cdot))$ is bilipschitz equivalent to $\mathbb{R}^d$ with the Euclidean metric when ω is an A_1 weight. If this is true then Problem 8.6 also has an affirmative answer, with ψ simply a reparameterization of ϕ. For that matter Problem 8.6 would have an affirmative answer if there is a regular mapping from $\mathbb{R}^d$ with the Euclidean metric onto $(\mathbb{R}^d, D_\omega(\cdot,\cdot))$.

The main result of [DS2] contains a result of the same type as Problem 8.6. Specifically, the image of a regular mapping from $(\mathbb{R}^d, D_\omega(\cdot,\cdot))$ into some $\mathbb{R}^n$ for ω a strong A_∞ weight is uniformly rectifiable and hence is contained in the image of a regular mapping associated to an A_1 weight. Thus A_1 weights are natural for Problem 8.6. (See also the discussion of open problems in [DS2, Section 21].)

A special case of the main result in [Se4] is that $(\mathbb{R}^d, D_\omega(\cdot,\cdot))$ is bilipschitz equivalent to a subset of some $\mathbb{R}^n$ (with the Euclidean metric) when ω is an A_1 weight. If, for a particular ω, this subset could be put inside the image of a regular mapping $\psi_0 : \mathbb{R}^d \to \mathbb{R}^m$ as in Problem 8.6, then the answer to Problem 8.6 would be affirmative in general for ω, *i.e.*, all other regular mappings from $(\mathbb{R}^d, D_\omega(\cdot,\cdot))$ into Euclidean space would also be good. To see this we need an auxiliary fact (which is useful to know anyway).

Proposition 8.7. *Suppose that $E \subseteq \mathbb{R}^n$ is closed and that $f : E \to \mathbb{R}^k$ is regular. Then there exists a regular mapping $F : \mathbb{R}^n \to \mathbb{R}^{k+n+1}$ such that F agrees with f on E, modulo the identification of $\mathbb{R}^k$ with a subspace of $\mathbb{R}^{k+n+1}$ in the obvious way.*

Assuming Proposition 8.7 for the moment let us apply it to the assertion in the preceding paragraph. Let ω be an A_1 weight on $\mathbb{R}^d$, and let $E \subseteq \mathbb{R}^n$ be chosen so that E (with the induced Euclidean metric) is bilipschitz equivalent to $(\mathbb{R}^d, D_\omega(\cdot,\cdot))$. Suppose also that E is contained in the image of a regular mapping $\psi_0 : \mathbb{R}^d \to \mathbb{R}^m$. We may as well assume that $m = n$, because each of m and n can be increased without difficulty. Let $\phi : (\mathbb{R}^d, D_\omega(\cdot,\cdot)) \to \mathbb{R}^k$ be some other regular map, so that we want to find another regular map ψ as in Problem 8.6. Since E is bilipschitz equivalent to $(\mathbb{R}^d, D_\omega(\cdot,\cdot))$ we get a regular mapping $f : E \to \mathbb{R}^k$ which is just a reparameterization of ϕ.

Proposition 8.7 provides us with an extension F of f to all of $\mathbb{R}^n$, and the composition $\psi = F \circ \psi_0$ is a regular mapping from $\mathbb{R}^d$ into $\mathbb{R}^{k+n+1}$ whose image contains the image of ϕ, as desired.

Proposition 8.7 is a simpler and cruder version of the result in [D2, Section 4]. As such it is very similar to [DS2, Proposition 17.4], and the proof below is essentially the same as the argument in [DS2, Section 17], modulo some additional simplifications which are possible in this case.

Lemma 8.8. *Given any closed set $E \subseteq \mathbb{R}^n$ we can find a Lipschitz function $\rho : \mathbb{R}^n \to \mathbb{R}^n$ which vanishes on E and which has the property that if B is a ball such that $2B \subseteq \mathbb{R}^n \setminus E$, then the restriction of ρ to B is regular, with a uniformly bounded constant.*

In fact we can do this in such a way that $\rho|_B$ is bilipschitz, with a uniform bound, at least if we allow ourselves to increase the dimension of the target space. This is basically what happens in [DS2, Section 17], but it is a little simpler to just get regularity.

Note that the mapping ρ in Lemma 8.8 cannot be orientation preserving, because (as pointed out to me by Juha Heinonen) quasiregular mapping theory would then force E to be discrete.

To prove Lemma 8.8 let $\{Q_i\}$ be a Whitney decomposition of $\mathbb{R}^n \setminus E$ into dyadic cubes, as described on [St1, p. 167ff]. We are going to build ρ by gluing together a bunch of little maps on the Whitney cubes Q_i. We shall do this via an induction on skeleta, with the next observation supplying the induction step.

Sublemma 8.9. *Given a j-dimensional cube Q in $\mathbb{R}^n$ and a regular mapping $\sigma : \partial Q \to \mathbb{R}^l$ with constant C_0 (as in Definition 8.1) we can find an extension $\Sigma : Q \to \mathbb{R}^{l+1}$ of σ (modulo the obvious identification of $\mathbb{R}^l$ with $\mathbb{R}^l \times \{0\}$) which is regular with constant $C\,C_0$, where C depends only on the dimensions. (Here ∂Q refers to the "polyhedral" boundary of Q, since Q will be its own topological boundary when $j < n$. Also, $j = 1$ and $l = 0$ are allowed here, with $\mathbb{R}^l$ interpreted as the trivial vector space with only the zero element when $l = 0$.)*

This is easy to prove. Assume for simplicity that Q is centered at the origin and that $0 \in \sigma(\partial Q)$. Let tQ denote the image of Q under the mapping $x \to t\,x$. If $x \in \partial(tQ)$, define $\Sigma(x)$ by taking the $\mathbb{R}^l$ part of $\Sigma(x)$ to be $t\,\sigma(t^{-1}x)$ (0 when $x = 0$) and the last coordinate to be $(1 - t)\operatorname{diam} Q$. It is easy to check that this defines a regular mapping with the correct estimate. Note that in the $j = 1$ case this gives a piecewise-linear mapping on a segment which vanishes at the endpoints, is positive in the middle, and does not preserve orientations.

To define ρ, we begin by setting $\rho = 0$ on E and also on all the vertices of the Whitney cubes Q_i. At this stage we can view ρ as a map into $\mathbb{R}^0$. We extend ρ to the various edges of the Whitney cubes using Sublemma 8.9. Actually, we have to be a little careful; let us call an edge of a Whitney cube *minimal* if it does not properly contain an edge of another Whitney cube, and let us call the collection of all these minimal edges the "minimal edges of the Whitney decomposition". Thus the edge of any Whitney cube is the union of (a bounded number of) minimal edges, disjoint except at the vertices, and the minimal edges contained in any Whitney cube Q_i cannot be smaller than a fixed constant times the sidelength of Q_i. This follows from the fact that if two Whitney cubes intersect, then the intersection must be a face of one of the two cubes (of some dimension j, $0 \leq j < n$), and the two cubes must have approximately the same size. (In fact one can take the "fixed constant" mentioned above to be $1/4$. See [St1, Proposition 1, p. 169].) We extend ρ to the edges of the Whitney cubes by applying Sublemma 8.9 to the minimal edges, which then takes care of all the others. This gives rise to a map into $\mathbb{R}^1$ which is regular on each minimal edge, and hence on any edge of any Whitney cube, with a uniformly bounded regularity constant. We then extend ρ to the various squares using Sublemma 8.9, so that this part of ρ takes values in $\mathbb{R}^2$. Note that a mapping on the boundary of a square is regular if it is continuous at the vertices and regular on each of the four sides of the square, so that we can apply Sublemma 8.9. Also, as before, we should really work with the "minimal squares of the Whitney decomposition", etc. Repeating this argument for the various dimensions up to n we get a map ρ which is defined and regular with a bounded constant on each of the Whitney cubes, and which is also continuous as one passes from a Whitney cube to its neighbor. Because ρ vanishes at the vertices of the Whitney cubes we also obtain that $|\rho(x)| \leq C \operatorname{dist}(x, E)$ for some constant C and all x. This implies that ρ is continuous across E, and it is not hard to check that ρ is Lipschitz on all of $\mathbb{R}^n$, and not just on the various Whitney cubes. The regularity property on balls stated in Lemma 8.8 follows easily from the fact that any such ball is covered by a bounded number of Whitney cubes. This proves Lemma 8.8.

Now let us prove Proposition 8.7. Let $f : E \to \mathbb{R}^k$ be given (and regular), and extend it to a Lipschitz map (also denoted by f) from $\mathbb{R}^n$ into $\mathbb{R}^k$. Identify $\mathbb{R}^{k+n+1}$ with $\mathbb{R}^k \times \mathbb{R}^n \times \mathbb{R}$, and identify $\mathbb{R}^k$ with $\mathbb{R}^k \times \{0\} \times \{0\}$. Let $\rho : \mathbb{R}^n \to \mathbb{R}^n$ be as in Lemma 8.8, and define $h : \mathbb{R}^n \to \mathbb{R}$ by $h(x) = \operatorname{dist}(x, E)$. Take F to be (f, ρ, h).

Let B be a ball in $\mathbb{R}^{k+n+1}$ with radius r, and let t denote the last coordinate of the center of B. We need to show that $F^{-1}(B)$ can be covered by a bounded number of balls of radius r.

Let β denote the ball in $\mathbb{R}^k$ which is the projection of B. By assumption we can cover $f^{-1}(M\beta) \cap E$ with at most $C(M)$ balls of radius r for any given $M > 0$. Let E_s denote the set of points in $\mathbb{R}^n$ at distance at most s from E.

It will be convenient to distinguish some cases concerning the relative values of t and r. If $t < -r$ then every point in B has negative last coordinate. In this case $F^{-1}(B) = \varnothing$, since h is nonnegative. Now suppose that $-r < t \leq 2r$. Then $F^{-1}(B) \subseteq E_{3r}$, by definition of h. For each point $z \in F^{-1}(B)$ there is a point $y \in E$ such that $|y - z| \leq 3r$, and so $|f(y) - f(z)| \leq 3Lr$, where L denotes the Lipschitz norm of f. That is to say, z lies within $3r$ of $f^{-1}((3L+1)\beta) \cap E$. From this it follows easily that $F^{-1}(B)$ can be covered by a bounded number of balls of radius r, since this is true for $f^{-1}((3L+1)\beta) \cap E$, by assumption.

It remains to deal with the case where $t > 2r$. In this case we have that $F^{-1}(B) \subseteq E_{2t} \setminus E_{t/2}$, again because of the definition of h. Let B' denote the ball in $\mathbb{R}^{k+n+1}$ with the same center as B but with radius t, so that $B' \supseteq B$. Then B' is a ball of the type considered in the previous case, and so $F^{-1}(B)$ is covered by a bounded number of balls of radius t. Since $F^{-1}(B) \subseteq \mathbb{R}^n \setminus E_{t/2}$ we obtain that $F^{-1}(B)$ is covered by a bounded number of balls whose doubles do not touch E. Because ρ is regular on each of these balls, with bounded constant, we conclude that $F^{-1}(B)$ is covered by a bounded number of balls with radius r. This proves Proposition 8.7.

Let us now prove Proposition 8.5. The "only if" part is trivial, since the metric spaces coming from strong A_∞ weights satisfy doubling conditions. Conversely, let $(M, d(\cdot,\cdot))$ be a metric space which satisfies a doubling condition and which admits a regular mapping into some Euclidean space. Fix $\alpha \in (0,1)$, *e.g.*, take $\alpha = 1/2$. By Assouad's Theorem 1.14 we can find a set E in some $\mathbb{R}^n$ such that $(M, d(\cdot,\cdot)^\alpha)$ is bilipschitz equivalent to E (with the Euclidean metric). Thus E is a snowflake of order $1/\alpha$, in the sense of Definition 6.1, and so we can apply Theorem 6.3 with $s = 1/\alpha$ to get a strong A_∞ weight ω such that $(E, D_\omega(\cdot,\cdot))$ is bilipschitz equivalent to $(M, d(\cdot,\cdot))$. Our assumption that M admits a regular mapping into some Euclidean space translates into the condition that there is a regular mapping $g : (E, D_\omega(\cdot,\cdot)) \to \mathbb{R}^k$ for some k. In order to prove Proposition 8.5 it suffices to show that there is a regular mapping $G : (\mathbb{R}^n, D_\omega(\cdot,\cdot)) \to \mathbb{R}^{k+n+1}$ which agrees with g on E, modulo the usual identification of $\mathbb{R}^n$ with a subspace of $\mathbb{R}^{k+n+1}$. (The sufficiency of this statement comes down to the fact that $(M, d(\cdot,\cdot))$ is bilipschitz equivalent to a subset of some Euclidean space if $(\mathbb{R}^n, D_\omega(\cdot,\cdot))$ is, since $(M, d(\cdot,\cdot))$ is bilipschitz equivalent to $(E, D_\omega(\cdot,\cdot))$.)

The construction of the mapping G is completely analogous to the proof of Proposition 8.7; we simply have make some adjustments for the weight.

Lemma 8.10. *There is a Lipschitz mapping $\tau : (\mathbb{R}^n, D_\omega(x,y)) \to (\mathbb{R}^n, |x-y|)$ which vanishes on E and which has the property that if B is a ball such that $2B \subseteq \mathbb{R}^n \setminus E$, then $\tau : (B, D_\omega(x,y)) \to (\mathbb{R}^n, |x-y|)$ is regular, with a uniformly bounded constant.*

This is proved in much the same way as Lemma 8.8 was. Let $\{Q_i\}$ be a Whitney decomposition of $\mathbb{R}^n \setminus E$, and define τ initially on E and on the vertices of the Whitney cubes Q_i as a map into $\mathbb{R}^0 = \{0\}$ which vanishes identically. We want to extend τ to the edges of the Whitney cubes, the higher-dimensional faces, and eventually to the Whitney cubes themselves as before, except that on these various faces we should build τ so that it is regular as a map with the D_ω-metric on the domain and the Euclidean metric on the range. In order to do this we need to have a version of Sublemma 8.9 where σ and Σ are regular as maps with the D_ω-metric on the domain and with the Euclidean metric on the range. This version of Sublemma 8.9 is true, at least when the cube Q in Sublemma 8.9 is contained in a Whitney cube (as it always is for our application). Indeed, $D_\omega(x,y)$ is comparable in size to $(\operatorname{diam} Q_i)^\delta |x-y|$ when x, y lie in a Whitney cube Q_i, where $\delta = 1/\alpha - 1$, and hence a mapping on a subset A of Q_i is regular with respect to $D_\omega(x,y)$ with a bounded constant if and only if it equals $(\operatorname{diam} Q_i)^\delta$ times a mapping which is regular with respect to the Euclidean metric with a bounded constant. This allows the aforementioned D_ω version of Sublemma 8.9 to be derived from the original Euclidean statement. Once we have this D_ω version of Sublemma 8.9 we can construct τ in the same manner as before, and we get a map into $(\mathbb{R}^n, |x-y|)$ which is regular with respect to $D_\omega(x,y)$ on each Whitney cube Q_i, and with a bounded constant. The regularity property on balls required in Lemma 8.10 again follows from the observation that any such ball is covered by a bounded number of Whitney cubes. On the other hand we have that $|\tau(x)| \le C\,(\operatorname{diam} Q_i)^{1/\alpha}$ for x in a Whitney cube Q_i, because τ vanishes on the vertices of Q_i (by construction) and is uniformly Lipschitz with respect to D_ω on Q_i, and this is the same as saying that $|\tau(x)|$ is bounded by a constant times the D_ω distance from x to E for any $x \in \mathbb{R}^n$ (because of Theorem 6.3, for instance). Using this bound it is not hard to show that τ has the correct Lipschitz property, and Lemma 8.10 follows.

Next define $h(x)$ for $x \in \mathbb{R}^n$ to be the D_ω-distance from x to E. This is about the same as the Euclidean distance to E raised to the power $1/\alpha$, by Theorem 6.3. Notice that h is Lipschitz as a map from $(\mathbb{R}^n, D_\omega(x,y))$ into $\mathbb{R}$ equipped with the Euclidean metric.

Extend g to a Lipschitz map from $(\mathbb{R}^n, D_\omega(\cdot,\cdot))$ to $\mathbb{R}^k$ equipped with the Euclidean metric. The existence of such an extension follows from the fact that $(\mathbb{R}^n, D_\omega(\cdot,\cdot))$ is a metric space, but in this case one could also use the methods of the Whitney extension theorem (as in [St1, Chapter VI]). Take G to be (g, τ, h), with $\mathbb{R}^{k+n+1}$ identified with $\mathbb{R}^k \times \mathbb{R}^n \times \mathbb{R}$, and identify $\mathbb{R}^k$ with $\mathbb{R}^k \times \{0\} \times \{0\}$.

The proof that G is regular is practically the same as in the proof of Proposition 8.7 above. Let B be a (Euclidean) ball in $\mathbb{R}^{k+n+1}$ with radius r, and let t denote the last coordinate of the center of B. We need to show that $G^{-1}(B)$ can be covered by a bounded number of ω-balls of ω-radius r. As before, there is nothing to do when $t < -r$.

For this proof let E_s denote the set of points $z \in \mathbb{R}^n$ such that the ω-distance from z to E is $\le s$, *i.e.*, $h(z) \le s$. If we assume now that $-r < t < 2r$, then $G^{-1}(B) \subseteq E_{3r}$. Thus for each $z \in G^{-1}(B)$ there is a $y \in E$ with $|g(y) - g(z)| \le L\, D_\omega(y,z) \le 3Lr$, where L is the Lipschitz norm of G. This implies that the ω-distance from z to $g^{-1}((3L+1)\beta) \cap E$ is $\le 3r$, where β is again the projection of B in $\mathbb{R}^k$. This implies that $G^{-1}(B)$ is contained in the union of a bounded number of ω-balls of radius $(3L+4)\,r$, since $g|_E$ is assumed to be regular, and the doubling property for the metric space $(\mathbb{R}^n, D_\omega(x,y))$ allows us to conclude that $G^{-1}(B)$ is covered by a bounded number of ω-balls of radius r, which is what we wanted.

Now suppose that $t > 2r$. In this case we have that $G^{-1}(B) \subseteq$

$E_{2t} \setminus E_{t/2}$, because of the definition of h. Let B' denote the (Euclidean) ball in $\mathbb{R}^{k+n+1}$ with the same center as B but with radius t, so that $B' \supseteq B$. Then B' is a ball of the type considered in the previous case, and so $G^{-1}(B)$ is covered by a bounded number of ω-balls of radius t. Since $G^{-1}(B) \subseteq \mathbb{R}^n \setminus E_{t/2}$ we obtain that $G^{-1}(B)$ is covered by a bounded number of Euclidean balls whose doubles do not touch E. This is not too hard to check, but one must be a little careful. (One way to do this uses the observation that for each $\varepsilon > 0$ we can cover $G^{-1}(B)$ by at most $C(\varepsilon)$ ω-balls of radius $\varepsilon\, t$, because of the doubling property for $(\mathbb{R}^n, D_\omega)$. If ε is small enough, then these smaller ω-balls will be contained in Euclidean balls whose doubles do not touch E. Alternatively, all the relevant ω-balls of radius t which cover $G^{-1}(B)$ lie in E_{4t}, and we can use the fact that that D_ω is comparable to a constant multiple of the Euclidean metric on $E_{10t} \setminus E_{t/5}$ in this special situation.) Therefore τ is regular on each of these balls, with bounded constant, and we conclude that $G^{-1}(B)$ is covered by a bounded number of ω-balls with radius r.

Thus we obtain that G is regular, which is what we wanted. This completes the proof of Proposition 8.5.

Acknowledgements. I am grateful to some topologists for telling me some things that I needed to know, but I am embarrassed for taking so long to realize their importance. I would also like to thank Juha Heinonen for pointing out the ridiculousness of one of my sublemmas before it was too late, and for other comments as well.

References.

[A1] Assouad, P., Espaces Métriques, Plongements, Facteurs. Thèse de Doctorat, Université de Paris XI, France, 1977.

[A2] Assouad, P., Étude d'une dimension métrique liée à la possibilité de plongement dans R^n. *C. R. Acad. Sci. Paris* **288** (1979), 731-734.

[A3] Assouad, P., Plongements Lipschitziens dans R^n. *Bull. Soc. Math. France* **111** (1983), 429-448.

[B] Blankinship, W. A., Generalization of a construction of Antoine. *Ann. of Math.* **53** (1951), 276-297.

[C1] Cannon, J., The characterization of topological manifolds of dimension $n \geq 5$. *Proceedings I.C.M.* (Helsinki, 1978), 449-454.

[C2] Cannon, J., Shrinking cell-like decompositions of manifolds: codimension 3. *Ann. of Math.* **110** (1979), 83-112.

[CW1] Coifman, R. Weiss, G., *Analyse Harmonique Non-commutative sur Certains Espaces Homogènes.* Lecture Notes in Math. **242** (1971), Springer-Verlag.

[CW2] Coifman, R., Weiss, G., Extensions of Hardy spaces and their use in analysis. *Bull. Amer. Math. Soc.* **83** (1977), 569-645.

[Da] Daverman, R., *Decompositions of Manifolds*, Academic Press, 1986.

[D1] David, G., Opérateurs intégraux singuliers sur certaines courbes du plan complexe. *Ann. Sci. École. Norm. Sup.* **17** (1984), 157-189.

[D2] David, G., Opérateurs d'intégrale singuliere sur les surfaces rǵulières. *Ann. Sci. École. Norm. Sup.* **21** (1988), 225-258.

[D3] David, G., Morceaux de graphes lipschitziens et intégrales singulières sur un surface. *Revista Mat. Iberoamericana* **4** (1988), 73-114.

[D4] David, G., *Wavelets and Singular Integrals on Curves and Surfaces.* Lecture Notes in Math. **1465** (1991), Springer-Verlag.

[DS1] David, G., Semmes, S., Strong A_∞-weights, Sobolev inequalities, and quasiconformal mappings. In *Analysis and Partial Differential Equations*, edited by C. Sadosky, Lecture Notes in Pure and Appl. Math. **122**, Marcel Dekker, 1990.

[DS2] David, G., Semmes, S., Singular Integrals and Rectifiable Sets in R^n: au-delà des graphes lipschitziens. *Astérisque* **193**, Soc. Math. France, 1991.

[DS3] David, G., Semmes, S., Quantitative rectifiability and Lipschitz mappings. *Trans. Amer. Math. Soc.* **337** (1993), 855-889.

[DS4] David, G., Semmes, S., *Analysis of and on uniformly rectifiable sets.* Math. Surveys and Monographs **38**, Amer. Math. Soc., 1993.

[DS5] David, G., Semmes, S., to exist.

[E] Edwards, R., The topology of manifolds and cell-like maps. *Proceedings I.C.M.* (Helsinki, 1978), 111-127.

[Fo] Folland, G., Subelliptic estimates and function spaces on nilpotent Lie groups. *Arkiv Math.* **13** (1975), 161-207.

[FS1] Folland, G., Stein, E., Estimates for the $\bar{\partial}_b$-complex and analysis on the Heisenberg group. *Comm. Pure Appl. Math.* **27** (1974), 429-522.

[FS2] Folland, G., Stein, E., Hardy Spaces on Homogeneous Groups. *Math. Notes* **28** (1982), Princeton University Press.

[Ga] Garnett, J., *Bounded Analytic Functions*, Academic Press, 1981.

[Ge] Gehring, F. W., The L^p integrability of the partial derivatives of a quasiconformal mapping. *Acta Math.* **130** (1973), 265-277.

[Gr] Gromov, M., *Structures Métriques pour les Variétés Riemanniennes.* (J. Lafontaine, P. Pansu, ed.), Cedic/Fernand Nathan, 1981.

[HK1] Heinonen, J., Koskela, P., Definitions of quasiconformality. *Invent. Math.* **120** (1995), 61-79.

[Je] Jerison, D., The Poincaré inequality for vector fields satisfying Hörmander's condition. *Duke Math. J.* **53** (1986), 503-523.

[Jr] Journé, J. L., *Calderón-Zygmund Operators, Pseudodifferential Operators, and the Cauchy Integral of Calderón.* Lecture Notes in Math. **994**, Springer-Verlag, 1983.

[Js] Jones, P., Lipschitz and bi-Lipschitz functions. *Revista Mat. Iberoamericana* **4** (1988), 115-122.

[LZ] Luukkainen, J., Vaisälä, J., Elements of Lipschitz topology. *Ann. Acad. Sci. Fenn. Ser. A I Math.* **3** (1977), 85-122.

[Mo] Moise, E. E., *Geometric Topology in Dimensions* 2 *and* 3. Graduate Texts in Math. **47**, Springer-Verlag, 1977.

[MRV] Martio, O., Rickman, S., Väisälä, J., Topological and metric properties of quasiregular mappings. *Ann. Acad. Sci. Fenn. Ser. A I Math.* **488** (1971), 1-31.

[P] Pansu, P., Métriques de Carnot-Carathéodory et quasiisométries des espaces symétriques de rang un. *Ann. of Math.* **129** (1989), 1-60.

[R] Rushing, T. B., Hausdorff dimension of wild fractals. *Trans. Amer. Math. Soc.* **334** (1992), 597-613.

[Se1] Semmes, S., Chord-arc surfaces with small constant I. *Advances in Math.* **85** (1991), 198-223.

[Se2] Semmes, S., Chord-arc surfaces with small constant II: Good parameterizations. *Advances in Math.* **88** (1991), 170 -199.

[Se3] Semmes, S., Hypersurfaces in $\mathbb{R}^n$ whose unit normal has small BMO norm. *Proc. Amer. Math. Soc.* **112** (1991), 403-412.

[Se4] Semmes, S., Bilipschitz mappings and strong A_∞ weights. *Ann. Acad. Sci. Fenn. Ser. A I Math.* **18** (1993), 211-248.

[Se5] Semmes, S., Finding structure in sets with little smoothness. *Proc. I.C.M.* (Zürich, 1994), 875-885. Birkhäuser, 1995.

[Se6] Semmes, S., Good metric spaces without good parameterizations. *Revista Mat. Iberoamericana* **12** (1996), 187-275.

[SS] Siebenmann, L., Sullivan, D., On complexes that are Lipschitz manifolds. In *Geometric Topology*, J. Cantrell, ed., Academic Press, 1979, 503-525.

[St1] Stein, E. M., *Singular Integrals and Differentiability Properties of Functions*, Princeton University Press, 1970.

[St2] Stein, E. M., *Harmonic Analysis: Real-Variable Methods, Orthogonality, and Oscillatory Integrals*, Princeton University Press, 1993.

[T1] Toro, T., Surfaces with generalized second fundamental form in L^2 are Lipschitz manifolds. *J. Differential Geom.* **39** (1994), 65-102.

[T2] Toro, T., Geometric conditions and existence of bilipschitz parameterizations. *Duke Math. J.* **77** (1995), 193-227.

Recibido: 15 de febrero de 1.995

Stephen Semmes*
Department of Mathematics
Rice University
Houston, Texas 77251, U.S.A.
semmes@math.rice.edu

and

I.H.E.S.
35 Route de Chartres
91440 Bures-sur-Yvette, FRANCE
semmes@ihes.fr

* The author is supported by the U.S. National Science Foundation.

Revista Matemática Iberoamericana
Vol. 12, N.° 2, 1996

Self-similar solutions in weak L^p-spaces of the Navier-Stokes equations

Oscar A. Barraza

Abstract. The most important result stated in this paper is a theorem on the existence of global solutions for the Navier-Stokes equations in $\mathbb{R}^n$ when the initial velocity belongs to the space weak $L^n(\mathbb{R}^n)$ with a sufficiently small norm. Furthermore, this fact leads us to obtain self-similar solutions if the initial velocity is, besides, an homogeneous function of degree -1. Partial uniqueness is also discussed.

1. Introduction.

We start our exposition by setting the central problem and establishing the framework of the principal ideas. First of all we are going to try to give to the reader a modest approach to the physical meaning.

Let us consider the Navier-Stokes equations of a viscous incompressible fluid which fills an infinite cylinder of cross section Ω, an open subset of $\mathbb{R}^n$. These equations govern the flow of the fluid which moves parallel to the plane of Ω when an external force $f = (f_1, f_2, \dots, f_n)$ is present.

The vector $\mathbf{v}(t, x)$ represents the velocity field of a particle at spatial position x and time t, and the function $\mathrm{p}(t,x)$ the pressure at x and time t, respectively. They are the unknows of the Navier-Stokes system

$$\begin{cases} \rho\left(\dfrac{\partial \mathbf{v}}{\partial t} + (\mathbf{v}\cdot\nabla)\mathbf{v}\right) - \mu\,\Delta\mathbf{v} + \nabla\mathrm{p} = f\,, \\ \nabla\cdot\mathbf{v} = 0\,, \end{cases} \tag{1}$$

where the constant $\rho > 0$ is the density of the fluid and $\mu > 0$ is the kinematic viscosity. As usual, $\nabla\mathrm{p} = (\partial_1\mathrm{p}, \partial_2\mathrm{p}, \dots, \partial_n\mathrm{p})$ denotes the gradient of p and $\nabla\cdot\mathbf{v} = \sum_{j=1}^{n}\partial_j\mathbf{v}_j$ the divergence of $\mathbf{v} = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n)$.

The first equation of system (1) is the momentum conservation equation and the second one is the mass conservation equation (incompressibility condition) [T].

We shall limit ourselves to the study of the existence of solutions of the Cauchy problem for the equation (1) in the case that Ω is the whole space $\mathbb{R}^n$. Then, there will not be any external force ($f \equiv 0$). Moreover, we will assume that the density $\rho = 1$ and the viscosity $\mu = 1$. The initial data of the velocity is a vectorial field $\mathbf{v}_0$ satisfying the condition $\nabla\cdot\mathbf{v}_0 = 0$ in the distributional sense.

Therefore, the problem in which we are interested is

$$\begin{cases} \dfrac{\partial \mathbf{v}}{\partial t} - \Delta\mathbf{v} + (\mathbf{v}\cdot\nabla)\mathbf{v} + \nabla\mathrm{p} = 0\,, \\ \nabla\cdot\mathbf{v} = 0\,, \\ \mathbf{v}(0,x) = \mathbf{v}_0(x)\,. \end{cases} \tag{2}$$

Of course, the pressure p will be disregarded. It will be automatically determined by mean of the first equation of (2) after computing the velocity $\mathbf{v}$, except for an additive function depending on time.

In his paper [K], Kato proved the existence and uniqueness of solutions for the problem (2) in the L^p-theory based on the technical details developed in [K-F]. In both works it was shown that such a global solution exists using the successive approximation method, applied to the integral equation formally equivalent to the initial value problem (2). This problem is recently solved by M. Cannone [C] for any abstract "adapted space" among the concrete applications we find the L^p spaces and some kind of Besov spaces. In Kozono-Yamazaki's paper [K-Y], the existence and the uniqueness of Cauchy problem (2) are treated too for a larger new function spaces constructed in the same way as the Besov ones, based on the Morrey spaces instead of the usual L^p spaces.

However, in the study of the attractors associated with the Navier-Stokes equations [T] it is necessary to be able to find "self-similar solutions". That is, solutions $\mathbf{v}(t,x)$ which satisfy $\mathbf{v}(t,x) = \lambda\mathbf{v}(\lambda^2 t, \lambda x)$

for all $x \in \mathbb{R}^n$, all $t > 0$ and all $\lambda > 0$. This kind of solutions are related to an asymptotic behaviour, for large time, of the global solutions of the Navier-Stokes equations. In the paper [G-M] the existence and uniqueness of self-similar mild-solutions are shown in the frame of the Morrey-type spaces of measures in $\mathbb{R}^3$, solving firstly the problem for the vorticity. In [F-M-T] the reader can find some implicit comments about the meaning of the self-similar solutions in the study of large time behaviour. If the functions $\mathbf{v}(t,x)$ and $\mathrm{p}(t,x)$ are global solutions of system (1), it is not difficult to show that, for each number $\lambda > 0$, the functions $\lambda\mathbf{v}(\lambda^2 t, \lambda x)$ and $\lambda^2\mathrm{p}(\lambda^2 t, \lambda x)$ are solutions of the same system too.

Moreover, it is possible to characterize the self-similary condition in the following way. A vector field $\mathbf{v}(t,x)$ has the *self-similar property* if and only if there exists a vector field $\mathbf{V} = (V_1, V_2, \dots, V_n)$ such that $\mathbf{v}(t,x) = \mathbf{V}(x/\sqrt{t})/\sqrt{t}$, for all $x \in \mathbb{R}^n$ and all $t > 0$. In fact, when the field $\mathbf{V}$ exists, this last equality gives the definition for $\mathbf{v}(t,x)$, and it is straightforward to see that it is self-similar. Conversely, when the self-similar solution $\mathbf{v}(t,x)$ is given, we define $\mathbf{V}(x) = \mathbf{v}(1,x)$, for all $x \in \mathbb{R}^n$. Then, the self-similar condition on $\mathbf{v}$ turns out the expected equality between $\mathbf{v}$ and $\mathbf{V}$ choosing $\lambda = 1/\sqrt{t}$.

But, the problem of finding self-similar solutions is not evident. The initial velocity $\mathbf{v}_0(x)$ must be an homogeneous function of degree -1, if a self-similar solution exists. On $\mathbb{R}^3$, the typical elementary example of such functions is given by any linear combination of the vector fields

$$\Big(0, -\frac{x_3}{|x|^2}, \frac{x_2}{|x|^2}\Big), \ \Big(\frac{x_3}{|x|^2}, 0, -\frac{x_1}{|x|^2}\Big) \ \text{ and } \ \Big(-\frac{x_2}{|x|^2}, \frac{x_1}{|x|^2}, 0\Big)$$

which are homogeneous of degree -1 and verify $\nabla \cdot \mathbf{v} = 0$ (see [G-M], [C]). This fact produces a complicated situation. We would hope that the L^p-spaces might be the natural mathematical setting. However, it is trivial that there is not any homogeneous function of degree -1 (or any other degree) in L^p, for any power p. The existence and uniqueness of self-similar solutions for the problem (2) are proved in [C] for a certain family of Besov spaces.

Professor Y. Meyer suggested I should study the same problem in a much more natural frame what is the weak L^p spaces. So, our interest is particulary centered on the existence of global solutions $\mathbf{v}(t,x)$ of problem (2) in the space weak $L^n(\mathbb{R}^n)$, when the norm of the initial velocity is sufficiently small.

For each $0 < p < \infty$ the space weak $L^p(\mathbb{R}^n)$, denoted $L^{(p,\infty)}(\mathbb{R}^n)$ or shortly $L^{(p,\infty)}$, is the set of all the complex-valued Lebesgue measurable functions f defined on $\mathbb{R}^n$ such that exists a constant $A > 0$ satisfying

$$m\{x \in \mathbb{R}^n : \ |f(x)| > s\} \leq \frac{A}{s^p} , \tag{3}$$

for all $s > 0$. Here m denotes the Lebesgue measure on $\mathbb{R}^n$.

For $p = \infty$, the space weak $L^\infty(\mathbb{R}^n)$ will be $L^\infty(\mathbb{R}^n)$, and it will be denoted $L^{(\infty,\infty)}(\mathbb{R}^n)$, or simply $L^{(\infty,\infty)}$.

The space $L^{(n,\infty)}$ has the advantage to be one of the most natural extensions of $L^n(\mathbb{R}^n)$ that contains, besides, the homogeneous functions of degree -1. It is easy to see that the functions given above belong to $L^{(3,\infty)}(\mathbb{R}^3)$ but they are not included in $L^3(\mathbb{R}^3)$. This fact is the key that resolves the problem of finding self-similar solutions of the Navier-Stokes equations in a special sense to be explained later. This subject is one of the principal results presented here. All the main theorems will be stated in Section 2.

We shall need to make use of a class space which contains L^p and weak L^p spaces. That is, the Lorentz spaces $L^{(p,q)}$. In Section 3 we shall recall the definition, notation and some properties following the works [L], [H], [S-W] and [O]. At last, in Section 4, the proofs and some technical results will be given.

2. Our main theorems.

The aim of this section is to describe the principal results and their mathematical setting. First of all we need to specify the frame of problem (2), that is, the space of solutions of this system.

Definition. *Let $n > 1$ a positive integer number and let q be any fixed real number such that $n < q < \infty$. Let us define E the space of all the complex-valued functions $\mathbf{v}(t,x)$, with $t > 0$ and $x \in \mathbb{R}^n$, such that the following conditions are satisfied*

$$\mathbf{v}(t,x) \in C((0,\infty), L^{(n,\infty)}) , \tag{4.1}$$

$$t^{(1-n/q)/2}\mathbf{v}(t,x) \in C((0,\infty), L^{(q,\infty)}) , \tag{4.2}$$

(4.3) *the map $t \mapsto \mathbf{v}(t,\cdot)$ from $(0,\infty)$ to $L^{(n,\infty)}$ is continuous at the origin.*

Through this work C denotes the class of bounded and continuous function, and therefore the norm of $\mathbf{v}$ is naturally defined by

$$(5)\qquad \|\mathbf{v}\|_E := \sup_{t>0} t^{(1-n/q)/2} \|\mathbf{v}(t,x)\|_{(q,\infty)} + \sup_{t>0} \|\mathbf{v}(t,x)\|_{(n,\infty)} .$$

Once for all the space $L^{(n,\infty)}$ will be equipped with its canonical $\sigma(L^{(n,\infty)}, L^{(n',1)})$ topology and the symbol $\|\cdot\|_{(r,s)}$ denotes the norm on the Lorentz space $L^{(r,s)}(\mathbb{R}^n)$. (See Section 3.)

The expression (5) defines a norm on E, and the pair $(E, \|\cdot\|_E)$ is a Banach space. All this is a straightforward consequence of results of Section 3. In what follows we shall find global solutions of the problem (2) in E when the initial velocity $\mathbf{v}_0$ belongs to $L^{(n,\infty)}$ with sufficiently small norm, and verifies $\nabla \cdot \mathbf{v}_0 = 0$.

As usual, the problem (2) is written under the following integral form:

$$(6)\qquad \mathbf{v}(t,x) = S(t)\mathbf{v}_0(x) + B(\mathbf{v},\mathbf{v})(t,x),$$

where

$$(7)\qquad S(t)\mathbf{v}_0(x) = \mathbb{P}e^{-t\Delta}\mathbf{v}_0(x)$$

and

$$(8)\qquad B(\mathbf{u},\mathbf{v})(t,x) = -\int_0^t \mathbb{P}S(t-s)(\mathbf{u}(s,x)\cdot\nabla)\mathbf{v}(s,x)\,ds .$$

Any vector field $\mathbf{v}(t,x)$ satisfying the equality (6) will be called a "mild-solution" of the initial value problem (2).

The bilinear map B is defined by a Bochner integral. $\mathbb{P}$ denotes the orthogonal projection of $L^2(\mathbb{R}^n)$ onto the subspace $\mathbb{P}L^2(\mathbb{R}^n) = \{f \in L^2(\mathbb{R}^n) : \nabla \cdot f = 0\}$. As is well known (see, for instance [F-M]), $\mathbb{P}$ can be extended to a continuous operator on $L^p(\mathbb{R}^n)$ to $\mathbb{P}L^p(\mathbb{R}^n) = \{f \in L^p(\mathbb{R}^n) : \nabla \cdot f = 0\}$, for $1 < p < \infty$ (see [K]). Furthermore, it is trivial to extend $\mathbb{P}$ on $L^p(\mathbb{R}^n)$ to $\mathbb{P}L^{(p,\infty)}(\mathbb{R}^n) = \{f \in L^{(p,\infty)}(\mathbb{R}^n) : \nabla\cdot f = 0\}$ because $L^p(\mathbb{R}^n) = L^{(p,p)}(\mathbb{R}^n)$. (See Section 3.)

We notice that, since $\mathbb{P}$ commutes with the Laplacian Δ, the operator $S(t)$ agrees with the heat operator on functions in $\mathbb{P}L^{(p,\infty)}(\mathbb{R}^n)$.

We are now ready to state our main results.

Theorem 1. *Let* $\mathbf{v}_0$ *be any function in* $L^{(n,\infty)}(\mathbb{R}^n)$ *satisfying* $\nabla\cdot\mathbf{v}_0 = 0$ *in the distributional sense. Then, there exists a constant* $\delta > 0$ *such that if* $\|\mathbf{v}_0\|_{(n,\infty)} < \delta$ *the initial value problem* (2) *for the Navier-Stokes equations has, at least, a global mild-solution* $\mathbf{v}(t,x)$ *belonging to* E. *Furthermore, this solution satisfies* $\sup_{t>0} t^{(1-n/q)/2}\,\|\mathbf{v}(t,x)\|_{(q,\infty)} < \eta$ *where* $\eta = \eta(\delta) \to 0$ *with* δ.

Moreover, there exists a constant $\eta_0 > 0$ *such that if there is a solution* $\mathbf{v}(t,x)$ *satisfying* $\sup_{t>0} t^{(1-n/q)/2}\,\|\mathbf{v}(t,x)\|_{(q,\infty)} < \eta_0$, *then it is unique.*

Theorem 2. (Corollary). *Let* $\mathbf{v}_0 \in L^{(n,\infty)}(\mathbb{R}^n)$ *such that* $\nabla\cdot\mathbf{v}_0 = 0$ *in the distributional sense. It will be assumed that* $\mathbf{v}_0$ *is an homogeneous function of degree* -1, *that is,* $\mathbf{v}_0(\lambda x) = \lambda^{-1}\mathbf{v}_0(x)$ *for all* $x \in \mathbb{R}^n, x \neq 0$ *and all* $\lambda > 0$.

Then, there exists a constant $\delta > 0$ *such that if* $\|\mathbf{v}_0\|_{(n,\infty)} < \delta$ *a global mild-solution* $\mathbf{v}(t,x)$ *of the problem* (2) *of the Navier-Stokes equations in* E, *given by Theorem* 1, *satisfies*

$$\mathbf{v}(t,x) = \lambda\,\mathbf{v}(\lambda^2 t, \lambda x)\,, \tag{9}$$

for all $x \in \mathbb{R}^n$, *all* $t > 0$ *and all* $\lambda > 0$.

REMARK. We remind the reader that any global solution $\mathbf{v}(t,x)$ of the problem (2) having the property (9) is called "self-similar solution" of (2).

Theorem 3. (Corollary: Existence and uniqueness of self-similar solutions of Navier-Stokes equations in E). *Let* $\mathbf{v}_0$ *be any function defined on* $\mathbb{R}^n$ *which is homogeneous of degree* -1 *and satisfies* $\nabla\cdot\mathbf{v}_0 = 0$ *in the distributional sense. Furthermore, it will be supposed that the restriction of* $\mathbf{v}_0$ *to the unitary sphere* S^{n-1} *of* $\mathbb{R}^n$, *denoted* $\mathbf{v}_0|_{S^{n-1}}$, *belongs to* $L^n(S^{n-1})$.

Then, there exists a constant $\delta > 0$ *such that if* $\|\mathbf{v}_0|_{S^{n-1}}\|_{L^n(S^{n-1})} < \delta$, *the Cauchy problem* (2) *for the Navier-Stokes equations has, at least, a self-similar solution in* E.

As in Theorem 1, *this system admits a unique self-similar solution if such solution belongs to the ball*

$$\{\mathbf{v}(t,x) \in E : \sup_{t>0} t^{(1-n/q)/2}\,\|\mathbf{v}(t,x)\|_{(q,\infty)} < \eta_0\}\,,$$

for some $\eta_0 > 0$ *sufficiently small.*

3. Facts on Lorentz spaces.

In this section we shall present a brief summary of the definition of Lorentz spaces and the principal properties we are going to use.

Let us consider a non-atomic measure space $(X, \mathcal{M}, m)$. For each complex-valued, m-measurable function f defined on X, its distribution function is defined by

$$\lambda(s) = m\{x \in X : |f(x)| > s\}, \qquad \text{for } s > 0, \tag{10}$$

which is non-increasing and continuous from the right. This function has a "quasi-inverse", called the non-increasing rearrangement of f onto $(0, \infty)$, whose definition is

$$f^*(t) = \inf\{s > 0 : \lambda(s) \leq t\}, \qquad \text{for } t > 0. \tag{11}$$

It should be noticed that $f^*(t)$ is the true inverse function of $\lambda(t)$ when this function is continuous and strictly decreasing. It results that $f^*(t)$ is also continuous from the right and has the same distribution function as f.

Thus, the Lorentz space $L^{(p,q)}$ on $(X, \mathcal{M}, m)$ will be the collection of all the complex-valued, m-measurable functions f defined on X such that $\|f\|^*_{(p,q)} < \infty$, where

$$\|f\|^*_{(p,q)} = \begin{cases} \left(\dfrac{p}{q} \displaystyle\int_0^\infty (t^{1/p} f^*(t))^q \, \dfrac{dt}{t}\right)^{1/q}, & \text{if } 0<p<\infty, 0<q<\infty, \\ \sup\limits_{t>0} t^{1/p} f^*(t), & \text{if } 0<p\leq\infty, q=\infty. \end{cases} \tag{12}$$

The case $p = \infty$, $0 < q < \infty$ is out of interest since $\|f\|^*_{(\infty,q)} < \infty$ implies $f = 0$ almost everywhere.

We shall only recall the following two elementary properties for the Lorentz spaces.

a) For any $p > 0$ and any q and r such that $0 < q \leq r \leq \infty$ we have

$$\|f\|^*_{(p,r)} \leq \|f\|^*_{(p,q)}, \tag{13}$$

and thus

$$L^{(p,q)} \subset L^{(p,r)}. \tag{14}$$

b) For any p such that $1 \leq p \leq \infty$ we have that $\|f\|^*_{(p,p)}$ is the usual L^p-norm, and then $L^{(p,p)} = L^p(X, \mathcal{M}, m)$.

Combining these two properties, we obtain in particular that for $0 < p \leq \infty$, $L^{(p,p)} = L^p \subset L^{(p,\infty)}$ with

$$\|f\|^*_{(p,\infty)} \leq \|f\|_p \, . \tag{15}$$

However, since the triangle inequality fails in general, $\|f\|^*_{(p,q)}$ is not a norm for $p < q$.

In view to build an "adequate" norm for the Lorentz spaces, it is necessary to define the function:

$$f^{**}(t) = \frac{1}{t} \int_0^t f^*(y)\, dy \,, \qquad \text{for} \quad t > 0 \, . \tag{16}$$

This function is closely related with the Hardy-Littlewood maximal function of f because it can be expressed by ([H])

$$f^{**}(t) = \sup_{\substack{E \in \mathcal{M} \\ m(E) \geq t}} \frac{1}{m(E)} \int_E |f(x)|\, dx \, . \tag{17}$$

Hence, the norm $\|f\|_{(p,q)}$ is defined as:

$$\|f\|_{(p,q)} = \begin{cases} \left(\dfrac{p}{q} \displaystyle\int_0^\infty (t^{1/p} f^{**}(t))^q \frac{dt}{t} \right)^{1/q}, & \text{if } 1 < p < \infty, 1 \leq q < \infty, \\ \sup\limits_{t>0} t^{1/p} f^{**}(t) \, , & \text{if } 1 < p \leq \infty \, , \ q = \infty \, . \end{cases} \tag{18}$$

These spaces $L^{(p,q)}$ with the norm $\|f\|_{(p,q)}$ are Banach spaces. In some sense $\|f\|_{(p,q)}$ is equivalent to $\|f\|^*_{(p,q)}$

$$\|f\|^*_{(p,q)} \leq \|f\|_{(p,q)} \leq \frac{p}{p-1} \|f\|^*_{(p,q)} \, , \tag{19}$$

for $1 < p < \infty$, $1 \leq q < \infty$. We observe that in definition (18) the case $p = 1$ has been excluded; although both expressions make sense, they do not define a norm. In this case property (19), or a similar one, is not valid. Besides, for $p = 1$ the second case in (18) would provide

$$\|f\|_{(1,\infty)} = \|f\|_1 \, , \tag{20}$$

which is inadequate.

Since expression (19), it will be sufficient to estimate $\|f\|^*_{(p,q)}$ instead of $\|f\|_{(p,q)}$, which is more complicated to manipulate.

The reader who is interested in the Lorentz spaces $L^{(p,q)}$ and their properties is referred to [S-W, chapter V], and to [L] and [H] for more details.

In [O] estimates for the product operators and convolution operators are given. From there we extract the following definitions and results.

Given three measure spaces (X,μ), $(\bar{X},\bar{\mu})$ and (Y,ν), a bilinear operator T, which maps complex-valued measurable functions on X and $\bar{X}$ into complex-valued measurable functions on Y, is called a "convolution operator" if

$$\|T(f,g)\|_1 \le \|f\|_1 \, \|g\|_1 \,, \tag{21.a}$$

$$\|T(f,g)\|_\infty \le \|f\|_1 \, \|g\|_\infty \,, \tag{21.b}$$

$$\|T(f,g)\|_\infty \le \|f\|_\infty \, \|g\|_1 \,. \tag{21.c}$$

As usual, $\| \,.\, \|_1$ and $\| \,.\, \|_\infty$ are respectively the L^1-norm and the L^∞-norm on any of the three spaces X, $\bar{X}$, Y.

Proposition 1 ([O]) (Generalized Young's inequality). *Let $1 < p_1, p_2, r < \infty$. If T is a convolution operator and if $f \in L^{(p_1,q_1)}$, $g \in L^{(p_2,q_2)}$ where $1/p_1 + 1/p_2 > 1$, then $h = T(f,g)$ belongs to $L^{(r,s)}$ where $1/r = 1/p_1 + 1/p_2 - 1$, and $s \ge 1$ is any number such that $1/q_1 + 1/q_2 \ge 1/s$.*

Moreover,

$$\|h\|_{(r,s)} \le 3\,r\,\|f\|_{(p_1,q_1)}\,\|g\|_{(p_2,q_2)} \,. \tag{22}$$

Otherwise, a bilinear operator P, which maps complex-valued measurable functions on X and $\bar{X}$ into complex-valued measurable functions on Y, is called a "product operator" if

$$\|P(f,g)\|_\infty \le \|f\|_\infty \, \|g\|_\infty \,, \tag{23.a}$$

$$\|P(f,g)\|_1 \le \|f\|_1 \, \|g\|_\infty \,, \tag{23.b}$$

$$\|P(f,g)\|_1 \le \|f\|_\infty \, \|g\|_1 \,. \tag{23.c}$$

Proposition 2 ([O]) (Generalized Hölder's inequality). *If P is a product operator, $h = P(f,g)$, and if $f \in L^{(p_1,q_1)}$, $g \in L^{(p_2,q_2)}$ where $1/p_1 + 1/p_2 < 1$, then $h \in L^{(r,s)}$ where $1/r = 1/p_1 + 1/p_2$, and $s \geq 1$ is any number such that $1/q_1 + 1/q_2 \geq 1/s$.*

Moreover,

$$\|h\|_{(r,s)} \leq r' \, \|f\|_{(p_1,q_1)} \, \|g\|_{(p_2,q_2)} \,, \tag{24}$$

being r' the conjugate index of r.

Before leaving this section, we extract the following proposition from Hunt's paper [H] about the duality of Lorentz spaces.

Proposition 3 ([H]). *The conjugate space of $L^{(p,1)}$ is $L^{(p',\infty)}$, where $1/p + 1/p' = 1$.*

The conjugate space of $L^{(p,q)}, 1 < p < \infty, 1 < q < \infty$, is $L^{(p',q')}$, where $1/p + 1/p' = 1$, $1/q + 1/q' = 1$, and hence, these spaces are reflexive.

REMARK. In the same way that L^1 is not the conjugate space of L^∞, $L^{(p,1)}$ is not the conjugate space of $L^{(p',\infty)}$. ([H]).

4. Proofs of the Theorems.

In this last section we shall develope the proofs of the stated in Section 2, and some intermediate results.

First of all, we begin reminding that the proyection $\mathbb{P}$ commutes with the Laplacian Δ, and then the operator $S(t)$, defined by (7), is essentially the heat operator. This fact allows us to estimate $S(t)$ without taking into account the operator $\mathbb{P}$.

The Weierstrass kernel, or heat kernel, is given by

$$w_t(x) = \left(\frac{1}{4\pi t}\right)^{n/2} e^{-|x|^2/(4t)} \,, \tag{25}$$

for all $t > 0$ and all $x \in \mathbb{R}^n$.

Thus, we can write

$$S(t)\mathrm{v}_0(x) = \mathbb{P}(w_t * \mathrm{v}_0)(x) \,, \tag{26}$$

and

$$(27)\qquad B(\mathbf{u},\mathbf{v})(t,x) = -\int_0^t \mathbb{P}\big(w_{t-s} * ((\mathbf{u}(s)\cdot\nabla)\mathbf{v}(s))\big)(x)\,ds\,.$$

Here $*$ denotes the usual convolution of functions.

We begin giving a meaning to the right hand side of (26) because, in our case, $\mathbf{v}_0$ is a distribution in $L^{(n,\infty)}(\mathbb{R}^n)$ which is not reflexive.

The operator $S(t)(\varphi) = \mathbb{P}(w_t * \varphi)$ is well defined for all functions φ in $\mathcal{S} = \mathcal{S}(\mathbb{R}^n)$ the Schwartz class. These functions are dense in each Lorentz space $L^{(p,q)}$ with $1 < p < \infty$ and $1 \le q < \infty$. Then, we can extend the operator $S(t)$ to these Lorentz spaces and this extension will be denoted $S(t)$ too.

Let us remember that $L^{(p,\infty)} = (L^{(p',1)})^*$ for any $1 \le p < \infty$, where p' is the conjugate index of p (see Proposition 3). So, by duality, we can define the formal transposed operator of $S(t)$ on each $L^{(p,\infty)}$, which we shall also denote $S(t)$. We shall deduce some continuity properties of this operator.

Lemma 1. *For any $t > 0$, any r and any p such that $1 < p \le r < \infty$ the operator*

$$(28)\qquad S(t) : L^{(p,\infty)}(\mathbb{R}^n) \longrightarrow L^{(r,\infty)}(\mathbb{R}^n)$$

is weak-star continuous.

In order to see this we must show the following continuity property of $S(t)$

$$(29)\qquad S(t) : L^{(r',1)} \longrightarrow L^{(p',1)}$$

is strong continuous for all $t > 0$. Moreover,

$$(30)\qquad \|S(t)\varphi\|_{(p',1)} \le c(n,p,r)\, t^{-(1/r'-1/p')n/2}\, \|\varphi\|_{(r',1)}\,.$$

Taking into account that $S(t)$ is essentially the convolution with w_t, the inequality (30), for the case $1 < p < r < \infty$, results directly from the generalized Young's inequality (22) and from the evident estimate

$$\|w_t\|_{(q_1,q_2)} \le c(n,q_1,q_2)\, t^{-(1-1/q_1)n/2} = c(n,p,r)\, t^{-(1/r'-1/p')n/2}\,,$$

where $q_1 > 1$ and $q_2 \geq 1$ must verify $1/p' = 1/r' + 1/q_1 - 1$ and $1 \leq 1 + 1/q_2$.

When $p = r > 1$ we obtain (30) from an elementary result over the homogeneous Banach space $L^{(p',1)}(\mathbb{R}^n)$ (see [Kn]).

Clearly, (30) implies (29). Then, a simple duality argument leads us to

$$\|S(t)f\|_{(r,\infty)} \leq c(n,p,r)\,\|f\|_{(p,\infty)}\, t^{-(1/p-1/r)n/2}\,.$$

completing the proof of Lemma 1.

We shall pass to study the behaviour of the operator $S(t)$ near the value $t = 0$. This fact will be closely related to the weak continuity at $t = 0$ of the mild-solution of the problem (2).

Once more we notice that the operator $S(t)$ is essentially the convolution with the heat kernel w_t so, taking into account that $w_t(x) \to 0$, when $t \to 0$, for all $x \in \mathbb{R}^n, x \neq 0$, one could hope that $S(t)$ tends to I, the identity operator. It is not difficult to obtain this convergence in the norm $\|\cdot\|_{(p,q)}$ when $S(t)$ acts on $L^{(p,q)}$ with $1 < p < \infty$, $1 < q < \infty$, that means, in the case that $L^{(p,q)}$ is a reflexive space (Proposition 3). But, unfortunately in our problem, the initial velocity $\mathbf{v}_0(x)$ belongs to $L^{(n,\infty)}$ which is not reflexive. However, for each $\varphi \in L^{(n',1)}$ we have by duality

$$|\langle S(t)\mathbf{v}_0 - \mathbf{v}_0, \varphi\rangle| = |\langle \mathbf{v}_0, S(t)\varphi - \varphi\rangle| \leq \|\mathbf{v}_0\|_{(n,\infty)}\,\|S(t)\varphi - \varphi\|_{(n',1)}$$

which tends to 0 with t. Hence, we have proved the following lemma.

Lemma 2. *If $\mathbf{v}_0 \in L^{(n,\infty)}$, then*

$$S(t)\mathbf{v}_0 \rightharpoonup \mathbf{v}_0\,, \quad \textit{when} \quad t \to 0^+\,.$$

In the next lemma our purpose will be the study of the continuity of the operator $S(t)$ as a function of t.

Lemma 3. *Let $f \in L^{(n,\infty)}(\mathbb{R}^n)$.*

a) *The function $t \mapsto S(t)f$ from $(0,\infty)$ to $L^{(n,\infty)}(\mathbb{R}^n)$ is continuous.*

b) *The function $t \mapsto t^{(1-n/q)/2}\, S(t)f$ from $(0,\infty)$ to $L^{(q,\infty)}(\mathbb{R}^n)$ is continuous.*

This result will be a consequence of the following estimate we shall suppose true for the moment. Hence we assume

$$\|\Delta S(t)f\|_{(r,\infty)} \le c(n,p,r)\,t^{-1-(1/p-1/r)n/2}\,\|f\|_{(p,\infty)}\,, \tag{31}$$

for $1 < p \le r < \infty$. Then, taking into account the mean-value theorem and

$$\frac{\partial}{\partial t}(S(t)f) = -\Delta S(t)f\,,$$

we obtain part a) from (31) giving, with $p = r = n$,

$$\left\|\frac{\partial}{\partial t}(S(t)f)\right\|_{(n,\infty)} \le \frac{c(n)}{t}\,\|f\|_{(n,\infty)}\,,$$

which is bounded in a neighbourhood of each $t > 0$.

Analogously, for part b) we apply again the mean-value theorem to evaluate the difference

$$\left\|t^{(1-n/q)/2}\,S(t)f - (t')^{(1-n/q)/2}\,S(t')f\right\|_{(q,\infty)}\,.$$

Thus it will be sufficient to estimate the derivative

$$\begin{aligned}\frac{\partial}{\partial t}\left(t^{(1-n/q)/2}\,S(t)f\right) &= \frac{1}{2}\left(1-\frac{n}{q}\right)t^{-(1+n/q)/2}S(t)f \\ &\quad + t^{(1-n/q)/2}\,\frac{\partial}{\partial t}\,(S(t)f)\,.\end{aligned} \tag{32}$$

Thanks to Lemma 1, the first term of the right hand side of (32) can be estimated by

$$\begin{aligned}&\left\|\frac{1}{2}\left(1-\frac{n}{q}\right)t^{-(1+n/q)/2}\,S(t)f\right\|_{(q,\infty)} \\ &\qquad \le c_1(n,q)\,t^{-(1+n/q)/2-(1-n/q)/2}\,\|f\|_{(n,\infty)} \\ &\qquad = c_1(n,q)\,t^{-1}\,\|f\|_{(n,\infty)}\,,\end{aligned} \tag{33}$$

where we have taken $p = n$ and $r = q$.

For the second term of the right hand side of (32) we use inequality (31) with $p = n$ and $r = q$ producing

$$\left\|t^{(1-n/q)/2}\,\frac{\partial}{\partial t}\,(S(t)f)\,\right\|_{(q,\infty)} \le c_2(n,q)\,t^{-1}\,\|f\|_{(n,\infty)}\,. \tag{34}$$

Finally, we have from (33), (34) and (32), that

$$\left\|\frac{\partial}{\partial t}\left(t^{(1-n/q)/2}\, S(t)f\right)\right\|_{(q,\infty)} \leq \frac{c(n,q)}{t}\, \|f\|_{(n,\infty)}\ ,$$

completing part b).

It remains to prove inequality (31). In the same way as we have done in Lemma 1, we must show that

$$\|\Delta S(t)\varphi\|_{(p',1)} \leq c(n,p,r)\, t^{-1-(1/r'-1/p')n/2}\, \|\varphi\|_{(r',1)}\ , \tag{35}$$

for all $\varphi \in L^{(r',1)}$. It will be enough to see (35) for any testing function φ.

Then, it can be rapidly seen that

$$\Delta S(t)\varphi = u_t^{(1)} * \varphi + u_t^{(2)} * \varphi\,, \tag{36}$$

where

$$u_t^{(1)}(x) = -\frac{n}{2t}\, w_t(x) \qquad \text{and} \qquad u_t^{(2)}(x) = \frac{|x|^2}{4t^2}\, w_t(x)\,.$$

Both terms will be treated separately. The first one is just equal to $(c/t)\,(w_t * \varphi)(x)$, and then, as in Lemma 1, we have

$$\left\|u_t^{(1)} * \varphi\right\|_{(p',1)} \leq \tilde{c}(n,r,p)\, t^{-1-(n/p-n/r)/2}\, \|\varphi\|_{(r',1)}\ , \tag{37}$$

for $1 < p \leq r < \infty$.

For the second term $(u_t^{(2)} * \varphi)(x)$ we can follow the way sketched in Lemma 1. Thus, we obtain

$$\left\|u_t^{(2)} * \varphi\right\|_{(p',1)} \leq \tilde{c}(n,r,p)\, t^{-1-(n/r'-n/p')/2}\, \|\varphi\|_{(r',1)}\ . \tag{38}$$

Finally, applying triangular inequality in $L^{(p',1)}$ to (36) and taking into account estimates (37) and (38) we get (35). Then, (31) is deduced from (38) by duality and therefore Lemma 3 is proved.

We can now state a result concerning the continuity of the operator $S(t)$ which collects the properties given in Lemmas 1, 2 and 3.

Proposition 4. *Let $t > 0$. The operator $S(t) : L^{(n,\infty)}(\mathbb{R}^n) \longrightarrow E$ is continuous.*

Next, we go on with the bilinear operator B and their continuity properties. We need to define two auxiliar spaces F_1 and F_2. Let F_1 be the set of all the complex-valued functions $\mathbf{v}(t,x)$, $(t,x) \in (0,\infty) \times \mathbb{R}^n$, such that $\mathbf{v}(t,\cdot) \in L^{(q,\infty)}(\mathbb{R}^n)$ and the number

$$\sup_{t>0} t^{(1-n/q)/2} \|\mathbf{v}(t)\|_{(q,\infty)}$$

is finite.

Analogously, let F_2 be the set of all the complex-valued functions $\mathbf{v}(t,x)$, $(t,x) \in (0,\infty) \times \mathbb{R}^n$, such that $\mathbf{v}(t,\cdot) \in L^{(n,\infty)}(\mathbb{R}^n)$ and the number $\sup_{t>0} \|\mathbf{v}(t,\cdot)\|_{(n,\infty)}$ is finite.

Lemma 4. *The bilinear operator B is continuous from $F_1 \times F_1$ to F_1 and from $F_1 \times F_1$ to F_2 too. More precisely, there exists a constant $\mathcal{K} > 0$, depending only on n and q, such that for all pair of functions $\mathbf{u}$ and $\mathbf{v}$ in E the following estimates are satisfied:*

$$\sup_{t>0} \|B(\mathbf{u},\mathbf{v})(t)\|_{(n,\infty)} \leq \frac{\mathcal{K}}{2}\Big(\sup_{t>0} t^{(1-n/q)/2} \|\mathbf{u}(t)\|_{(q,\infty)}\Big) \cdot \Big(\sup_{t>0} t^{(1-n/q)/2} \|\mathbf{v}(t)\|_{(q,\infty)}\Big), \tag{39}$$

and

$$\sup_{t>0} t^{(1-n/q)/2} \|B(\mathbf{u},\mathbf{v})(t)\|_{(q,\infty)} \leq \frac{\mathcal{K}}{2}\Big(\sup_{t>0} t^{(1-n/q)/2} \|\mathbf{u}(t)\|_{(q,\infty)}\Big) \cdot \Big(\sup_{t>0} t^{(1-n/q)/2} \|\mathbf{v}(t)\|_{(q,\infty)}\Big). \tag{40}$$

We are going to deduce the estimates (39) and (40) on each component of the matrix operator $B(\mathbf{u},\mathbf{v})(t,x)$. When the fields $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ and $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ are mild-solutions of (2), they must satisfy $\nabla \cdot \mathbf{u} = \nabla \cdot \mathbf{v} = 0$. Consequently,

$$\begin{aligned}(\mathbf{u} \cdot \nabla)\mathbf{v} &= \nabla(\mathbf{u} \otimes \mathbf{v}) \\ &= \Big(\sum_{j=1}^{n} \frac{\partial}{\partial x_j}(u_j v_1), \sum_{j=1}^{n} \frac{\partial}{\partial x_j}(u_j v_2), \ldots, \sum_{j=1}^{n} \frac{\partial}{\partial x_j}(u_j v_n)\Big).\end{aligned}$$

Then, putting this in the definition (8) of B, and after an integration by parts, it is not difficult to see that each component of B, denoted $b(f,g)(t)$, can be expressed as

$$b(f,g)(t) = c_n \int_0^t (t-s)^{-(n+1)/2}\, \Theta\Big(\frac{\cdot}{2\sqrt{t-s}}\Big) * (fg)(s)\, ds\,, \tag{41}$$

being $\Theta(z)$ a smooth function which is $O(|z|^{-n-1})$ when $|z| \to \infty$, and c_n is a constant depending only on n. The symbol $*$ means the convolution in the variable x.

The totality of the Lemma 5 will be obtained from a unique estimate. Let $r \geq n$, frozen for the moment. According to Proposition 1, we infer from (41) that

$$\begin{aligned} &\|b(f,g)(t)\|_{(r,\infty)} \\ &\leq c_n \int_0^t (t-s)^{-(n+1)/2} \Big\|\Theta\Big(\frac{\cdot}{2\sqrt{t-s}}\Big) * (fg)(s)\Big\|_{(r,\infty)} ds \\ &\leq c(n,r) \int_0^t (t-s)^{-(n+1)/2} \Big\|\Theta\Big(\frac{\cdot}{2\sqrt{t-s}}\Big)\Big\|_p \|f(s)\, g(s)\|_{(q/2,\infty)}\, ds\,, \end{aligned} \tag{42}$$

with $1/r = 1/p + 2/q - 1$.

Now, due to Proposition 2, we write

$$\|f(s)\, g(s)\|_{(q/2,\infty)} \leq \frac{q}{q-2}\, \|f(s)\|_{(q,\infty)}\, \|g(s)\|_{(q,\infty)}\,, \tag{43}$$

where $q/(q-2)$ is the conjugate index of $q/2$.

Otherwise, after a simple change of variables, we have

$$\Big\|\Theta\Big(\frac{\cdot}{2\sqrt{t-s}}\Big)\Big\|_p = c(p)\, \|\Theta\|_p\, \big(2\sqrt{t-s}\big)^{n/p}\,. \tag{44}$$

Replacing (43) and (44) in (42), we get

$$\begin{aligned} \|b(f,g)(t)\|_{(r,\infty)} \leq c(n,r,p,q)\, \|\Theta\|_p & \\ \cdot \Big(\int_0^t (t-s)^{-(n+1)/2+n/(2p)}\, s^{-(1-n/q)}\, ds\Big)& \\ \cdot \Big(\sup_{0<t<T_0} t^{(1-n/q)/2}\, \|f(t)\|_{(q,\infty)}\Big)& \\ \cdot \Big(\sup_{0<t<T_0} t^{(1-n/q)/2}\, \|g(t)\|_{(q,\infty)}\Big)&\,, \end{aligned} \tag{45}$$

for all $T_0 > t$, including $T_0 = \infty$. Notice that the integral on s in (45) is convergent since the numbers $q > n > 1$ are fixed parameters and the choice of p gives $1/p > (n-1)/n$, and then it is evident that the exponent $-(n+1)/2 + n/(2p) > -1$.

The very well known relation between Beta and Gamma functions yields

$$\int_0^t (t-s)^{-(n+1)/2+n/(2p)}\, s^{-(1-n/q)}\, ds = c(n,r,q)\, t^{-(1-n/r)/2}\,, \tag{46}$$

where

$$c(n,r,q) = \frac{\Gamma\Big(\frac{1}{2}+\frac{n}{2r}-\frac{n}{q}\Big)\,\Gamma\Big(\frac{n}{q}\Big)}{\Gamma\Big(\frac{1}{2}+\frac{n}{2r}\Big)}\,.$$

Case i): Taking $r = n$ and setting (46) in (45), it results

$$\begin{aligned} \|b(f,g)(t)\|_{(n,\infty)} \leq c_1(n,q) \Big(\sup_{0<t<T_0} t^{(1-n/q)/2}\, \|f(t)\|_{(q,\infty)} \Big) \\ \cdot \Big(\sup_{0<t<T_0} t^{(1-n/q)/2}\, \|g(t)\|_{(q,\infty)} \Big)\,. \end{aligned} \tag{47}$$

In order to obtain (39), we take the supremum over all $t > 0$ in (47) and choose $T_0 = \infty$.

Case ii): Considering $r = q$, and following the same proceeding, we arrive to

$$\begin{aligned} \|b(f,g)(t)\|_{(q,\infty)} \leq c_2(n,q)\, t^{-(1-n/q)/2} \Big(\sup_{0<t<T_0} t^{(1-n/q)/2}\, \|f(t)\|_{(q,\infty)} \Big) \\ \cdot \Big(\sup_{0<t<T_0} t^{(1-n/q)/2}\, \|g(t)\|_{(q,\infty)} \Big)\,. \end{aligned} \tag{48}$$

In the same way as before we get (40).

The continuity properties of B are direct consequences of the estimates (39) and (40), completing the proof of Lemma 4.

Another property on the continuity of the bilinear operator B is given in the following result.

Lemma 5. *Let* $\mathbf{u}(t,x)$ *and* $\mathbf{v}(t,x)$ *belong to* E*. Then,*

a) *The function* $t \mapsto B(\mathbf{u},\mathbf{v})(t,x)$ *from* $(0,\infty)$ *to* $L^{(n,\infty)}(\mathbb{R}^n)$ *is continuous.*

b) *The function* $t \mapsto B(\mathbf{u},\mathbf{v})(t,x)$ *from* $(0,\infty)$ *to* $L^{(q,\infty)}(\mathbb{R}^n)$ *is continuous.*

c) *The function* $t \mapsto t^{(1-n/q)/2}\, B(\mathbf{u},\mathbf{v})(t,x)$ *from* $(0,\infty)$ *to* $L^{(q,\infty)}(\mathbb{R}^n)$ *is continuous.*

We are going to show parts a) and b). The third one is a consequence of b).

Let $t > 0$ and h be sufficiently small (take $|h| < t/2$). We suppose $h > 0$. The case $h < 0$ can be analogously treated. We evaluate the difference of B between t and $t+h$. Then we write

$$\begin{aligned}
B(\mathbf{u},\mathbf{v})&(t+h,x) - B(\mathbf{u},\mathbf{v})(t,x)\\
&= \int_0^{t+h} S(t+h-s)\,\nabla(\mathbf{u}(s,x)\otimes\mathbf{v}(s,x))\,ds\\
&\quad - \int_0^{t} S(t-s)\,\nabla(\mathbf{u}(s,x)\otimes\mathbf{v}(s,x))\,ds\\
&= \int_t^{t+h} S(t+h-s)\,\nabla(\mathbf{u}(s,x)\otimes\mathbf{v}(s,x))\,ds\\
&\quad + \int_0^{t} (S(t+h-s)-S(t-s))\,\nabla(\mathbf{u}(s,x)\otimes\mathbf{v}(s,x))\,ds\\
&:= B^{(1)}(h) + B^{(2)}(h)\,.
\end{aligned}$$

We have to see

$$\|B^{(1)}(h)\|_{(n,\infty)} \to 0\,, \qquad \text{when } h \to 0\,, \tag{49}$$

$$\|B^{(1)}(h)\|_{(q,\infty)} \to 0\,, \qquad \text{when } h \to 0\,, \tag{50}$$

$$\|B^{(2)}(h)\|_{(n,\infty)} \to 0\,, \qquad \text{when } h \to 0\,, \tag{51}$$

$$\|B^{(2)}(h)\|_{(q,\infty)} \to 0\,, \qquad \text{when } h \to 0\,. \tag{52}$$

Let $r \geq n$. As in Lemma 4, denoting $b^{(1)}(f,g)(h)$ the entries of $B^{(1)}(h)$ and after applying generalized Young and Hölder inequalities, we arrive to

$$
(53)\qquad
\begin{aligned}
&\|b^{(1)}(f,g)(h)\|_{(r,\infty)} \\
&\leq c(n,r,p,q)\int_t^{t+h}(t+h-s)^{-(n+1)/2+n/(2p)}\,s^{-(1-n/q)}\,ds \\
&\quad\cdot\Big(\sup_{t>0} t^{(1-n/q)/2}\,\|f\|_{(q,\infty)}\Big) \\
&\quad\cdot\Big(\sup_{t>0} t^{(1-n/q)/2}\,\|g\|_{(q,\infty)}\Big),
\end{aligned}
$$

where $1/p = 1/r + 1 - 2/q$. A simple computation (by homogeneity) gives

$$
(54)\qquad
\begin{aligned}
&\int_t^{t+h}(t+h-s)^{-(n+1)/2+n/(2p)}\,s^{-(1-n/q)}\,ds \\
&\qquad\qquad = c_1(n,q,r)\,t^{-(1-n/q)}\,h^{n/(2r)-n/q+1/2},
\end{aligned}
$$

for all $r \geq n$ such that $n/(2r) - n/q + 1/2 > 0$. As it was noticed in the proof of Lemma 4, the last integral is finite. If we take $r = n$, we have from (53) and (54)

$$
\begin{aligned}
\|b^{(1)}(f,g)(h)\|_{(n,\infty)} \leq \tilde{c}(n,q)\,t^{-(1-n/q)}\,h^{1-n/q}
&\cdot\Big(\sup_{t>0} t^{(1-n/q)/2}\,\|f\|_{(q,\infty)}\Big) \\
&\cdot\Big(\sup_{t>0} t^{(1-n/q)/2}\,\|g\|_{(q,\infty)}\Big),
\end{aligned}
$$

and then $\|B^{(1)}(h)\|_{(n,\infty)} \to 0$ with h.

Otherwise, choosing $r = q$ we obtain from (53) and (54)

$$
\begin{aligned}
\|b^{(1)}(f,g)(h)\|_{(q,\infty)} \leq \tilde{\tilde{c}}(n,q)\,t^{-(1-n/q)}\,h^{(1-n/q)/2}
&\cdot\Big(\sup_{t>0} t^{(1-n/q)/2}\,\|f\|_{(q,\infty)}\Big) \\
&\cdot\Big(\sup_{t>0} t^{(1-n/q)/2}\,\|g\|_{(q,\infty)}\Big),
\end{aligned}
$$

yielding $\|B^{(1)}(h)\|_{(q,\infty)} \to 0$ with h. Thus, we have just proved (49) and (50).

In order to see (51) and (52) we take again the entries of the operator $B^{(2)}(h)$ which we denote by $b^{(2)}(f,g)(h)$. Hence, using the same notation as in Lemma 5 we have

$$\begin{aligned} b^{(2)}(f,g)(h) = c_n \int_0^t \Big((t+h-s)^{-(n+1)/2}\, \Theta\Big(\frac{\cdot}{2\sqrt{t+h-s}}\Big) \\ - (t-s)^{-(n+1)/2}\, \Theta\Big(\frac{\cdot}{2\sqrt{t-s}}\Big)\Big) * (fg)(s)\, ds\,. \end{aligned} \tag{55}$$

Let $r \geq n$. We are interested on $r = n$ and $r = q$. It is easy to show that the expression under the integral sign in (55) has bounded $L^{(r,\infty)}$-norm. Indeed, if $1/r = 1/p + 2/q - 1$, after applying generalized Young and Hölder inequalities we obtain

$$\begin{aligned} &\Big\| \Big((t+h-s)^{-(n+1)/2} \Theta\Big(\frac{\cdot}{2\sqrt{t+h-s}}\Big) \\ &\quad - (t-s)^{-(n+1)/2}\, \Theta\Big(\frac{\cdot}{2\sqrt{t-s}}\Big)\Big) * (fg)(s) \Big\|_{(r,\infty)} \\ &\qquad \leq c(n,r,q) \Big\| \Big((t+h-s)^{-(n+1)/2}\, \Theta\Big(\frac{\cdot}{2\sqrt{t+h-s}}\Big) \\ &\qquad\qquad - (t-s)^{-(n+1)/2}\, \Theta\Big(\frac{\cdot}{2\sqrt{t-s}}\Big)\Big) \Big\|_p \|(fg)(s)\|_{(q/2,\infty)} \\ &\qquad \leq \tilde{c}(n,r,q) \Big\| \Big((t+h-s)^{-(n+1)/2}\, \Theta\Big(\frac{\cdot}{2\sqrt{t+h-s}}\Big) \\ &\qquad\qquad - (t-s)^{-(n+1)/2}\, \Theta\Big(\frac{\cdot}{2\sqrt{t-s}}\Big)\Big) \Big\|_p \\ &\qquad\quad \cdot \|f(s)\|_{(q,\infty)}\, \|g(s)\|_{(q,\infty)}\,. \end{aligned} \tag{56}$$

Taking into account (44) and reminding that $h > 0$, we can write

$$\begin{aligned} &\Big\| \Big((t+h-s)^{-(n+1)/2}\, \Theta\Big(\frac{\cdot}{2\sqrt{t+h-s}}\Big) \\ &\quad - (t-s)^{-(n+1)/2}\, \Theta\Big(\frac{\cdot}{2\sqrt{t-s}}\Big)\Big) * (fg)(s) \Big\|_{(r,\infty)} \\ &\qquad \leq c'(n,r,q) \Big((t+h-s)^{-(n+1)/2+n/(2p)} + (t-s)^{-(n+1)/2+n/(2p)} \Big) \\ &\qquad\quad \cdot \|\Theta\|_p\, \|f(s)\|_{(q,\infty)}\, \|g(s)\|_{(q,\infty)} \end{aligned}$$

$$
\begin{aligned}
&\le c''(n,r,q)\,(t-s)^{-(n+1)/2+n/(2p)}\, s^{-(1-n/q)}\, \|\Theta\|_p \\
&\quad\cdot \Big(\sup_{t>0} t^{(1-n/q)/2}\, \|f(s)\|_{(q,\infty)} \Big) \\
&\quad\cdot \Big(\sup_{t>0} t^{(1-n/q)/2}\, \|g(s)\|_{(q,\infty)} \Big)\,,
\end{aligned}
$$

which, due to $-(n+1)/2 + n/(2p) > -1$ once more, belongs to $L^1([0,t], ds)$ and it is independent of h. It remains to prove that the function under the integral sign in (55) tends to zero with h in the $L^{(r,\infty)}$-norm. But from (56) this is evident since the continuity on t of the function $(t-s)^{-(n+1)/2}\,\Theta(\,\cdot\,/2\sqrt{t-s})$ from $(0,\infty)$ to L^p and the behaviour of $\Theta(z)$ for $|z| \to +\infty$. Thus, thanks to the Lebesgue dominated convergence theorem we get $\lim_{h\to 0^+} \|b^{(2)}(f,g)(h)\|_{(r,\infty)} = 0$. In particular we have shown (51) and (52). Then, the proof of Lemma 5 is complete.

Lemma 6. *If* $\mathbf{u}$ *and* $\mathbf{v}$ *belong to* E*, then*

$$
(57) \qquad B(\mathbf{u},\mathbf{v})(t,x) \rightharpoonup 0\,, \qquad \text{when } t \to 0^+\,.
$$

As it was noticed in Lemma 4, it is not difficult to see that if $u_j(s)$ and $v_k(s)$ denote respectively the j-th and the k-th components of the vectors $\mathbf{u}(s)$ and $\mathbf{v}(s)$, then $B_k(\mathbf{u},\mathbf{v})(t,x)$, the k-th component of $B(\mathbf{u},\mathbf{v})(t,x)$, is written as

$$
(58) \qquad B_k(\mathbf{u},\mathbf{v})(t,x) = \sum_{j=1}^{n} \int_0^t Q_{jk}(t-s)\,(u_j(s)v_k(s))(x)\,ds\,,
$$

where $Q_{jk}(t-s)$ are pseudodifferential operators with symbols

$$
(59) \qquad \sigma(Q_{jk}(t-s))(\xi) = i\,\xi_j \left(\delta_{jk} - \frac{\xi_j \xi_k}{|\xi|^2} \right) e^{-(t-s)|\xi|^2}
$$

for $j,k = 1,2,\ldots,n$. Here, δ_{jk} denotes the Kronecker's delta function and $i^2 = -1$.

More generally, taking into account (58) and (59), we are going to consider a pseudodifferential operator $Q(t-s)$ on $\mathbb{R}^n$ whose symbol has the form $\sigma(Q(t-s))(\xi) = q_1(\xi)q_2(\xi)e^{-(t-s)|\xi|^2}$, where $q_1(\xi)$ and $q_2(\xi)$ are homogeneous functions on $|\xi| > 0$ of degree $1/2$. Note that in (59) we can take $q_1(\xi) = \xi_j/|\xi|^{1/2}$ and $q_2(\xi) = i\,(\delta_{jk}\,|\xi|^2 - \xi_j\,\xi_k)/|\xi|^{3/2}$.

We define $T^{(1)}(\tau)$ and $T^{(2)}(\tau)$ as the pseudodifferential operators whose symbols are

$$\sigma(T^{(1)}(\tau))(\xi) = q_1(\xi)\, e^{-\tau|\xi|^2/2}$$

and

$$\sigma(T^{(2)}(\tau))(\xi) = q_2(\xi)\, e^{-\tau|\xi|^2/2}\,,$$

and such that $Q(t-s) = T^{(1)}(t-s)\,T^{(2)}(t-s)$.

We consider too two functions f and g satisfying $\|f\|_E < \infty$ and $\|g\|_E < \infty$. In fact we shall only need that $\sup_{t>0} \|f(t)\|_{(n,\infty)} < \infty$ and $\sup_{t>0} \|g(t)\|_{(n,\infty)} < \infty$. Then, we shall prove (57) for the operator $\int_0^t Q(t-s)\,(f(s)g(s))(x)\,ds$.

Let φ be a testing function ($\varphi \in \mathcal{S}$). Hence we have

$$\begin{aligned} \Big|\Big\langle \int_0^t Q(t-s)\,(f(s)g(s))\,ds, \varphi \Big\rangle\Big| &= \Big| \int_0^t \langle Q(t-s)\,(f(s)g(s)), \varphi\rangle\, ds\Big| \\ &\leq \int_0^t \big|\langle T^{(1)}(t-s)\,T^{(2)}(t-s)\,(f(s)g(s)), \varphi\rangle\big|\, ds \\ &= \int_0^t \big|\langle T^{(2)}(t-s)\,(f(s)g(s)), T^{(1)}(t-s)\,\varphi\rangle\big|\, ds \\ &\leq \int_0^t \big\|T^{(2)}(t-s)\,(f(s)g(s))\big\|_{(n,\infty)} \big\|T^{(1)}(t-s)\,\varphi\big\|_{(n',1)}\, ds\,. \end{aligned} \tag{60}$$

Thanks to Propositions 1 and 2 we get

$$\begin{aligned} &\|T^{(2)}(t-s)\,(f(s)g(s))\|_{(n,\infty)} \\ &\qquad \leq c(n)\,\|k_2(t-s)\|_{p_2}\,\|f(s)\|_{(n,\infty)}\,\|g(s)\|_{(n,\infty)} \end{aligned} \tag{61}$$

with $1/n = 1/p_2 + 2/n - 1$ and the function $k_2(t-s) \in \mathcal{S}$ verifies

$$\widehat{k_2(t-s)}(\xi) = q_2(\xi)\, e^{-(t-s)|\xi|^2/2}\,.$$

It is clear that $p_2 = n'$. As usual, we denote by $\widehat{h}(\xi)$ the Fourier transformation of a function h at the point ξ.

Analogously,

$$\big\|T^{(1)}(t-s)\varphi\big\|_{(n',1)} \leq c(n,r)\,\|k_1(t-s)\|_{p_1}\,\|\varphi\|_{(r,1)}\,, \tag{62}$$

where $1/n' = 1/p_1 + 1/r - 1$, $p_1 > 1$, $r > 1$, and the function $k_1(t-s) \in \mathcal{S}$ satisfies

$$\widehat{k_1(t-s)}(\xi) = q_1(\xi)\, e^{-(t-s)|\xi|^2/2}\,.$$

Then, for each function $k_j(t-s)$, $j = 1, 2$, we can rapidly see that

$$\begin{aligned} \|k_j(t-s)\|_{p_j} &= \left\|(t-s)^{-1/4-n/2}\, \tilde{k}_j\Big(\frac{\cdot}{\sqrt{t-s}}\Big)\right\|_{p_j} \\ &= (t-s)^{-1/4-n/2+n/(2p_j)}\, \|\tilde{k}_j\|_{p_j} \\ &= c(n,p_j)\,(t-s)^{-1/4-n/2+n/(2p_j)}\,, \end{aligned} \tag{63}$$

with $\widehat{\tilde{k}_j}(\xi) = q_j(\xi)\, e^{-|\xi|^2/2}$.

Due to (63) and (62) we obtain

$$\begin{aligned} \|T^{(1)}(t-s)\varphi\|_{(n',1)} &\le \tilde{c}(n,r)\,(t-s)^{-1/4-n/2+n/(2p_1)}\, \|\varphi\|_{(r,1)} \\ &= \tilde{c}(n,r)\,(t-s)^{-3/4+n/2-n/(2r)}\, \|\varphi\|_{(r,1)}\,. \end{aligned} \tag{64}$$

Besides, from (63) and (61) we have

$$\begin{aligned} &\|T^{(2)}(t-s)\,(f(s)g(s))\|_{(n,\infty)} \\ &\quad\le \tilde{c}(n)\,(t-s)^{-3/4}\, \|f(s)\|_{(n,\infty)}\, \|g(s)\|_{(n,\infty)} \\ &\quad\le \tilde{c}(n)\,(t-s)^{-3/4}\, \sup_{\tau>0} \|f(\tau)\|_{(n,\infty)} \\ &\qquad\cdot \sup_{\tau>0} \|g(\tau)\|_{(n,\infty)}\,. \end{aligned} \tag{65}$$

Therefore, replacing (64) and (65) in (60) we get

$$\begin{aligned} &\left|\left\langle \int_0^t Q(s-t)\,(f(s)g(s))\,ds, \varphi \right\rangle\right| \\ &\qquad\le \tilde{\tilde{c}}(n,r)\, \|\varphi\|_{(r,1)} \int_0^t (t-s)^{-3/2+(n-n/r)/2}\, ds \\ &\qquad\quad\cdot \sup_{\tau>0} \|f(\tau)\|_{(n,\infty)}\, \sup_{\tau>0} \|g(\tau)\|_{(n,\infty)}\,. \end{aligned}$$

For instance, choosing $r = 3$, it results that $r > n'$ since $n \ge 2$ and then the s-integral is convergent and it is like $t^{-1/2+(n-n/r)/2}$ which tends to zero with t. Finally we have that

$$\lim_{t\to 0} \left|\left\langle \int_0^t Q(s-t)\,(f(s)g(s))\,ds, \varphi \right\rangle\right| = 0\,.$$

This proves Lemma 6.

The conclusions of Lemmas 4, 5 and 6 can be briefly summarized in the following terms.

Proposition 5. *The bilinear operator $B : E \times E \longrightarrow E$ is continuous.*

Now, we are ready to prove Theorem 1. The method we shall apply takes the approach on succesive approximations (called Picard's method) developed in [K-F], [F-K], [K] and [C]. Picard's sequence is defined as follows.

$$\begin{aligned} \mathbf{v}_1(t,x) &= (S(t)\mathbf{v}_0)(x)\,, \\ \mathbf{v}_{m+1}(t,x) &= \mathbf{v}_1(t,x) + B(\mathbf{v}_m, \mathbf{v}_m)(t,x)\,, \end{aligned} \tag{66}$$

for $m = 1, 2, 3, \ldots$ Most of the estimates given here are the weak versions of those presented in [K]. In [K-F] and [F-K] some techniques based on the fractional powers of the Laplace operator are used. We are not going to employ them here.

Previously to explain the details of the proof, we need the following abstract result extracted from [C].

Lemma 7 ([C]). *Let X be a Banach space with norm $\|\cdot\|_X$, and $B : X \times X \longrightarrow X$ a continuous bilinear map. That is, there exists a constant $\mathcal{K} > 0$ such that for all x_1 and x_2 in X we have*

$$\|B(x_1, x_2)\|_X \leq \mathcal{K}\, \|x_1\|_X\, \|x_2\|_X\, .$$

Then, for any vector $y \in X$, $y \neq 0$, such that $4\mathcal{K}\,\|y\|_X < 1$, there exists a solution $x \in X$ for the equation $x = y + B(x,x)$. Moreover, this solution x satisfies $\|x\|_X \leq 2\,\|y\|_X$.

This Lemma will be proved in the same way as in [C]. Let

$$R = \frac{1 - \sqrt{1 - 4\mathcal{K}\,\|y\|_X}}{2\mathcal{K}}\,,$$

which is the smallest solution of the equation $\|y\|_X + \mathcal{K}\,R^2 = R$. We can easily observe that $R \leq 2\,\|y\|_X$.

In X, we consider the closed ball $B_R = \{x \in X : \|x\|_X \leq R\}$. Let us define $F : X \longrightarrow X$ by $F(x) = y + B(x,x)$. First, we note that for all $x \in B_R$,

$$\|F(x)\|_X \leq \|y\|_X + \mathcal{K}\,\|x\|_X^2 \leq \|y\|_X + \mathcal{K}\,R^2 = R\,.$$

That is, F maps B_R into B_R. Besides, for all x and x' in B_R we have

$$\begin{aligned}\|F(x) - F(x')\|_X &= \|B(x,x) - B(x',x')\|_X \\ &\le \|B(x - x', x)\|_X + \|B(x', x - x')\|_X \\ &\le 2\mathcal{K} R \|x - x'\|_X \,.\end{aligned}$$

From the definition of R, it becomes

$$0 < 2\mathcal{K} R = 1 - \sqrt{1 - 4\mathcal{K}\|y\|_X} < 1\,.$$

Hence, $\|F(x) - F(x')\|_X \le c\|x - x'\|_X$, with $0 < c < 1$. Thus, the map $F : B_R \longrightarrow B_R$ is a contraction. Consequently, from the Picard contractions Theorem applied to the sequence

$$\begin{aligned} x_1 &= y\,,, \\ x_{m+1} &= F(x_m)\,, \qquad \text{for } m = 1, 2, \ldots, \end{aligned}$$

we obtain a unique solution x in B_R, but perhaps not unique in X. The last estimates is trivial, since as it was observed

$$\|x\|_X \le R \le 2\|y\|_X \,. \tag{67}$$

PROOF OF THEOREM 1. For the existence of mild-solutions of problem (2) we want to apply Lemma 7 to the integral equation (6) with the Banach space $X = E$ and the vector $y = S(t)\mathbf{v}_0$. This fact leads us to check the condition

$$4\mathcal{K}\|S(t)\mathbf{v}_0\|_E < 1\,, \tag{68}$$

being $\mathcal{K}$ the constant given in Lemma 4. Besides, from Proposition 4 there exists a constant $c > 0$ such that

$$\|S(t)\mathbf{v}_0\|_E \le c\|\mathbf{v}_0\|_{(n,\infty)} \,. \tag{69}$$

Choosing $0 < \delta < (4\mathcal{K} c)^{-1}$ and making use of (69), inequality (68) is satisfied provided that $\|\mathbf{v}_0\|_{(n,\infty)} < \delta$. Thus, Lemma 7 guarantees the existence of a global mild-solution $\mathbf{v}(t,x)$ in E. Moreover, we know from (67) that $\|\mathbf{v}\|_E \le 2\|S(t)\mathbf{v}_0\|_E \le 2c\delta$, which goes to 0 with δ.

Now, we shall pass to consider the uniqueness. Let us suppose that $\mathbf{u}$ and $\mathbf{v}$ are two mild-solutions of the Navier-Stokes equations in E with the same initial data $\mathbf{v}_0$, and such that there exists $0 < \eta_0 < 1/\mathcal{K}$ for which the following inequalities

$$\sup_{t>0} t^{(1-n/q)/2} \|\mathbf{u}(t)\|_{(q,\infty)} < \eta_0$$

and

$$\sup_{t>0} t^{(1-n/q)/2} \|\mathbf{v}(t)\|_{(q,\infty)} < \eta_0$$

hold. Thanks to the estimate (48) of the proof of Lemma 4, we note that the difference

$$\begin{aligned}\mathbf{w}(t,x) &:= \mathbf{u}(t,x) - \mathbf{v}(t,x)\\ &= B(\mathbf{u},\mathbf{u})(t,x) - B(\mathbf{v},\mathbf{v})(t,x)\\ &= B(\mathbf{w},\mathbf{u})(t,x) - B(\mathbf{v},\mathbf{w})(t,x)\end{aligned}$$

satisfies (with $T_0 = \infty$)

$$\begin{aligned}&\sup_{t>0} t^{(1-n/q)/2} \|\mathbf{w}(t)\|_{(q,\infty)}\\ &\quad \le \frac{\mathcal{K}}{2}\Big(\sup_{t>0} t^{(1-n/q)/2} \|\mathbf{u}(t)\|_{(q,\infty)} + \sup_{t>0} t^{(1-n/q)/2} \|\mathbf{v}(t)\|_{(q,\infty)}\Big)\\ &\quad\quad \cdot \sup_{t>0} t^{(1-n/q)/2} \|\mathbf{w}(t)\|_{(q,\infty)}\\ &\quad < \mathcal{K}\,\eta_0 \sup_{t>0} t^{(1-n/q)/2} \|\mathbf{w}(t)\|_{(q,\infty)}\\ &\quad < \sup_{t>0} t^{(1-n/q)/2} \|\mathbf{w}(t)\|_{(q,\infty)}\,,\end{aligned}$$

since for $\eta_0 > 0$ it was supposed $\mathcal{K}\eta_0 < 1$. This fact implies that $\mathbf{w}(t) = 0$ for all $t > 0$, and then, $\mathbf{v}(t) = \mathbf{u}(t)$. This concludes the proof.

PROOF OF THEOREM 2. From Theorem 1 we know that the system (2) with initial data $\mathbf{v}_0(x)$ admits, at least, a global mild-solution $\mathbf{v}(t,x) \in E$. In fact, this solution was founded as the limit of the sequence $\{\mathbf{v}_m(t,x)\}$ defined by (66). It is straightforward to verify that $\mathbf{v}_1(t,x) = (S(t)\mathbf{v}_0)(x)$ satisfies $\lambda\,\mathbf{v}_1(\lambda^2 t, \lambda x) = \mathbf{v}_1(t,x)$ by a change of variables. Also, it is evident to see by induction on m that all the functions $\mathbf{v}_m(t,x)$ have this property. Then, it is clear that the limit $\mathbf{v}(t,x)$ must verify

$$\lambda\,\mathbf{v}(\lambda^2 t, \lambda x) = \mathbf{v}(t,x)$$

for all $\lambda > 0$, all $t > 0$ and all $x \in \mathbb{R}^n$. Theorem 2 is proved.

PROOF OF THEOREM 3. Theorem 3 follows directly from Theorem 2 and the following characterization.

Lemma 8 (A characterization of homogeneous functions). *Let $0 < d < n$ and let f be a complex-valued function defined on $\mathbb{R}^n$. Suppose that f is homogeneous of degree $-d$. The function $f \in L^{(n/d,\infty)}(\mathbb{R}^n)$ if and only if its restriction $f|_{S^{n-1}}$ to the sphere S^{n-1} of $\mathbb{R}^n$ belongs to $L^{n/d}(S^{n-1})$.*

The key of this Lemma is centered in the following computation. Let $s > 0$. Recall that m denotes the Lebesgue measure in $\mathbb{R}^n$.

Therefore, taking polar coordinates and reminding the homogeneity assumption on f, we can write

$$\begin{aligned} m\{x \in \mathbb{R}^n : \ |f(x)| > s\} &= \int_{|f(x)|>s} 1\, dx \\ &= \int_0^\infty \int_{\{\xi \in S^{n-1} : |f(\xi)| > r^d s\}} r^{n-1}\, d\sigma_\xi\, dr\,. \end{aligned}$$

Here, $d\sigma_\xi$ denotes the differential area over the sphere. Applying now Fubini's Theorem we conclude

$$\begin{aligned} m\{x \in \mathbb{R}^n : \ |f(x)| > s\} &= \int_{S^{n-1}} \int_0^{(|f(\xi)|/s)^{1/d}} r^{n-1}\, dr\, d\sigma_\xi \\ &= \int_{S^{n-1}} \frac{1}{n} \frac{|f(\xi)|^{n/d}}{s^{n/d}}\, d\sigma_\xi\,. \end{aligned}$$

Finally, we get

$$\begin{aligned} \|f\|^*_{(n/d,\infty)} &= \sup_{s>0} s\, m\{x \in \mathbb{R}^n : \ |f(x)| > s\}^{d/n} \\ &= \frac{1}{n^{d/n}} \|f|_{S^{n-1}}\|_{L^{n/d}(S^{n-1})}\,. \end{aligned}$$

Then, one of both norms is finite if and only if it is finite the other. It is necessary to take into account the "equivalence" (19) and the hypothesis $0 < d < n$ to complete the proof of Lemma 8.

REMARK. It is easy to observe after Fubini's Theorem and a standard change of variables that there is not any homogeneous function defined on $\mathbb{R}^n$ belonging to $L^p(\mathbb{R}^n)$, for any $p > 0$. This fact leads us to try to find " adequate" extensions of $L^p(\mathbb{R}^n)$ as , for instance, the $L^{(p,\infty)}(\mathbb{R}^n)$ spaces.

FINAL REMARK. The results shown throughout the present paper can be obtained as a consequence of the general theory exposed by M. Cannone in [C], for the case of spatial dimension $n = 3$. It is not true that the Lorentz space $L^{(3,\infty)}(\mathbb{R}^3)$ is a subset of the Besov space $\dot{B}_3^{0,\infty}$ but, on the other hand, we have that

$$L^{(3,\infty)}(\mathbb{R}^3) \subset \dot{B}_q^{-\alpha,\infty}, \tag{70}$$

for $q > 3$ and $\alpha = 1 - 3/q$. So, the Besov space $\dot{B}_q^{-\alpha,\infty}$ can be useful as the auxiliary space to build the artificial norm in Kato's theory (see, for instance, [F-K], [K-F], [K]).

From [C] we learn that the bilinear operator B defined by (8) is continuous from $\dot{B}_q^{-\alpha,\infty} \times \dot{B}_q^{-\alpha,\infty}$ to $L^3(\mathbb{R}^3)$. Therefore, taking into account (70), if the initial condition v_0 satisfies $\|v_0\|_{(3,\infty)} < \alpha$, for some $\alpha > 0$ small enough, we have $\|v_0\|_{\dot{B}_q^{-\alpha,\infty}} < c\,\alpha$, and then, thanks to Cannone's results [C] once more, we know that the solution

$$\mathbf{v}(t,x) = S(t)\mathbf{v}_0(x) + B(\mathbf{v},\mathbf{v})(t,x),$$

where $\|B(\mathbf{v},\mathbf{v})(t)\|_3 \leq c_1$. On the other hand $L^{(3,\infty)}$ is a translation invariant space which yields $S(t)\mathbf{v}_0 \in L^{(3,\infty)}$ uniformly on t.

References.

[C] Cannone, M., Ondelettes, paraproduits et Navier-Stokes. Thesis. Université Paris IX-Dauphine, 1994.

[F-M-T] Foias, C., Manley, O. P., Temam, R., New representation of the Navier-Stokes equations governing self-similar homogeneous turbulence. *Phys. Rev. Lett.* **51** (1983), 617-620.

[F-K] Fujita, H., Kato, T., On the Navier-Stokes initial value problem, I. *Arch. Rational Mech. Anal.* **16** (1964), 269-315.

[F-M] Fuyiwara, D., Morimoto, H., An L_r-theorem of the Helmholtz decomposition of vector fields. *J. Fac. Sci. Univ. Tokyo, Sec. I* **24** (1977), 685-700.

[G-M] Giga, Y., Miyakawa, T., Navier-Stokes flow in $\mathbb{R}^3$ with measures as initial vorticity and Morrey spaces. *Comm. Partial Diff. Equations* **14** (1989), 577-618.

[H-L-P] Hardy, G. H., Littlewood, J. E., Pólya, G., *Inequalities.* Cambridge Univ. Press, 1934.

[H] Hunt, R., On $L(p,q)$ spaces. *L'Enseignement Mathématique* **12** (1966), 249-276.

[K] Kato, T., Strong L^p-solutions of the Navier-Stokes equation in $\mathbb{R}^m$, with applications to weak solutions. *Math. Z.* **187** (1984), 471-480.

[K-F] Kato, T., Fujita, H., On the non-stationary Navier-Stokes system. *Rend. Sem. Math. Univ. Padova* **32** (1962), 243-260.

[Kn] Katznelson, Y., *An Introduction to Harmonic Analysis.* J. Wiley & Sons, 1968.

[K-Y] Kozono, H., Yamazaki, M., Semilinear heat equations and the Navier-Stokes equation with distributions in new function spaces as initial data. *Comm. Partial Diff. Equations* **19** (1994), 959-1014.

[L] Lorentz, G. G., Some new functional spaces. *Ann. of Math.* **1** (1950), 37-55.

[O] O'Neil, R., Convolution operators and $L(p,q)$ spaces. *Duke Math. J.* **30** (1963), 129-142.

[S-W] Stein, E., Weiss, G., *Introduction to Fourier Analysis on Euclidean Spaces.* Princeton Univ. Press, 1971.

[T] Temam, R., *Infinite-dimensional dynamical system in Mechanics and Physics.* Springer-Verlag, 1988.

Recibido: 30 de marzo de 1.995

Oscar A. Barraza*
CEREMADE
Université Paris IX-Dauphine
FRANCE
barraza@paris9.dauphine.fr

and

Departamento de Matemática (UNLP)
C.C. 172 (1900) - La Plata
ARGENTINA
oscar@cacho.mate.unlp.edu.ar

* This research is part of a fellowship at CEREMADE-UNIVERSITE PARIS IX--DAUPHINE, FRANCE, under the direction of Professor Yves Meyer. It was supported by CONICET, ARGENTINA.

Revista Matemática Iberoamericana
Vol. 12, N.° 2, 1996

The spectrum of singularities of Riemann's function

Stéphane Jaffard

Abstract. We determine the Hölder regularity of Riemann's function at each point; we deduce from this analysis its spectrum of singularities, thus showing its multifractal nature.

1. Introduction.

According to the tradition, Riemann would have proposed the function

$$\varphi(x) = \sum_{1}^{\infty} \frac{1}{n^2} \sin \pi n^2 x$$

as an example of continuous nowhere differentiable function. It turned out that, unlike lacunary series, the regularity of this function varies strongly from point to point. Let $x_0 \in \mathbb{R}$; by definition, a function f is $C^\alpha(x_0)$ if there exists a polynomial P of order at most α such that

$$(1) \qquad |f(x) - P(x - x_0)| \leq C\, |x - x_0|^\alpha .$$

Let us recall the main steps of the determination of the Hölder regularity of φ at every point.

Hardy and Littlewood proved that φ is nowhere $C^{3/4}$ except perhaps at the rational points of the form $(2p+1)/(2q+1)$, $p, q \in \mathbb{Z}$ (see [9]). Their proof is interesting under many respects; for instance it anticipates wavelet methods; they remark that the function

$$C(a,b) = \frac{a}{2}\left(\theta(b+ia) - 1\right)$$

(where θ is the Jacobi function $\theta(z) = \sum_{n\in\mathbb{Z}} e^{i\pi n^2 z}$) is the convolution of φ with contractions (by a factor a) of

$$\psi(x) = \frac{1}{\pi(x-i)^2}\,.$$

Since ψ has a vanishing integral, an Abel type theorem (which we will state precisely in Proposition 1) shows that, if φ is smooth at x_0, $C(a,b)$ must have a certain decay when $a \to 0$ and $b \to x_0$. This method will thus yield upper bounds for the function

$$\alpha(x_0) = \sup\{\beta :\ \varphi \in C^{\beta}(x_0)\}\,.$$

This method actually yields the following more precise result which relates the pointwise behavior of φ at x_0 to the Diophantine approximation properties of x_0. Let $x_0 \notin \mathbb{Q}$, p_n/q_n be the sequence of its approximations by continued fractions and define

$$\tau(x_0) = \sup\left\{\tau :\ \left|x_0 - \frac{p_m}{q_m}\right| \le \frac{1}{q_m^{\tau}}\right\}$$

for infinitely many m's such that p_m and q_m are not both odd, then

$$\alpha(x_0) \le \frac{1}{2} + \frac{1}{2\,\tau(x_0)}\,.$$

a result which is actually stated by J. J. Duistermaat [4] where a more direct proof is given (this paper was actually one of the main motivations for writing the present one).

Converse results, which would yield an information about the pointwise regularity of φ from estimates on its convolutions are more difficult to obtain since they are of tauberian type, and were of course unavailable at the time of Hardy and Littlewood. This is why they had results concerning only the *irregularity* of Riemann's function and not its regularity. The Tauberian-type result we need is stated in Proposition 1.

Finally Gerver proved the differentiability at the rational points of the form $(2p+1)/(2q+1)$ [7] (where we now know that φ is exactly

$C^{3/2}$, see [13]). The analysis of the behavior of φ near such a rational point x_0 has been considerably sharpened since; Y. Meyer exhibited a complete "chirp" asymptotic expansion which describes the oscillations of φ:

$$\varphi(x) = u(x) + \sum_{n\geq 0} (x - x_0)^{n+3/2}\, v_+^n\Big(\frac{1}{x - x_0}\Big), \qquad \text{if } x \geq x_0 ,$$

$$\varphi(x) = u(x) + \sum_{n\geq 0} |x - x_0|^{n+3/2}\, v_-^n\Big(\frac{1}{|x - x_0|}\Big), \qquad \text{if } x \leq x_0 ,$$

where u is C^∞, the $v_\pm^n$ are 2π periodic, with a vanishing integral, and are $C^{n+1/2}$, see [13] (we will actually show that these points are the only ones where a chirp expansion exists).

The results of Hardy and Gerver left open the problem of the determination of the exact regularity of φ at irrational points; one of our purposes is to do this determination using the wavelet method we sketched.

Let the spectrum of singularities of φ be the function $d(\beta)$ which associates to each β the Hausdorff dimension of the set of points x where $\alpha(x) = \beta$ (conventionally the dimension of the empty set is $-\infty$). We will deduce form our study this spectrum which will be nonconstant on a whole interval. The Hölder singularities of φ are located on a whole collection of sets of different dimensions, so that φ is truly a "multifractal function". More precisely the determination of the spectrum of singularities is motivated by the following problem, referred to in the litterature as the "Multifractal Formalism for functions".

Let

$$L^{p,s} = \{f \in L^p : \ (-\Delta)^{s/2} f \in L^p\}.$$

If g is a one variable function, define

$$\eta(p) = \sup\{s : \ g \in L^{p,s/p}\}.$$

Frisch and Parisi in [6] conjectured that the spectrum of singularities of g is given by the following Legendre transform formula

$$d(\alpha) = \inf_p (\alpha\, p - \eta(p) + 1). \tag{2}$$

Though one easily finds counterexamples, the exact mathematical range of validity of this formula is a fascinating problem (see [1], [3] and [12]).

Since φ has a nontrivial spectrum, it was natural to test this conjecture, and we will show that it is correct in this case. The interesting point here is that φ has a very different structure from the previous cases where the Multifractal Formalism was known to hold (see [3] and [12]).

Our main results are stated in the following theorem.

Theorem 1. *Let $x \notin \mathbb{Q}$ and let p_n/q_n be the sequence of its approximations by continued fractions. Let*

$$\tau(x) = \sup \left\{ \tau : \ \left| x - \frac{p_m}{q_m} \right| \leq \frac{1}{q_m^{\tau}} \right\}$$

for infinitely many m's such that p_m and q_m are not both odd.

Then

$$\alpha(x) = \frac{1}{2} + \frac{1}{2\,\tau(x)} \,.$$

The spectrum of singularities of φ is given by

$$d(\alpha) = \begin{cases} 4\alpha - 2\,, & \text{if } \alpha \in \left[\frac{1}{2}, \frac{3}{4}\right], \\ 0\,, & \text{if } \alpha = \frac{3}{2}, \\ -\infty\,, & \text{else}; \end{cases}$$

and if $\alpha \leq 3/4$, $d(\alpha)$ satisfies (2).

The existence of this sets of smooth points is not only a consequence of the kind of lacunarity introduced by the frequencies n^2, but also of the very special coefficents that are chosen, which creates an exceptional behavior, as shown by the following remark: if the coefficients were multiplied by independant identically distributed Gaussians or Rademacher series (± 1), [14, Chapter 8, Theorem 4] shows that the corresponding random function would be almost surely nowhere $C^{1/2}$.

2. Pointwise regularity and wavelet transform.

Because of Hardy's result, we will only consider Hölder exponents smaller than 3/4.

Suppose that a function ψ is nonvanishing and satisfies the following asumptions

$$(3)\qquad |\psi(x)| + |\psi'(x)| \leq C\,(1+|x|)^{-2} \qquad \text{and} \qquad \int \psi(x)\,dx = 0\,,$$

and either

$$(4)\qquad \int_0^\infty |\hat{\psi}(\xi)|^2\,\frac{d\xi}{\xi} = \int_0^\infty |\hat{\psi}(-\xi)|^2\,\frac{d\xi}{\xi} = 1$$

or

$$(5)\qquad \hat{\psi}(\xi) = 0\,, \quad \text{if } \xi < 0 \qquad \text{and} \qquad \int_0^\infty |\hat{\psi}(\xi)|^2\,\frac{d\xi}{\xi} = 1\,.$$

In the last case the wavelet is said to be analytic. The wavelet transform of an L^∞ function f is defined by

$$C(a,b)(f) = \frac{1}{a}\int f(t)\,\bar{\psi}\Big(\frac{t-b}{a}\Big)\,dt\,.$$

We will consider the three following settings: In the first one the analyzed function f is real valued, and the wavelet satisfies (5); in the second one, f is complex valued and the wavelet satisfies (4); in the third one, $\hat{f}(\xi) = 0$ if $\xi < 0$ and the wavelet satisfies (5).

In each case, the following results concerning the relationships between the size of the wavelet transform and the regularity of the function hold.

Proposition 1. *Suppose that* $0 < \alpha < 1$. *Under the previous hypotheses if a function* f *is* $C^\alpha(x_0)$,

$$(6)\qquad |C(a,b)(f)| \leq C\,a^\alpha \left(1 + \frac{|b-x_0|}{a}\right)^\alpha .$$

Conversely, if

$$(7)\qquad |C(a,b)(f)| \leq C\,a^\alpha \left(1 + \frac{|b-x_0|}{a}\right)^{\alpha'} \quad \textit{for an } \alpha' < \alpha\,,$$

and if $|x - x_0| \leq 1/2$, *then*

$$(8)\qquad |f(x) - f(x_0)| \leq C\,|x - x_0|^\alpha\,.$$

The first assertion was first stated in the wavelet terminology in [10] or [11], but it is only fair to say that it is at least implicitely contained in Hardy's paper [8]; and only the second assertion (see [11]) will have new implications on Riemann's function. Let us recall the proof of this proposition for the reader's convenience when $0 < \alpha < 1$ (the only case we will be interested in here).

PROOF OF PROPOSITION 1. If $f \in C^\alpha(x_0)$,

$$\begin{aligned}
|C(a,b)(f)| &= \frac{1}{a}\Big|\int f(x)\,\psi(\frac{x-b}{a})\,dx\Big| \\
&= \frac{1}{a}\Big|\int (f(x)-f(x_0))\,\psi(\frac{x-b}{a})\,dx\Big| \\
&\le \frac{C}{a}\int |x-x_0|^\alpha \left(\frac{1}{1+\left|\frac{x-b}{a}\right|}\right)^2 dx \\
&\le \frac{C}{a}\int \frac{|x-b|^\alpha}{\left(1+\left|\frac{x-b}{a}\right|\right)^2}\,dx + |b-x_0|^\alpha\,\frac{C}{a}\int \frac{dx}{\left(1+\left|\frac{x-b}{a}\right|\right)^2} \\
&\le C\,a^\alpha \left(1+\left|\frac{b-x_0}{a}\right|\right)^\alpha .
\end{aligned}$$

Suppose now that we are in the first or the third case. In that case, f is reconstructed from its wavelet transform by

$$f(x) = \iint C(a,b)(f)\,\psi(\frac{x-b}{a})\,\frac{da\,db}{a^2} .$$

Let

$$\omega(a,x) = \int C(a,b)(f)\,\psi(\frac{x-b}{a})\,\frac{db}{a} ;$$

if (7) holds,

$$|\omega(a,x)| \le C\,a^\alpha \left(1+\frac{|x-x_0|}{a}\right)^{\alpha'}$$

and

$$\left|\frac{\partial\omega(a,x)}{\partial x}\right| \le C\,a^{\alpha-1}\left(1+\frac{|x-x_0|}{a}\right)^{\alpha'} .$$

Using the second estimate (and the mean value theorem) for $a \ge |x-x_0|$ and the first estimate for $a \le |x-x_0|$, we obtain

$$|f(x)-f(x_0)| \le \int_{a\ge|x-x_0|} C\,a^{\alpha-1}\left(1+\left|\frac{x-x_0}{a}\right|\right)^{\alpha'} |x-x_0|\,\frac{da}{a}$$

$$+ \int_{a \le |x-x_0|} C\, a^{\alpha} \left(1 + \left|\frac{x-x_0}{a}\right|\right)^{\alpha'} \frac{da}{a}$$
$$\le C\, |x - x_0|^{\alpha}\,.$$

This also implies the result in the second case by superposing reconstruction formulas for $\xi \ge 0$ and $\xi \le 0$.

Using Cauchy's formula, we obtain that (using the wavelet $\psi(x) = (x-i)^{-2}$) the wavelet transform of $\varphi(x)$ is $2ia(\theta(b+ia)-1)/2$. Since we want to determine Hölder exponents between $1/2$ and $3/4$, because of (7) we can add a term ia and the study of the pointwise regularity of φ reduces to obtaining estimates similar to (6) for the function

$$C(a,b) = a\,\theta(b+ia)\,. \tag{9}$$

3. Theta Jacobi function and continued fractions.

The Theta modular group is obtained by composing the two transforms

$$x \longmapsto x+2 \qquad \text{and} \qquad x \longmapsto -\frac{1}{x}\,.$$

It is composed of the fractional linear transformations

$$\gamma(x) = \frac{rx+s}{qx-p}\,,$$

where $rp + sq = -1$, r, s, p, q are integers and the matrix

$$\begin{pmatrix} r & s \\ q & p \end{pmatrix} \text{ is of the form } \begin{pmatrix} \text{even} & \text{odd} \\ \text{odd} & \text{even} \end{pmatrix} \text{ or } \begin{pmatrix} \text{odd} & \text{even} \\ \text{even} & \text{odd} \end{pmatrix}. \tag{10}$$

When γ belongs to the Theta modular group, θ is transformed following the formula (*cf.* [2] or [15])

$$\theta(z) = \theta(\gamma(z))\, e^{im\pi/4}\, q^{-1/2} \left(z - \frac{p}{q}\right)^{-1/2}, \tag{11}$$

where m is an integer which depends on r, s, p, q.

Let $\rho \notin \mathbb{Q}$ and p_n/q_n the sequence of its approximations by continued fractions. The idea of the proof of Theorem 1 is to use (11), which will allow us to deduce the behavior of $\theta(z)$ near p_n/q_n (hence near ρ) from its behavior near 0 or 1. Because of (10), we will have to separate two cases depending whether p_n and q_n are both odd or not; but let us first derive some straightforward estimates for θ near 0 and 1.

First remark that

$$|\theta(z) - 1| \le \frac{1}{2}\,, \qquad \text{if } \operatorname{Im} z \ge 1\,, \tag{12}$$

because in this case,

$$|\theta(z) - 1| \le 2 \sum_{n\ge 1} e^{-\pi n^2 \operatorname{Im} z} \le \frac{2\,e^{-\pi \operatorname{Im} z}}{1 - e^{-\pi \operatorname{Im} z}} \le \frac{1}{2}\,.$$

We also have

$$|\theta(z)| \le C\,|\operatorname{Im} z|^{-1/2}\,, \qquad \text{if } \operatorname{Im} z \le 1\,, \tag{13}$$

because $|\theta(z)| \le \sum e^{-n^2 \operatorname{Im} z}$; the sum for $n \le (\operatorname{Im} z)^{-1/2}$ is bounded trivially by $(\operatorname{Im} z)^{-1/2} + 1$ and the same bound holds for $n > (\operatorname{Im} z)^{-1/2}$ (by comparison with an integral).

Let us now obtain the behavior of θ near the point 1. Recall that θ satisfies (see [2])

$$\theta(1+z) = \sqrt{\frac{i}{z}}\,\Big(\theta\big(-\frac{1}{4z}\big) - \theta\big(-\frac{1}{z}\big)\Big)\,, \tag{14}$$

so that

$$\theta(1+z) = 2\sqrt{\frac{i}{z}}\,\big(A(4z) - A(z)\big)\,,$$

where $A(z) = \sum_1^\infty e^{-i\pi n^2/z}$. If $\operatorname{Im}(-1/z) \ge 1$,

$$|A(z)| \le 2\,\exp\Big(-\pi \operatorname{Im}\big(\frac{-1}{z}\big)\Big)\,,$$

so that in that case,

$$|\theta(1+z)| \le C\,|z|^{-1/2}\,\exp\Big(-\pi \operatorname{Im}\big(-\frac{1}{z}\big)\Big)\,. \tag{15}$$

Proposition 2. *Let $\{p_n/q_n\}$ be the sequence of approximations of ρ by continued fractions; let τ_n be defined by*

$$\left|\rho - \frac{p_n}{q_n}\right| = \left(\frac{1}{q_n}\right)^{\tau_n}. \tag{16}$$

For each n, if

$$3\left|\rho - \frac{p_n}{q_n}\right| \le |b - \rho + ia| \le 3\left|\rho - \frac{p_{n-1}}{q_{n-1}}\right| \tag{17}$$

the following estimates hold:

If p_n and q_n are not both odd but p_{n-1} and q_{n-1} are both odd, then

$$|C(a,b)| \le C\, a^{(1+1/\tau_n)/2}\left(1 + \frac{|b-\rho|}{a}\right)^{1/2\tau_n}. \tag{18}$$

If p_n and q_n are not both odd and p_{n-1} and q_{n-1} are not both odd, then

$$|C(a,b)| \le C\, a^{(1+1/\tau_n)/2}\left(1 + \frac{|b-\rho|}{a}\right)^{1/2\tau_n} \tag{19}$$

or

$$|C(a,b)| \le C\, a^{(1+1/\tau_{n-1})/2}\left(1 + \frac{|b-\rho|}{a}\right)^{1/2\tau_{n-1}}. \tag{20}$$

If p_n and q_n are both odd

$$|C(a,b)| \le C\, a^{(1+1/\tau_{n-1})/2}\left(1 + \frac{|b-\rho|}{a}\right)^{1/2\tau_{n-1}}. \tag{21}$$

Furthermore, if p_n and q_n are not both odd, these estimates are optimal, which means that there exists a point in the domain (17) *where* (18) *or* (19) *are equalities.*

Remark that, since $\tau_n \ge 2$, this result together with Proposition 1 implies Hardy's result that $\varphi(x) - \varphi(x_0)$ is nowhere $o(|x - x_0|)^{3/4}$ except perhaps at the rational points quotient of two odd numbers. More precisely, we have

Corollary 1. *Let $\rho \notin \mathbb{Q}$; If there exists an infinity of integers n such that p_n and q_n are not both odd and $\tau_n \ge \tau$, then*

$$\varphi(x) - \varphi(x_0) \ \textit{ is not } \ o(|x - x_0|)^{(1+1/\tau)/2}$$

but if there exists N such that for any $n \geq N$ (verifying p_n and q_n are not both odd), $\tau_n \leq \tau$, then

$$\varphi(x) - \varphi(x_0) = O(|x - x_0|)^{(1+\tau)/2}.$$

Define $\Gamma^\alpha(x_0)$ as the set of functions f such that

$$\begin{cases} \text{for all } \beta > \alpha, & f \notin C^\beta(x_0), \\ \text{for all } \beta < \alpha, & f \in C^\beta(x_0). \end{cases}$$

If $\eta(\rho) = \limsup \tau_n(\rho)$, where the lim sup bears only on the n's such that p_n and q_n are not both odd, this result implies that $\varphi \in \Gamma^{(1+1/\eta(\rho))/2}(\rho)$. We will prove this proposition in Sections 5 and 6, and in the next section, we will show how to derive the spectrum of singularities from this result.

Let us now recall a few properties of approximations by continued fractions.

Since

$$p_n\, q_{n-1} - p_{n-1}\, q_n = (-1)^{n-1} \tag{22}$$

thus

$$\frac{1}{q_n\, q_{n+1}} = \left|\frac{p_n}{q_n} - \frac{p_{n+1}}{q_{n+1}}\right| \geq \left|\rho - \frac{p_n}{q_n}\right|,$$

because (see [10]) p_n/q_n and p_{n+1}/q_{n+1} are not on the same side of ρ; and

$$\frac{1}{q_n\, q_{n+1}} = \left|\frac{p_n}{q_n} - \frac{p_{n+1}}{q_{n+1}}\right| \leq 2\left|\rho - \frac{p_n}{q_n}\right|,$$

so that

$$\left(\frac{1}{q_n}\right)^{\tau_n - 1} \leq \frac{1}{q_{n+1}} \leq 2\left(\frac{1}{q_n}\right)^{\tau_n - 1}. \tag{23}$$

4. Spectrum of singularities and Multifractal Formalism.

Recall that if

$$E_\tau = \left\{\rho : \left|\rho - \frac{p_n}{q_n}\right| \leq \frac{C}{q_n^\tau} \text{ for infinitely many } n\text{'s}\right\},$$

the Hausdorff dimension of E_τ is $2/\tau$ and the $\mathcal{H}^{2/\tau}$-measure of E_τ is positive (a direct consequence of [5, Propositions 10.4 and 8.5]).

Let

$$F_\tau = \left\{\rho : \left|\rho - \frac{p_n}{q_n}\right| \leq \frac{C}{q_n^\tau}\right\},$$

for infinitely many n's such that p_n and q_n are not both odd, and let

$$G_\tau = \{\rho : \left|\rho - \frac{p_n}{q_n}\right| \leq \frac{C}{q_n^\tau}\right\},$$

for infinitely many n's such that p_n and q_n are both odd. Of course, because of the best approximation properties of continued fractions, we have

$$E_\tau = F_\tau \cup G_\tau .$$

We will need the following lemma proved in [17].

Lemma 1. *Let $\rho \in \mathbb{R}$; If p and q have no common factor and*

$$|q\,\alpha - p| < \frac{1}{2\,q},$$

then p/q is a continued fraction approximation of ρ.

This lemma implies that if p/q is a continued fraction which approximates ρ and such that p and q are odd and $|\rho - p/q| \leq q^{-\tau}$ with $\tau > 2$, $p/(2q)$ is a continued fraction which approximates $\rho/2$.

Let us prove that the $\mathcal{H}^{2/\tau}$-measure of F_τ is positive. If the $\mathcal{H}^{2/\tau}$-measure of G_τ vanishes, we have nothing to prove. Else, the remark we just made shows that if $\tau > 2$ and $x \in G_\tau$ then $x/2 \in F_\tau$; thus, if the $\mathcal{H}^{2/\tau}$-measure of G_τ is positive, the $\mathcal{H}^{2/\tau}$-measure of F_τ is also positive (if $\tau = 2$, $F_\tau = \mathbb{R}$).

Consider the set

$$F_\tau \setminus \bigcup_{\tau' > \tau} E_{\tau'} .$$

The $\mathcal{H}^{2/\tau}$-measure of $\cup_{\tau'>\tau} E_{\tau'}$ vanishes; since F_τ has a positive $\mathcal{H}^{2/\tau}$-measure, $F_\tau \setminus \cup_{\tau'>\tau} E_{\tau'}$ has dimension $2/\tau$.

If $\rho \in F_\tau \setminus \cup_{\tau'>\tau} E_{\tau'}$, since $\rho \in F_\tau$, Proposition 2 implies that φ is not smoother than $(1 + 1/\tau)/2$ at ρ and since $\rho \notin \cup_{\tau'>\tau} E_{\tau'}$, φ is $C^{(1+1/\tau)/2-\varepsilon}(\rho)$ for all $\varepsilon > 0$; thus $\varphi \in \Gamma^{(1+1/\tau)/2}(\rho)$ and the dimension of $\{\rho : \varphi \in \Gamma^{(1+1/\tau)/2}(\rho)\}$ is at least $2/\tau$.

Suppose that ρ is such that $\varphi \in \Gamma^{(1+1/\tau)/2}(\rho)$; then φ is $C^{(1+1/\tau)/2-\varepsilon}(\rho)$ for all $\varepsilon > 0$ and thus $\rho \in E_{\tau'}$ for all $\tau' < \tau$; thus

$$\{\rho : \ \varphi \in \Gamma^{(1+1/\tau)/2}(\rho)\} \subset \bigcup_{\tau' > \tau} E_{\tau'}$$

and the dimension of $\{\rho : \ \varphi \in \Gamma^{(1+1/\tau)/2}(\rho)\}$ is bounded by $2/\tau$, hence the second part of Theorem 1 follows.

Let us now check that the Multifractal Formalism is true for Riemann's function. Let

$$S_n(x) = \sum_{m=1}^{n} e^{im^2\pi x}\,.$$

In [18], Z. Zalcwasser proves that

$$\int_0^1 |S_n(x)|^p \, dx \sim \begin{cases} n^{p/2}\,, & \text{if } 0 < p < 4\,, \\ n^2 \log(n+1)\,, & \text{if } p = 4\,, \\ n^{p-2}\,, & \text{if } p > 4\,. \end{cases}$$

Thus, taking D-adic blocs (for a D large enough),

$$\Big\| \sum_{D^j \le m^2 < D^{j+1}} e^{im^2\pi x} \Big\|_{L^p} \sim \begin{cases} D^{j/4}\,, & \text{if } 0 < p < 4\,, \\ D^{j(p-2)/(2p)}\,, & \text{if } p > 4\,. \end{cases}$$

Let $\Phi = \sum e^{in^2\pi x}/n^2$; we have $\Phi' \in B_p^{-1/4,\infty}$ if $p < 4$ and $\Phi' \in B_p^{-1/2+1/p,\infty}$ if $p > 4$. So that $\Phi \in B_p^{3/4,\infty}$ if $p < 4$ and $\Phi \in B_p^{1/2+1/p,\infty}$ if $p > 4$, and these estimates are optimal. Because of the continuity of the Hilbert transform on Besov spaces, the same result holds for φ. Since $\eta(p)$ can also be defined by

$$\eta(p) = \sup\{s : \ \varphi \in B_p^{s/p,\infty}\}\,,$$

we have

$$\eta(p) = \begin{cases} 3p/4\,, & \text{if } 0 < p \le 4\,, \\ 1 + p/2\,, & \text{if } p \ge 4\,. \end{cases}$$

If $\alpha < 1/2$, $\inf_p(\alpha p - \eta(p) + 1) = -\infty$ and if $1/2 \le \alpha \le 3/4$, $\inf_p(\alpha p - \eta(p) + 1) = 4\alpha - 2$; we recover thus the increasing part of the spectrum, thus showing the validity of the Multifractal Formalim in that case.

Remark that Propositions 1 and 2 imply that if $x_0 \in F_\tau$, φ is $C^{(1+1/\tau)/2}(x_0)$.

We now prove Proposition 2.

5. The case when p_n and q_n are not both odd.

We first determine $\gamma_n = (r_n x + s_n)/(q_n x - p_n)$ in the theta modular group such that the pole of γ_n is p_n/q_n. Because of (22) if p_{n-1} and q_{n-1} are not both odd, we can choose

$$r_n = (-1)^n q_{n-1}\,, \qquad s_n = (-1)^{n+1} p_{n-1}\ ;$$

the corresponding transform satisfies (10) and thus belongs to the theta modular group; and if p_{n-1} and q_{n-1} are both odd, we can choose

$$r_n = (-1)^n q_{n-1} + q_n\,, \qquad s_n = (-1)^{n+1} p_{n-1} - p_n\,.$$

Since

$$\gamma_n\Big(\frac{p_n}{q_n} + z\Big) = \frac{r_n}{q_n} - \frac{1}{q_n^2 z}\,,$$

applying (11) to $p_n/q_n + z$ and γ_n we obtain

$$\Big|\theta\Big(\frac{p_n}{q_n} + z\Big)\Big| = \Big|\theta\Big(\frac{r_n}{q_n} - \frac{1}{q_n^2 z}\Big)\Big| \frac{1}{\sqrt{q_n |z|}}\,. \tag{24}$$

Since $\operatorname{Im}(-1/q_n^2 z) = \operatorname{Im}(z)/q_n^2|z|^2$, we consider the two following cases.

First case: $\operatorname{Im}(z)/q_n^2|z|^2 \geq 1$; then (24) and (12) imply that

$$\Big|\theta\Big(\frac{p_n}{q_n} + z\Big)\Big| \sim \frac{1}{\sqrt{q_n|z|}}$$

so that

$$|C(a,b)| \sim \frac{C\,a}{\sqrt{q_n(a + |b-\rho|)}}$$

(note that here and hereafter, $\sim$ means that the two quantities are equivalent, the constants in the equivalence being independant of n). Because of (17),

$$a + |b - \rho| \geq \frac{1}{q_n^{\tau_n}}\,,$$

so that

$$|C(a,b)| \le C\,a^{(1+1/\tau_n)/2}\Big(1+\frac{|b-\rho|}{a}\Big)^{(1/\tau_n-1)/2};$$

and because of (12) this upper bound becomes an equality if we choose $a = 1/q_n^{\tau_n}$, $b = 0$, hence (18) and (19) in this case, and we also have proved their optimality.

Second case: $\mathrm{Im}\,(z)/q_n^2|z|^2 \le 1$; we separate this case into two subcases:

First subcase: p_{n-1} and q_{n-1} are not both odd; then

$$\Big|\theta\big(\frac{p_n}{q_n}+z\big)\Big| \le \frac{1}{\sqrt{q_n|z|}}\Big(\frac{\mathrm{Im}\,(z)}{q_n^2\,|z|^2}\Big)^{-1/2} = \sqrt{\frac{q_n|z|}{Im(z)}}\,, \tag{25}$$

so that, since $|z| \ge 2\,|\rho - p_n/q_n|$,

$$|C(a,b)| \le 2\sqrt{a\,q_n(a+|b-\rho|)}\,. \tag{26}$$

Because of (17),

$$a+|b-\rho| \le 6\Big|\rho-\frac{p_{n-1}}{q_{n-1}}\Big| \le 6\Big(\frac{1}{q_{n-1}}\Big)^{\tau_{n-1}} \le 6\Big(\frac{1}{q_n}\Big)^{\tau_{n-1}/(\tau_{n-1}-1)};$$

thus

$$\begin{aligned} |C(a,b)| &\le C\,a\,q_n^{1/2}\Big(1+\frac{|b-\rho|}{a}\Big)^{1/2}\\ &\le C\,a\Big(\frac{1}{a+|b-\rho|}\Big)^{(\tau_{n-1}-1)/(2\tau_{n-1})}\Big(1+\frac{|b-\rho|}{a}\Big)^{1/2}\\ &\le C\,a^{1/2+1/(2\tau_{n-1})}\Big(1+\frac{|b-\rho|}{a}\Big)^{1/(2\tau_{n-1})}; \end{aligned} \tag{27}$$

hence (20).

Second Subcase: p_{n-1} and q_{n-1} are both odd; then

$$r_n = (-1)^n q_{n-1} + q_n\,, \qquad s_n = (-1)^{n+1}p_{n-1} - p_n\,.$$

We now want to estimate θ near the points p_n/q_n where p_n and q_n are both odd; we will deduce this estimate from (15).

We see that (24) becomes

$$\Big|\theta\big(\frac{p_n}{q_n}+z\big)\Big| = \Big|\theta\big(\frac{(-1)^n q_{n-1}}{q_n} - \frac{1}{q_n^2 z} + 1\big)\Big| \frac{1}{\sqrt{q_n|z|}}$$

let

$$g_n(z) = \frac{(-1)^n q_n q_{n-1} z - 1}{q_n^2 z}$$

From (14), we get

$$\Big|\theta\big(\frac{p_n}{q_n}+z\big)\Big| = \frac{1}{\sqrt{q_n|z|\,|g_n(z)|}} \Big|\theta\big(\frac{-1}{4\,g_n(z)}\big) - \theta\big(\frac{-1}{g_n(z)}\big)\Big|.$$

Remark that

$$\operatorname{Im}\big(\frac{-1}{g_n(z)}\big) = \frac{\operatorname{Im}(g_n(z))}{|g_n(z)|^2} = \frac{\operatorname{Im}(z)}{q_n^2\,|z|^2\,|g_n(z)|^2}.$$

If $\operatorname{Im}(-1/g_n(z)) \geq 1$, using (15),

$$\begin{aligned}\Big|\theta\big(\frac{p_n}{q_n}+z\big)\Big| &\leq \frac{1}{\sqrt{q_n\,|z|\,|g_n(z)|}} \exp\Big(-\pi\,\frac{\operatorname{Im}(z)}{q_n^2\,|z|^2|\,g_n(z)|^2}\Big)\\ &\leq \frac{1}{\sqrt{q_n\,|z|\,|g_n(z)|}} \Big(\frac{q_n^2\,|z|^2\,|g_n(z)|^2}{\operatorname{Im}(z)}\Big)^{1/4}\\ &\leq \Big(\frac{1}{\operatorname{Im}(z)}\Big)^{1/4}\end{aligned}$$

so that $|C(a,b)| \leq a^{3/4}$; hence (18) in that case.

Suppose now that $\operatorname{Im}(-1/g_n(z)) \leq 1$. Then

$$\begin{aligned}\Big|\theta\big(\frac{p_n}{q_n}+z\big)\Big| &\leq \frac{1}{\sqrt{q_n\,|z|\,|g_n(z)|}} \Big(\operatorname{Im}\big(\frac{-1}{g_n(z)}\big)\Big)^{-1/2}\\ &\leq \frac{1}{\sqrt{q_n\,|z|\,|g_n(z)|}}\,\frac{q_n\,|z|\,|g_n(z)|}{\sqrt{\operatorname{Im}(z)}}\\ &\leq \sqrt{\frac{q_n\,|z|\,|g_n(z)|}{\operatorname{Im}(z)}}\,.\end{aligned}$$

Because of (17), $|z| \leq 6/(q_n q_{n-1})$, so that $|g_n(z)| \leq 7/(q_n^2|z|)$ and

$$\Big|\theta\big(\frac{p_n}{q_n}+z\big)\Big| \leq \sqrt{\frac{q_n\,|z|}{q_n^2\,|z|\operatorname{Im}(z)}} \leq \frac{1}{\sqrt{q_n\operatorname{Im}(z)}}\,.$$

Thus

$$|C(a,b)| \le \frac{\sqrt{a}}{\sqrt{q_n}} .$$

From (17), we have $(1/q_n)^{\tau_n} \le (a + |b - \rho|)$ so that

$$|C(a,b)| \le a^{1/2+1/(2\tau_n)} \Big(1 + \frac{|b-\rho|}{a}\Big)^{1/(2\tau_n)} ;$$

Hence (18) in that case.

6. The case when p_n and q_n are both odd.

Following the same procedure as in the previous section, we first determine $\gamma_n = (ax+b)/(cx+d)$ such that $\gamma_n(p_n/q_n) = 1$. We choose either $r_n = q_{n+1}$, $s_n = p_{n+1}$ or $r_n = -q_{n+1}$, $s_n = -p_{n+1}$ such that

$$p_n\, r_n - s_n\, q_n = 1$$

(which is possible because of (22)). Now γ_n is defined by the coefficients

$$a = q_n + r_n\,, \quad b = -p_n - s_n\,, \quad c = r_n\,, \quad d = -s_n\,.$$

One easily checks that (10) holds and thus γ_n belongs to the Theta modular group;

$$\gamma_n\Big(\frac{p_n}{q_n} + z\Big) = \frac{1 + q_n\,(r_n + q_n)\,z}{1 + r_n\,q_n\,z} = 1 + f_n(z)$$

with

$$f_n(z) = \frac{q_n^2\, z}{1 + r_n\, q_n\, z}$$

so that, because of (11),

$$\Big|\theta\Big(\frac{p_n}{q_n} + z\Big)\Big| = |\theta(1 + f_n(z))|\, \frac{1}{r_n^{1/2}\Big(z + \frac{p_n}{q_n} - \frac{p_{n+1}}{q_{n+1}}\Big)^{1/2}}$$

but (17) implies

$$|z| \ge 3\Big|\rho - \frac{p_n}{q_n}\Big| \ge \frac{3}{2}\Big|\frac{p_n}{q_n} - \frac{p_{n+1}}{q_{n+1}}\Big| ,$$

so that

$$\left|\theta\left(\frac{p_n}{q_n}+z\right)\right| \le \frac{|\theta(1+f_n(z))|}{r_n^{1/2}\,|z|^{1/2}}\,.$$

Remark that the condition $\mathrm{Im}\,(-1/f_n(z)) \ge 1$ is equivalent to $\mathrm{Im}\,(z) \ge q_n^2\,|z|^2$. Thus we now consider the two following cases.

First Case: $\mathrm{Im}\,(z) \ge q_n^2\,|z|^2$. In that case, because of (15),

$$\begin{aligned}\left|\theta\left(\frac{p_n}{q_n}+z\right)\right| &\le C\,\exp\left(\frac{-\pi\,\mathrm{Im}\,(z)}{q_n^2\,|z|^2}\right)\left(\frac{1}{|f_n(z)\,r_n\,z|}\right)^{1/2}\\ &\le C\,\exp\left(\frac{-\pi\,\mathrm{Im}\,(z)}{q_n^2\,|z|^2}\right)\frac{|1+r_n\,q_n\,z|^{1/2}}{(|r_n z|)^{1/2}|q_n^2 z|^{1/2}}\,.\end{aligned}$$

Because of (16),

$$1/|r_n q_n| \le 2/q_n^{\tau_n}\,,,$$

and thus

$$|r_n q_n z| \ge \frac{3}{2}\,.$$

Thus

$$\begin{aligned}\left|\theta\left(\frac{p_n}{q_n}+z\right)\right| &\le C\,\exp\left(\frac{-\pi\,\mathrm{Im}\,(z)}{q_n^2|z|^2}\right)\frac{1}{|z|^{1/2}q_n^{1/2}}\\ &\le C\left(\frac{q_n^2|z|^2}{\mathrm{Im}\,(z)}\right)^{1/4}\frac{1}{|z|^{1/2}\,q_n^{1/2}}=\frac{C}{\mathrm{Im}\,(z)^{1/4}}\end{aligned}$$

so that

$$|C(a,b)| \le a^{3/4}\,;$$

hence (21) in that case.

Second Case: $\mathrm{Im}\,(z) \le q_n^2|z|^2$. In that case, from (13), (14) and (28), we obtain

$$\begin{aligned}\left|\theta\left(\frac{p_n}{q_n}+z\right)\right| &\le \frac{1}{\sqrt{|f_n(z)r_n z|}}\left|\theta\left(\frac{-1}{4\,f_n(z)}\right)-\theta\left(\frac{-1}{f_n(z)}\right)\right|\\ &\le C\,\frac{q_n\,|z|^{1/2}}{\sqrt{r_n\,|f_n(z)|\,\mathrm{Im}\,(z)}}\\ &\le C\,\frac{|1+r_n\,q_n\,z|^{1/2}}{\sqrt{r_n\,\mathrm{Im}\,(z)}}\\ &\le C\,\frac{|q_n\,z|^{1/2}}{\sqrt{\mathrm{Im}\,(z)}}\end{aligned}$$

as above $|r_n\, q_n\, z| \geq 3/2$, so that

$$|C(a,b)| \leq Ca^{1/2}\sqrt{q_n(a+|b-\rho|)}\,.$$

Because of (17),

$$a + |b-\rho| \leq 3\,(1/q_n)^{\tau_{n-1}/(\tau_{n-1}-1)}\,,$$

so that

$$|C(a,b)| \leq C\,a^{1/2+1/(2\tau_{n-1})}\left(1+\frac{|b-\rho|}{a}\right)^{1/(2\tau_{n-1})}.$$

Thus (21) holds in this case and Proposition 3 is proved.

The fact that (18) and (21) cannot be improved in a cone

$$\mathrm{Im}\,(z-\rho) \geq C\,\mathrm{Re}\,(z-\rho)$$

yields a slightly more precise information than the fact that φ is not smoother than $1/2 + 1/(2\eta(\rho))$ at ρ because it shows that φ has no chirp expansion at an irrational point ρ (see [9]); thus the only points where φ has a chirp expansion are the rationals of the form odd / odd. It also shows that fractional integrals of φ of order s will be exactly $\Gamma^{s+1/2+1/(2\eta(\rho))}(\rho)$. Actually, from Proposition 2, and the chirp characterization given in [13], one easily obtains the following corollary.

Corollary 2. *Let*

$$\varphi_s(x) = \sum_1^{\infty} \frac{1}{n^{2+2s}} \sin \pi n^2 x\,.$$

If $s \in (-1/2, +\infty)$ and if ρ is not a rational quotient of two odd numbers, $\varphi_s \in \Gamma^{s+1/2+1/(2\eta(\rho))}(\rho)$ and the spectrum of singularities of φ_s is given by

$$d(\alpha) = \begin{cases} 4(\alpha - s) - 2\,, & \text{if } \alpha \in \left[s+\frac{1}{2}, s+\frac{3}{4}\right], \\ 0\,, & \text{if } \alpha = 2s + \frac{3}{2}\,, \\ -\infty\,, & \text{else;}\,. \end{cases}$$

Acknowledgement. The author is particularly thankful to Yves Meyer for many remarks and comments on this text.

References.

[1] Arneodo, A., Bacry, E., Muzy, J. F., The thermodynamics of fractal revisited with wavelets. *Physica A* **213** (1995), 232-275.

[2] Apostol, T. M., *Modular functions and Dirichlet series in number theory.* Springer Verlag, 1976.

[3] Daubechies, I., Lagarias, J., On the thermodynamic formalism for multifractal functions. In *The State of Matter* (dedicated to Elliot Lieb). Eds. M. Aizenman and H. Araki, World Scientific (1994), 213-265.

[4] Duistermaat, J. J., Selfsimilarity of Riemann's non-differentiable function. *Nieuw Archief voor Wiskunde* **9** (1991), 303-337.

[5] Falconer, K., *Fractal geometry.* John Wiley and sons, 1990.

[6] Frisch, U., Parisi, Fully developped turbulence and intermittency. *Proc. International Summer school Phys. Enrico Fermi*, 84-88, North Holland, 1985.

[7] Gerver, J., The differentiability of the Riemann function at certain rational multiples of π. *Amer. J. Math.* **92** (1970), 33-55.

[8] Hardy, G. H., Weierstrass's non-differentiable function. *Trans. Amer. Math. Soc.* **17** (1916), 301-325.

[9] Hardy, G. H., Littlewood, J. E., Some problems of Diophantine approximation. *Acta Math.* **37** (1914), 193-239.

[10] Holschneider, M., Tchamitchian, P., Pointwise analysis of Riemann's "non differentiable" function. *Invent. Math.* **105** (1991), 157-176.

[11] Jaffard S., Pointwise smoothness, two-microlocalization and wavelet coefficients. *Publ. Mat.* **35** (1991), 155-168.

[12] Jaffard S., Multifractal Formalism for functions Part I: Results valid for all functions et Part II: Selfsimilar functions. To appear in *SIAM Math. Anal.*

[13] Jaffard S., Meyer Y., Wavelet methods for pointwise regularity and local oscillations of functions. To appear in *Memoirs Amer. Math. Soc.*

[14] Kahane, J. P., *Some Random Series of Functions.* Cambridge Studies in Advanced Mathematics **5**, Cambridge Univ. Press, 1993.

[15] Knopp, M., *Modular functions in analytic number theory.* Markham Publ., 1970.

[16] Valiron, G., *Théorie des fonctions.* Masson, 1942.

[17] Van der Porten, An introduction to continued fractions, in *Diophantine Analysis,* Cambridge Univ. Press, London Lecture notes series **109** (1986), 99-138.

[18] Zalcwasser, Z., Sur les polynomes associés aux fonctions modulaires θ. *Studia Math.* **7** (1938), 16-35.

Recibido: 7 de mayo de 1.995

S. Jaffard
Departement de Mathematiques
Université Paris XII
61 Av. du General de Gaulle
94010 Creteil Cedex, FRANCE

and

C.M.L.A.
E.N.S. Cachan
61 Av. du Président Wilson
94235 Cachan Cedex, FRANCE
jaffard@cmla.ens-cachan.fr

Revista Matemática Iberoamericana
Vol. 12, N.° 2, 1996

On the uniqueness problem for quasilinear elliptic equations involving measures

Tero Kilpeläinen and Xiangsheng Xu

Abstract. We discuss the uniqueness of solutions to problems like

$$\begin{cases} \lambda\,|u|^{s-1}u - \operatorname{div}(|\nabla u|^{p-2}\,\nabla u) = \mu & \text{on } \Omega\,,\\ u = 0 & \text{in } \partial\Omega\,, \end{cases}$$

where $\lambda \geq 0$ and μ is a signed Radon measure.

1. Introduction.

Throughout this paper we let Ω be a bounded open set in $\mathbb{R}^n$ and $1 < p \leq n$ a fixed number with $p > 2 - 1/n$.[1] Suppose that μ is a signed Radon measure in Ω with finite total variation. We consider the solutions $u \in W^{1,1}_{\text{loc}}(\Omega)$ of the equation

$$B(u) - \operatorname{div}\mathcal{A}(x, \nabla u) = \mu\,,$$

[1] The restriction $p>2-1/n$ could be removed by using a generalized derivative as in [5] or a different concept of a solution as *e.g.* in [1] or [9].

where $B\colon \mathbb{R} \longrightarrow \mathbb{R}$ is a continuous increasing function with $B(0) = 0$ and $\mathcal{A}\colon \mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}^n$ is a mapping that satisfies the following assumptions for some numbers $0 < \alpha \le \beta < \infty$:

(1.1) the function $x \mapsto \mathcal{A}(x,\xi)$ is measurable for all $\xi \in \mathbf{R}^n$, and the function $\xi \mapsto \mathcal{A}(x,\xi)$ is continuous for a.e. $x \in \mathbf{R}^n$;

for all $\xi \in \mathbf{R}^n$ and almost every $x \in \mathbf{R}^n$

$$\mathcal{A}(x,\xi)\cdot\xi \ge \alpha\,|\xi|^p\,, \tag{1.2}$$

$$|\mathcal{A}(x,\xi)| \le \beta\,|\xi|^{p-1}\,, \tag{1.3}$$

$$(\mathcal{A}(x,\xi) - \mathcal{A}(x,\zeta))\cdot(\xi-\zeta) > 0\,, \tag{1.4}$$

whenever $\xi \ne \zeta$.

Solutions are understood in the sense of distributions, and we fix weak zero boundary values. More precisely, we consider the problem

$$\begin{cases} B(u) - \operatorname{div}\mathcal{A}(x,\nabla u) = \mu\,, \\ B(u) \in L^1(\Omega)\,, \\ u \in W^{1,\max\{p-1,1\}}_{\text{loc}}(\Omega)\,, \\ T_k(u) \in W^{1,p}_0(\Omega) \ \text{ for } k > 0\,, \end{cases} \tag{1.5}$$

where T_k is the double side truncating operator at the level k,

$$T_k(t) = \max\big\{\min\{t,k\}, -k\big\}\,.$$

Here the first line in (1.5) means that

$$\int_\Omega B(u)\,\varphi\,dx + \int_\Omega \mathcal{A}(x,\nabla u)\cdot\nabla\varphi\,dx = \int_\Omega \varphi\,d\mu$$

for each $\varphi \in C^\infty_0(\Omega)$.

The prime examples of such equations arise from the p-Laplacian operator

$$\Delta_p u = \operatorname{div}(|\nabla u|^{p-2}\,\nabla u)\,.$$

Keeping this example in mind one easily convinces oneself that, for an arbitrary measure μ, there is no hope to find a solution from the "natural" Sobolev space $W^{1,p}_0(\Omega)$. Indeed, existence of a solution in this space automatically implies that μ is in the dual of $W^{1,p}_0(\Omega)$. Moreover, it is well known that this dual does not contain point measures for $1 < p \leq n$ (see *e.g.* the discussion before Theorem 3.5 below).

Therefore we only require that the truncations of a solution be in $W_0^{1,p}(\Omega)$. Then, using compactness arguments we find that the solution itself lies in $W_0^{1,q(p-1)}(\Omega)$ for each $1 \leq q < n/(n-1)$.

There are several papers, where the authors discuss the existence of problems like (1.5) in different senses, see *e.g.* [7], [2], [5]. In the nonlinear case, there are a few results aiming at the treatment of the question of uniqueness: Lions and Murat have announced an existence and uniqueness result for renormalized solutions in the case when $p = 2$ and $\mu \in L^1$ (see [8]); unfortunately, we haven't seen their proof. Two different approaches to the general case with $\mu \in L^1$ are given in [1] and in [9]. Rakotoson [9] uses renormalized solutions, and Bénilan *et. al.* [1] an "entropy condition" which we shall adopt and modify. We shall consider measures μ that are absolutely continuous with respect to p-capacity (see 2.1 below), in particular, L^1-functions are particular cases of our consideration. We prove:

1.6. Theorem. *Let μ be a finite signed measure in Ω that is absolutely continuous with respect to p-capacity. Then there is a unique solution u of* (1.5) *such that for $\sigma \in \{+,-\}$*

$$\int_\Omega \mathcal{A}(x,\nabla u)\cdot\nabla T_k^\sigma(u-\varphi)\,dx + \int_\Omega B(u)\,T_k^{6\sigma}(u-\varphi)\,dx = \int_\Omega T_k^\sigma(u-\varphi)\,d\mu\,,$$

whenever $\varphi \in C_0^\infty(\Omega)$ and $k > 0$. Moreover, $u \in W_0^{1,q(p-1)}(\Omega)$ for each $1 \leq q < n/(n-1)$.

Here

$$T_k^+(t) = \max\big\{\min\{t,k\},0\big\}$$

and

$$T_k^-(t) = \min\big\{\max\{t,-k\},0\big\}$$

are the positive and negative truncating operators. Take notice that here and in what follows we always take the quasicontinuous, hence Borel, representatives of Sobolev functions; hence there are no problems with measurability.

To display a simple example that motivates the use of a constraint for the solutions, consider the p-Laplacian

$$\Delta_p u = 0$$

in the punctured ball $\Omega = B(0,1)\setminus\{0\}$. Then the identical zero function is a trivial solution of (1.5) in Ω (there $B = 0$, $\mu = 0$ and $\mathcal{A}(x,\xi) = |\xi|^{p-2}\xi$). Another solution is given by

$$u(x) = \begin{cases} |x|^{(p-n)/(p-1)} - 1\,, & \text{if } p < n\,, \\ \log|x|\,, & \text{if } p = n\,. \end{cases}$$

Observe that these functions both are SOLAs (solutions obtained as limits of approximations) in the sense of [3].

Note that the assumption that $p \le n$ is no restriction, for any finite Radon measure belongs to the dual of the Sobolev space $W_0^{1,q}(\Omega)$ if $q > n$ and then the unique solvability of (1.5) is well known. On the contrary, the assumption that $p > 2 - 1/n$ is partly essential and partly purely technical. It is a simple matter to construct measures μ for which there cannot be any solutions with locally integrable distributional derivatives if $p \le 2 - 1/n$. There are at least two different ways out of this trouble: either one could consider a generalized gradient as it was done in [5], or to leave distributional solutions and work with renormalized solutions as in [9] or [11]. We leave these technicalities to the interested reader.

2. Uniqueness.

To begin with, we recall that the Sobolev space $W^{1,q}(\Omega)$, $1 \le q < \infty$, consists of all q-integrable functions u whose first distributional derivative ∇u is also q-integrable in Ω; equipped with the norm

$$\|u\|_{1,q} = \Big(\int_\Omega (|u|^q + |\nabla u|^q)\,dx\Big)^{1/q},$$

$W^{1,q}(\Omega)$ is a Banach space. The corresponding local space is marked as $W^{1,q}_{\mathrm{loc}}(\Omega)$. Moreover, $W_0^{1,q}(\Omega)$ stands for the closure of $C_0^\infty(\Omega)$ in $W^{1,q}(\Omega)$.

Next we define the *p-capacity of the set* $E \subset \mathbb{R}^n$ to be the number

$$\mathrm{cap}_p(E) = \inf \int_{\mathbb{R}^n} (|\varphi|^p + |\nabla\varphi|^p)\,dx\,,$$

where the infimum is taken over all $\varphi \in W^{1,1}_{\mathrm{loc}}(\mathbb{R}^n)$ such that $\varphi = 1$ on an open set containing E. Then cap_p defines an outer measure, but there are only very few measurable sets. The p-capacity is intimately connected with Sobolev spaces $W^{1,p}$ and with p-type equations (1.5), see *e.g.* [4], [12]. In particular each $u \in W^{1,p}(\Omega)$ has a quasicontinuous version, *i.e.* there is v such that $u = v$ almost everywhere and for each $\varepsilon > 0$ there is an open set G such that $\mathrm{cap}_p(G) < \varepsilon$ and the restriction to $\Omega \setminus G$ of v is continuous and real-valued.

We say that μ is *absolutely continuous with respect to p-capacity* if

$$\mu(E) = 0 \quad \text{whenever} \quad \mathrm{cap}_p(E) = 0\,. \tag{2.1}$$

Note that the Hausdorff dimension of a set of p-capacity zero is at most $n - p$, while a set with finite $n - p$ dimensional Hausdorff measure is of p-capacity zero, see *e.g.* [4].

In this section we establish uniqueness under a slightly weaker condition than was stated in Theorem 1.6. We say that a solution u of (1.5) satisfies the *entropy condition* if for $\sigma \in \{+, -\}$

$$\begin{aligned} \int_\Omega \mathcal{A}(x, \nabla u) \cdot \nabla T_k^\sigma(u - \varphi)\, dx + \int_\Omega B(u)\, T_k^\sigma(u - \varphi)\, dx \\ \leq \int_\Omega T_k^\sigma(u - \varphi)\, d\mu \end{aligned} \tag{2.2}$$

for all $\varphi \in C_0^\infty(\Omega)$ and $k > 0$;

In particular, we have that

$$\int_\Omega \mathcal{A}(x, \nabla u) \cdot \nabla T_k(u - \varphi)\, dx + \int_\Omega B(u)\, T_k(u - \varphi)\, dx \leq \int_\Omega T_k(u - \varphi)\, d\mu$$

whenever T_k is the double side truncating operator.

2.3. Lemma. *If u is a solution that satisfies the entropy condition* (2.2), *then for each $M > 0$ and $k > 0$*

$$\int_{\{k \leq u \leq k+M\}} |\nabla u|^p\, dx \leq c\, M\, |\mu|(\{|u| > k\}) + c\, M \int_{\{|u|>k\}} |B(u)|\, dx\,.$$

PROOF. An easy approximation shows that one can replace φ in (2.2) by any bounded function from $W_0^{1,p}$ (see [1, Lemma 3.3]). In particular,

$$\begin{aligned} c \int_{\{k\le|u|\le k+M\}} |\nabla u|^p\,dx &\le \int_\Omega \mathcal{A}(x,\nabla u)\cdot\nabla T_M(u-T_k u)\,dx \\ &\le \int_\Omega T_M(u-T_k u)\,d\mu \\ &\quad - \int_\Omega B(u)\,T_M(u-T_k u)\,dx \\ &\le M\,|\mu|(\{|u|>k\}) \\ &\quad + M\int_{\{|u|>k\}} |B(u)|\,dx\,, \end{aligned}$$

as desired.

2.4. Corollary. *Let u be a solution that satisfies the entropy condition* (2.2). *If $|\mu|(\{|u|=\infty\})=0$, then*

$$\lim_{k\to\infty} \int_{\{k\le|u|\le k+M\}} |\nabla u|^p\,dx = 0\,.$$

Corollary 2.4 is in general false if $|\mu|(\{|u|=\infty\})>0$. Take, for instance, $\mu=$ the Dirac measure. Then if $\mathcal{A}(x,\xi)=|\xi|^{p-2}\xi$ is the p-Laplacian, we have

$$\lim_{k\to\infty} \int_{\{k\le u\le k+M\}} |\nabla u|^p\,dx = M\,.$$

In this paper, we restrict our consideration to measures which are absolutely continuous with respect to p-capacity. Then $|\mu|(\{|u|=\infty\})=0$ for p-quasicontinuous u.

2.5. Theorem. *Let μ_1 and μ_2 be finite signed Radon measures that are absolutely continuous with respect to p-capacity such that $\mu_1\le\mu_2$. If u and v are solutions of* (1.5) *with measures μ_1 and μ_2, respectively, that satisfy the entropy condition* (2.2)*, then $u\le v$.*

PROOF. By approximation,

$$\int_\Omega \mathcal{A}(x,\nabla u)\cdot\nabla T_k^+(u-T_lv)\,dx \le \int_\Omega T_k^+(u-T_lv)\,d\mu_1 - \int_\Omega B(u)T_k^+(u-T_lv)\,dx$$

and

$$\int_\Omega \mathcal{A}(x,\nabla v)\cdot\nabla T_k^-(v-T_lu)\,dx \le \int_\Omega T_k^-(v-T_lu)\,d\mu_2 - \int_\Omega B(v)T_k^-(v-T_lu)\,dx\,.$$

If we add these inequalities up and let $l\to\infty$, the right hand side is treated by the aid of the dominated convergence theorem and its limit is

$$\int_\Omega T_k^+(u-v)\,d\mu_1 - \int_\Omega T_k^+(u-v)\,d\mu_2 - \int_\Omega \big(B(u)-B(v)\big)T_k^+(u-v)\,dx \le 0\,,$$

since $\mu_1\le\mu_2$ and B is increasing. The set of integration on the left hand side is splitted into four parts:

$$\begin{aligned} G_1 &= \{|u-v|\le k\,,\ |v|\le l\,,\ \text{and } |u|\le l\}\,,\\ G_2 &= \{|u-v|>k\}\,,\\ B_1 &= \{|u-v|\le k\,,\ |v|\le l\,,\ \text{and } |u|>l\}\,,\\ B_2 &= \{|u-v|\le k\,,\ |v|>l\,,\ \text{and } |u|\le l\}\,. \end{aligned}$$

The parts B_1 and B_2 are symmetric and they tend to zero as is seen with an estimation like

$$\Big|\int_{B_1}\mathcal{A}(x,\nabla u)\cdot\nabla T_k^+(u-T_lv)\,dx\Big| \le c\int_{B_1}|\nabla u|^p\,dx + c\int_{B_1}|\nabla u|^{p-1}|\nabla v|\,dx$$

$$\le c\int_{\{l\le|u|\le l+k\}}|\nabla u|^p\,dx$$

$$
\begin{aligned}
&+ c\Big(\int\limits_{\{l\le|u|\le l+k\}} |\nabla u|^p\,dx\Big)^{p/(p-1)} \\
&\cdot\Big(\int\limits_{\{l-k\le|v|\le l\}} |\nabla v|^p\,dx\Big)^{1/p} \\
&\to 0+\,,
\end{aligned}
$$

as $l \to \infty$ by Corollary 2.4. Further,

$$
\Big|\int_{B_1} \mathcal{A}(x,\nabla v)\cdot\nabla T_k^-(v - T_l u)\,dx\Big| \le c \int\limits_{\{l-k\le|v|\le l\}} |\nabla v|^p\,dx \to 0\,,
$$

as $l \to \infty$. Next we estimate the integrals over G_2. For instance

$$
\Big|\int_{G_2} \mathcal{A}(x,\nabla u)\cdot\nabla T_k^+(u - T_l v)\,dx\Big| \le c \int\limits_{\{l-k\le|u|\le l+k\}} |\nabla u|^p\,dx \to 0
$$

and the other integral is treated similarly.

Hence we conclude that the integral over G_1 tends to a nonpositive number as $l \to \infty$, and hence

$$
\int\limits_{\{|u-v|\le k,\,u>v\}} \big(\mathcal{A}(x,\nabla u) - \mathcal{A}(x,\nabla v)\big)\cdot\big(\nabla u - \nabla v\big)\,dx \le 0\,.
$$

Since this last integrand strictly positive if $\nabla u \ne \nabla v$, we have $\nabla u = \nabla v$ almost everywhere in the set where $|u - v| \le k$ and $u > v$. Letting $k \to \infty$ we find that $u \le v$ in Ω in the view of the weak boundary values. The proof is complete.

2.6. Corollary. *If μ is absolutely continuous with respect to p-capacity, then there is at most one solution u of* (1.5) *that satisfies the entropy condition.*

3. Existence.

There are various proofs for the existence of solutions to problem (1.5). Because we want that a particular solution satisfies the entropy condition, we have to give a proof that results in the entropy equality as well.

We start our investigation with a compactness lemma.

3.1. Lemma. *Let μ_j be a sequence of signed Radon measures that belong to the dual of $W_0^{1,p}(\Omega)$ such that*

$$|\mu_j|(\Omega) \le M < \infty$$

for each j. Let $u_j \in W_0^{1,p}(\Omega)$ be such that $B(u_j) \in L^1(\Omega)$ and

$$B(u_j) - \operatorname{div} \mathcal{A}(x, \nabla u_j) = \mu_j$$

in Ω. Then there is a subsequence u_j and a function u such that $u_j \to u$ pointwise almost everywhere and weakly in $W^{1,q(p-1)}$ whenever $1 \le q < n/(n-1) = n'$. Furthermore, $B(u_j)$ is bounded in $L^1(\Omega)$ and $\nabla u_j(x) \to \nabla u(x)$ for almost every x, $\mathcal{A}(x, \nabla u_j) \to \mathcal{A}(x, \nabla u)$ in $L^q(\Omega)$ and for each $k > 0$, the sequence of truncations $\nabla T_k(u_j)$ is bounded in $L^p(\Omega)$.

PROOF. By using the test functions $T_1(u_j/\varepsilon)$, $\varepsilon > 0$, we find that

$$\int_\Omega |B(u_j)|\, dx = \limsup_{\varepsilon \to 0} \left(\int_\Omega T_1(u_j/\varepsilon)\, d\mu_j - \frac{1}{\varepsilon} \int_{\{0<|u_j|<\varepsilon\}} \mathcal{A}(x, \nabla u_j) \cdot \nabla u_j\, dx \right)$$

$$\le |\mu_j|(\Omega) \le M < \infty\,. \tag{3.2}$$

Similarly, the use of the test function $T_k(u_j)$ shows that

$$\int_\Omega |\nabla T_k(u_j)|^p\, dx \le c\, k\, M\,, \tag{3.3}$$

so that, by the usual compactness arguments (see *e.g.* [4, 7.43]), the sequence $|\nabla u_j|^{p-1}$ is bounded in $L^q(\Omega)$ for all $1 \le q < n'$. Then there is $u \in W_0^{1,q(p-1)}(\Omega)$ such that $u_j \to u$ weakly in $W_0^{1,q(p-1)}(\Omega)$. By the aid of the Rellich compactness theorem we can extract a subsequence u_j that converges pointwise to u almost everywhere in Ω.

It remains to show that $\nabla u_j \to \nabla u$ pointwise almost everywhere. Fix $\varepsilon > 0$ and let

$$E_{j,k} = \{x \in \Omega\colon\ \big(\mathcal{A}(x, \nabla u_j) - \mathcal{A}(x, \nabla u_k)\big) \cdot \big(\nabla u_j - \nabla u_k\big) > \varepsilon\}\,.$$

We estimate the measure of $E_{j,k}$:

$$|E_{j,k}| \le |E_{j,k} \cap \{|u_j - u_k| \ge \varepsilon^2\}| + \frac{1}{\varepsilon} \int_{E_{j,k}\cap\{|u_k-u_j|<\varepsilon^2\}} \big(\mathcal{A}(x,\nabla u_j) - \mathcal{A}(x,\nabla u_k)\big) \cdot \big(\nabla u_j - \nabla u_k\big)\,dx\,.$$

Using the test function $T_{\varepsilon^2}(u_j - u_k)$ we find the estimate

$$\begin{aligned}\int_{E_{j,k}\cap\{|u_k-u_j|<\varepsilon^2\}} &\big(\mathcal{A}(x,\nabla u_j) - \mathcal{A}(x,\nabla u_k)\big) \cdot \big(\nabla u_j - \nabla u_k\big)\,dx \\ &\le \int_\Omega T_{\varepsilon^2}(u_j - u_k)\,d\mu_j - \int_\Omega T_{\varepsilon^2}(u_j - u_k)\,d\mu_k \\ &\quad - \int_\Omega B(u_j)\,T_{\varepsilon^2}(u_j - u_k)\,dx + \int_\Omega B(u_k)\,T_{\varepsilon^2}(u_j - u_k)\,dx \\ &\le c\,\varepsilon^2\end{aligned}$$

by what we proved above. Hence we arrive at the estimate

$$|E_{j,k}| \le c\,\varepsilon + |E_{j,k} \cap \{|u_j - u_k| \ge \varepsilon^2\}|\,, \tag{3.4}$$

where the constant c is independent of j, k, and ε.

Since $u_j \to u$ almost everywhere we easily infer from (3.4) and the monotonicity and continuity assumptions on $\mathcal{A}$ that ∇u_j converges pointwise almost everywhere to a function that must coincide with ∇u.

Now we consider a nonnegative finite Radon measure μ on Ω. We may as well assume that μ is defined on the whole of $\mathbb{R}^n$ with $\mu(\mathbb{R}^n \setminus \Omega) = 0$. Then $\mu \in \big(W_0^{1,p}(\Omega)\big)^*$ if and only if

$$\int_{\mathbb{R}^n} \mathbf{W}^\mu_{1,p}(x,1)\,d\mu < \infty\,,$$

where

$$\mathbf{W}^\mu_{1,p}(x,1) = \int_0^1 \left(\frac{\mu(B(x,r))}{r^{n-p}}\right)^{1/(p-1)} \frac{dr}{r}$$

is the Wolff potential (see *e.g.* [12, Theorem 4.7.5]).

Now we find a solution for which the entropy inequality (2.2) is an equality.

3.5. Theorem. *Let μ be a nonnegative finite measure in Ω with*

$$\mu(\{x\colon\ \mathbf{W}^{\mu}_{1,p}(x,1)=\infty\})=0\,.$$

Then there is a solution u of (1.5) *such that for $\sigma\in\{+,-\}$*

$$\int_\Omega \mathcal{A}(x,\nabla u)\cdot\nabla T^\sigma_k(u-\varphi)\,dx+\int_\Omega B(u)\,T^\sigma_k(u-\varphi)\,dx=\int_\Omega T^\sigma_k(u-\varphi)\,d\mu\,,$$

whenever $\varphi\in C^\infty_0(\Omega)$ and $k>0$.

PROOF. For a nonnegative integer j, let

$$E_j=\{x\colon\ \mathbf{W}^{\mu}_{1,p}(x,1)\le j\}\,,$$

and let μ_j be the restriction to E_j of μ,

$$\mu_j(E)=\mu(E\cap E_j)\,.$$

Then $0\le\mu_j\le\mu_{j+1}\le\mu$ and $\mu_j\to\mu$ weakly, for

$$\mu(\{x\colon\ \mathbf{W}^{\mu}_{1,p}(x,1)=\infty\})=0\,.$$

Since

$$\int_{\mathbb{R}^n}\mathbf{W}^{\mu_j}_{1,p}(x,1)\,d\mu_j\le\int_{\mathbb{R}^n} j\,d\mu_j\le j\,\mu(\Omega)<\infty\,,$$

we have $\mu_j\in\big(W^{1,p}_0(\Omega)\big)^*$. Hence there is a unique $u_j\in W^{1,p}_0(\Omega)$ such that $B(u_j)\in L^1(\Omega)$ and

$$B(u_j)-\operatorname{div}\mathcal{A}(x,\nabla u_j)=\mu_j \tag{3.6}$$

in Ω (see *e.g.* [10] or [7]). Using Lemma 3.1 we find a subsequence of u_j increasing to a function u such that $B(u)\in L^1(\Omega)$ and

$$B(u)-\operatorname{div}\mathcal{A}(x,\nabla u)=\mu$$

in Ω with weak boundary values.

The entropy equality for u is verified as follows: fix $\varphi\in C^\infty_0(\Omega)$. Then for each $k>0$ we have

$$\int_\Omega \mathcal{A}(x,\nabla u_j)\cdot\nabla T^\sigma_k(u-\varphi)\,dx+\int_\Omega B(u_j)T^\sigma_k(u-\varphi)\,dx$$
$$=\int_\Omega T^\sigma_k(u-\varphi)\,d\mu_j\,.$$

Letting $j \to \infty$ this gives us the desired equality. Indeed, the second integral does not cause any troubles, for $B(u_j) \to B(u)$ in L^1 since u_j increases to u. The first integral is treated by the aid of (3.3): for $M \geq k + \sup|\varphi|$ we have

$$\begin{aligned}\int_\Omega &\mathcal{A}(x,\nabla u_j)\cdot\nabla T_k^\sigma(u-\varphi)\,dx\\ &= \int_{\{u\leq M\}} \mathcal{A}(x,\nabla T_M(u_j))\cdot\nabla T_k^\sigma(u-\varphi)\,dx\\ &\to \int_{\{u\leq M\}} \mathcal{A}(x,\nabla T_M(u))\cdot\nabla T_k^\sigma(u-\varphi)\,dx\\ &= \int_\Omega \mathcal{A}(x,\nabla u)\cdot\nabla T_k^\sigma(u-\varphi)\,dx\,,\end{aligned}$$

since the sequence u_j is increasing and $\nabla u_j \to \nabla u$ pointwise almost everywhere. Finally,

$$\int_\Omega T_k^\sigma(u-\varphi)\,d\mu_j = \int_\Omega T_k^\sigma(u-\varphi)\,\chi_{E_j}\,d\mu \to \int_\Omega T_k^\sigma(u-\varphi)\,d\mu\,,$$

where χ_{E_j} stands for the characteristic function of the set E_j.

3.7. REMARK. If μ is in the dual of $W_0^{1,p}(\Omega)$, then

$$\mu(\{x:\ \mathbf{W}_{1,p}^\mu(x,1)=\infty\})=0\,.$$

Consequently, if μ is such that

$$\mu(\{x:\ \mathbf{W}_{1,p}^\mu(x,1)=\infty\})>0\,,$$

or equivalently[2], if μ is not absolutely continuous with respect to p-capacity, then there does not exist any increasing sequence of nonnegative Radon measures $\mu_j \in (W_0^{1,p}(\Omega))^*$ with $\mu_j \to \mu$ weakly.

---

[2] Indeed, the set where $\mathbf{W}_{1,p}^\mu(x,1)=\infty$ is of p-capacity zero by [6]. On the other hand, if $\mu(\{x:\ \mathbf{W}_{1,p}^\mu(x,1)=\infty\})=0$, then as in the previous proof, we find an increasing sequence μ_j of measures from the dual of $W_0^{1,p}(\Omega)$ such that $\mu_j \to \mu$ weakly. Then, since μ_j are absolutely continuous with respect to p-capacity, the same holds for the measure μ.

3.8. Corollary. *Let μ be a nonnegative finite measure in Ω that is absolutely continuous with respect to p-capacity. Then there is a unique solution u of* (1.5) *such that for $\sigma \in \{+,-\}$*

$$\int_\Omega \mathcal{A}(x,\nabla u)\cdot \nabla T_k^\sigma(u-\varphi)\,dx + \int_\Omega B(u)\,T_k^\sigma(u-\varphi)\,dx = \int_\Omega T_k^\sigma(u-\varphi)\,d\mu$$

whenever $\varphi \in C_0^\infty(\Omega)$ and $k > 0$.

PROOF. The uniqueness follows from Corollary 2.6, the existence from Theorem 3.5, for the set

$$E = \{x\colon\ \mathbf{W}_{1,p}^{\mu}(x,1) = \infty\}$$

is of p-capacity zero (there is a p-superharmonic function u such that $u = \infty$ on E by [6]; thus $\operatorname{cap}_p(E) = 0$ by [4, 10.1]).

Next we sketch the existence proof for signed measures.

PROOF OF THEOREM 1.6. The uniqueness was established in Corollary 2.6.

To prove the existence, let $\mu = \mu^+ - \mu^-$, where μ^+ and μ^- are nonnegative measures. Let $\sigma \in \{+,-\}$ and as in the proof of Theorem 3.5, write μ_j^σ for the restriction of μ to the set where $\mathbf{W}_{1,p}^{\mu^\sigma}(x,1) \le j$. Since, for fixed i the measure $\mu_{j,i} = \mu_j^+ - \mu_i^- \in \big(W_0^{1,p}(\Omega)\big)^*$, there is a unique $u_{j,i} \in W_0^{1,p}(\Omega)$ such that $B(u_{j,i}) \in L^1(\Omega)$ and

$$B(u_{j,i}) - \operatorname{div}\mathcal{A}(x,\nabla u_{j,i}) = \mu_{j,i}$$

in Ω. By Lemma 3.1 there is $v_i \in W^{1,q(p-1)}(\Omega)$ such that the truncations $T_k(v_i)$ belong to $W_0^{1,p}(\Omega)$ and $u_{j,i} \to v_i$ weakly in $W^{1,q(p-1)}(\Omega)$ as $j \to \infty$. By the Rellich compactness theorem we have that (a subsequence of) $u_{j,i}$ converges to v_i a.e. and $\mathcal{A}(x,\nabla u_{j,i}) \to \mathcal{A}(x,\nabla v_i)$ weakly in $L^q(\Omega)$. Then, since $u_{j,i}$ increases to v_i, we infer that v_i is a solution of

$$B(v_i) - \operatorname{div}\mathcal{A}(x,\nabla v_i) = \mu^+ - \mu_i^-$$

with $B(v_i) \in L^1(\Omega)$, and the entropy equality is proved almost verbatim as in Theorem 3.5.

Now Theorem 2.5 implies that the sequence v_i is decreasing. Repeating the analysis above one easily sees that the limit function $u = \lim_{i\to\infty} v_i$ is the desired solution; we leave the details to the reader.

3.9. REMARK. Suppose that u is a solution of (1.5). When does it automatically satisfy the entropy condition? The example we gave in the introduction shows that this is not always the case. Suppose that the sets

$$E_j = \{x \in \Omega : \ |u(x)| \geq j\}$$

are compact for j large enough (see the estimates in [6] for the pointwise behavior of u in terms of the potential $\mathbf{W}^{\mu}_{1,p}$). Then $T_k(u-\varphi)$ can be approximated in $W_0^{1,p}(\Omega)$ by $C_0^{\infty}(\Omega)$ functions whose gradients vanish on E_j. Thus it follows that we can plug $T_k(u-\varphi)$ in as a test function, and u therefore satisfies the entropy condition.

References.

[1] Bénilan, P., Boccardo, L., Gallouët, T., Gariepy, R., Pierre, M., and Vázquez, J. L., An L^1-theory of existence and uniqueness of solutions of nonlinear elliptic equations. *Ann. Scuola Norm. Sup. Pisa. Cl. Science, Ser. IV* **22** (1995), 241-273.

[2] Boccardo, L., Gallouët, T., Nonlinear elliptic equations with right-hand side measure. *Comm. Partial Diff. Equations* **17** (1992), 641-655.

[3] Boccardo, L., Murat, F., A property of nonlinear elliptic equations when the right-hand side is a measure. *Potential Analysis* **3** (1994), 257-263.

[4] Heinonen, J., Kilpeläinen, T., Martio, O., *Nonlinear potential theory of degenerate elliptic equations.* Oxford Univ. Press, 1993.

[5] Kilpeläinen, T., Malý, J., Degenerate elliptic equations with measure data and nonlinear potentials. *Ann. Scuola Norm. Sup. Pisa. Cl. Science, Ser. IV* **19** (1992), 591-613.

[6] Kilpeläinen, T., Malý, J., The Wiener test and potential estimates for quasilinear elliptic equations. *Acta Math.* **172** (1994), 137-161.

[7] Leray, J., Lions, J. L., Quelques résultats de Višik sur les problèmes nonlinéaires par les méthodes de Minty-Browder. *Bull. Soc. Math. France* **93** (1965), 97-107.

[8] Lions, P. L., Murat, F., Sur les solutions renormalisées d'équations elliptiques non linéaires. To appear in *C. R. Acad. Sci. Paris Sér. I Math.*

[9] Rakotoson, J. M., Uniqueness of renormalized solutions in a T-set for the L^1-data problem and the link between various formulations. *Indiana Univ. Math. J.* **43** (1994), 685-702.

[10] Xu, X., Existence and convergence theorems for doubly nonlinear partial differential equations of elliptic-parabolic type. *J. Math. Anal. Appl.*

150 (1990), 205-223.

[11] Xu, X., Some results on functional capacity and their applications to p-Laplacian problems involving measure. To appear in *Nonlinear Analysis, TMA*.

[12] Ziemer, W. P., *Weakly differentiable functions: Sobolev spaces and functions of bounded variation.* Springer-Verlag, 1989.

Recibido: 20 de mayo de 1.995

Tero Kilpeläinen
University of Jyväskylä
Department of Mathematics
P.O. Box 35
40351 Jyväskylä, FINLAND
terok@math.jyu.fi

Xiangsheng Xu*
University of Arkansas
Department of Mathematical Sciences
Fayetteville, AR 72701, U.S.A.

and

Department of Mathematics and Statistics
Mississippi State University
Drawer MA
Mississippi State, MS 39762, U.S.A.
xxu@math.msstate.edu

* Research supported partially by an NSF grant (DMS9424448)

Revista Matemática Iberoamericana
Vol. 12, N.° 2, 1996

Goldbach numbers represented by polynomials

Alberto Perelli

1. Introduction.

Let N be a large positive real number. It is well known that almost all even integers in the interval $[N, 2N]$ are Goldbach numbers, *i.e.* a sum of two primes. The same result also holds for short intervals of the form $[N, N+H]$, see Mikawa [4], Perelli-Pintz [7] and Kaczorowski-Perelli-Pintz [3] for the choice of admissible values of H and the size of the exceptional set in several problems in this direction.

One may ask if similar results hold for thinner sequences of integers in $[N, 2N]$, of cardinality smaller than the upper bound for the exceptional set in the above problems. In this paper we deal with the polynomial case.

Let $L = \log N$, $F \in \mathbb{Z}[x]$ with $\deg F = k \geq 1$ and with positive leading coefficient,

$$R(n) = \sum_{r+s=n} \Lambda(r)\, \Lambda(s)$$

and

$$\mathfrak{S}(n) = \prod_{p\nmid n} \Big(1 - \frac{1}{(p-1)^2}\Big) \prod_{p|n} \Big(1 + \frac{1}{p-1}\Big) \, .$$

Our main result is the following

Theorem 1. *Let $k \in \mathbb{N}$, $A, \varepsilon > 0$ and $N^{1/(3k)+\varepsilon} \le H \le N^{1/k}$. Then*

$$\sum_{N^{1/k} \le n \le N^{1/k}+H} |R(F(n)) - F(n)\,\mathfrak{S}(F(n))|^2 \ll_{A,\varepsilon,F} H\,N^2\,L^{-A}\,.$$

For $d \in \mathbb{N}$ let

$$\varrho_F(d) = |\{m\,(\mathrm{mod}\ d):\ F(m) \equiv 0\,(\mathrm{mod}\ d)\}|\,.$$

If $\varrho_F(2) \neq 0$ then, writing

$$A_F(N,H) = \{N^{1/k} \le n \le N^{1/k} + H:\ F(n) \equiv 0\,(\mathrm{mod}\ 2)\}\,,$$

by standard techniques we have that

$$|A_F(N,H)| = \frac{\varrho_F(2)}{2}\,H + O(1)\,.$$

Hence from Theorem 1 we easily obtain the following

Corollary 1. *Let $k \in \mathbb{N}$, $A, \varepsilon > 0$ and $N^{1/(3k)+\varepsilon} \le H \le N^{1/k}$. Then*

$$R(F(n)) = F(n)\,\mathfrak{S}(F(n)) + O_{A,\varepsilon,F}(N\,L^{-A})\,,$$

for all $n \in [N^{1/k}, N^{1/k} + H]$ but $O_{A,\varepsilon,F}(HL^{-A})$ exceptions. In particular, if $\varrho_F(2) \neq 0$, for all $n \in A_F(N,H)$ but $O_{A,\varepsilon,F}(HL^{-A})$ exceptions, $F(n)$ is a Goldbach number.

Since k can be chosen arbitrarily large, the above results provide examples of thin sequences in $[N, 2N]$ of cardinality $O(N^{\delta})$ with $\delta > 0$ arbitrarily small, having the property that almost all their elements are Goldbach numbers.

No attempt is made here to obtain results which are uniform in the coefficients of F. This problem would be of some interest, expecially in the case $k = 1$.

Theorem 1 is obtained by an extension of the techniques used in [7] and hence we shall refer to [7] at several places in the proof, to avoid merely repeating the arguments there. We also note that the technique in [7] of localizing the primes involved can be used in this paper too. This implies that the second statement of Corollary 1 remains valid for the shorter interval with $N^{7/(36k)+\varepsilon} \le H \le N^{1/k}$. Moreover, we can get stronger results under the assumption of the Generalized Riemann Hypothesis, by techniques similar to those in [3]. In particular we can obtain rather good uniformity in the coefficients in the case $k = 1$.

Theorem 1 deals with a short interval mean square estimate of the error term for the number of Goldbach representations of $F(n)$. By similar, but simpler, techniques we can also obtain the asymptotic formula for the average of $R(F(n))$ over shorter intervals. Writing $F(x) = a_k x^k + \cdots + a_0$ with $a_k > 0$,

$$C(F) = a_k \frac{\varrho_F(2)}{2} \mathfrak{S} \prod_{p>2} \Big(1 + \frac{\varrho_F(p)}{p(p-2)}\Big)$$

and

$$\mathfrak{S} = 2 \prod_{p>2} \Big(1 - \frac{1}{(p-1)^2}\Big),$$

we have

Theorem 2. *Let $k \in \mathbb{N}$, $A, \varepsilon > 0$ and $N^{1/(6k)+\varepsilon} \le H \le N^{1/k-\varepsilon}$. Then*

$$\sum_{N^{1/k} \le n \le N^{1/k}+H} R(F(n)) = C(F)\, H\, N + O_{A,\varepsilon,F}(H\, N\, L^{-A}).$$

An easy consequence of Theorem 2 is

Corollary 2. *Let $k \in \mathbb{N}$, $\varepsilon > 0$, $N^{1/(6k)+\varepsilon} \le H \le N^{1/k-\varepsilon}$ and $\varrho_F(2) \ne 0$. Then there exists $n \in [N^{1/k}, N^{1/k} + H]$ such that $F(n)$ is a Goldbach number.*

The above remarks concerning uniformity in the coefficients, localization of the prime summands and conditional results also apply, in an appropriate form, to Theorem 2 and Corollary 2. We finally note that the N^{ε} in the above results may be replaced by a suitable power of L.

We wish to thank the referee for having pointed out several inaccuracies in the paper.

2. Proof of Theorem 1.

We first note that we may assume that $H = N^{1/(3k)+\varepsilon}$, $\varepsilon > 0$ is sufficiently small, $A > 0$ is a sufficiently large and $N \geq N_0(A,\varepsilon)$, a large constant.

Let $P = L^B$, where $B > 0$ is a suitable constant which will be chosen later on in terms of A and F, $\overline{Q} = H^k L^{-B/4}$ and $Q = \overline{Q}^{1/2}/2$. Denote by $\mathfrak{M}(q,a)$ and $\overline{\mathfrak{M}}(q,a)$ the Farey arc with centre at a/q of the Farey dissections of order Q and $\overline{Q}$ respectively, and let

$$\mathfrak{M}'(q,a) = \left\{\alpha : \ \left|\alpha - \frac{a}{q}\right| \leq \frac{L^B}{N}\right\},$$

$$\mathfrak{M} = \bigcup_{q \leq P} \ \mathop{{\bigcup}^*}_{a=1}^{q} \mathfrak{M}'(q,a), \qquad \overline{\mathfrak{M}} = \bigcup_{q \leq L^{B/4}} \ \mathop{{\bigcup}^*}_{a=1}^{q} \overline{\mathfrak{M}}(q,a),$$

$$\mathfrak{m} = [0,1] \setminus \mathfrak{M} (\mathrm{mod}\ 1) \qquad \text{and} \qquad \overline{\mathfrak{m}} = [0,1] \setminus \overline{\mathfrak{M}} (\mathrm{mod}\ 1),$$

where * means that $(a,q) = 1$.

Writing

$$S(\alpha) = \sum_{n \leq c_1 N} \Lambda(n)\, e(n\alpha)$$

and

$$K(F,\alpha) = \sum_{N^{1/k} \leq n \leq N^{1/k}+H} e(F(n)\,\alpha),$$

where $c_1 = c_1(F) > 0$ is a suitable constant, we have

$$\begin{aligned} &\sum_{N^{1/k} \leq n \leq N^{1/k}+H} |R(F(n)) - F(n)\,\mathfrak{S}(F(n))|^2 \\ &\ll \sum_{N^{1/k} \leq n \leq N^{1/k}+H} \left| \int_{\mathfrak{M}} S(\alpha)^2\, e(-F(n)\,\alpha)\, d\alpha - F(n)\,\mathfrak{S}(F(n)) \right|^2 \\ &\qquad + \sum_{N^{1/k} \leq n \leq N^{1/k}+H} \left| \int_{\mathfrak{m}} S(\alpha)^2\, e(-F(n)\,\alpha)\, d\alpha \right|^2 \\ &:= \sum_{\mathfrak{M}} + \sum_{\mathfrak{m}}, \end{aligned} \tag{1}$$

say.

The quantity $\sum_{\mathfrak{M}}$ can be estimated by standard methods. Using the arguments of Vaughan [10, Chapter 3] we obtain that

$$\begin{aligned} &\int_{\mathfrak{M}} S(\alpha)^2\, e(-F(n)\,\alpha)\,d\alpha - F(n)\,\mathfrak{S}(F(n)) \\ &\qquad \ll N\Big|\sum_{q>P} \frac{\mu(q)^2}{\varphi(q)^2}\, c_q(-F(n))\Big| + N\,L^{-B+1} + N\,L^{-A/2}\,, \end{aligned} \tag{2}$$

where $c_q(m)$ is the Ramanujan sum. By the well known formula

$$c_q(m) = \varphi(q)\,\frac{\mu(q/(q,m))}{\varphi(q/(q,m))}$$

we have that

$$\begin{aligned} \sum_{q>P} \frac{\mu(q)^2}{\varphi(q)^2}\, c_q(-F(n)) &\ll \sum_{q>P} \varphi(q)^{-1}\,\varphi\Big(\frac{q}{(q,F(n))}\Big)^{-1} \\ &\ll \sum_{d|F(n)} \varphi(d)^{-1} \sum_{r>P/d} \varphi(r)^{-2} \\ &\ll P^{-1} \sum_{d|F(n)} \frac{d}{\varphi(d)} \\ &\ll \frac{F(n)\,\tau(F(n))}{P\,\varphi(F(n))}\,, \end{aligned} \tag{3}$$

where τ is the divisor function. Hence from (2), (3) and the Theorem of Nair [6] we get

$$\sum_{\mathfrak{M}} \ll H\,N^2\,L^{-2B+c_2} + H\,N^2\,L^{-A}\,, \tag{4}$$

where $c_2 = c_2(F) > 0$ is suitable constant.

From Parseval's identity we have that

$$\begin{aligned} \sum_{\mathfrak{m}} &= \int_{\mathfrak{m}} S(\xi)^2 \int_{\mathfrak{m}} \overline{S(\alpha)}^2\, K(F,\alpha-\xi)\,d\alpha\,d\xi \\ &\ll N\,L \sup_{\xi\in\mathfrak{m}} \int_{\mathfrak{m}} |S(\alpha)|^2\,|K(F,\alpha-\xi)|\,d\alpha\,. \end{aligned} \tag{5}$$

We need the following slight variant of Weyl's inequality.

Lemma. *Let $|\alpha - a/q| \le 1/q^2$ and $(a,q)=1$. Then for any $D>0$ we have that*

$$K(F,\alpha) \ll_{F,D} H\left(\frac{1}{q}+\frac{1}{H}+\frac{q}{H^k}+L^{-2D+k^2-2}\right)^{1/K} L^{(D+1)/K},$$

where $K = 2^{k-1}$.

PROOF. Arguing as in Lemma 2.4 of [10] we get

$$|K(F,\alpha)|^K \ll H^{K-k}\Big(H^{k-1} + \sum_{h=1}^{c_3H^{k-1}} \tau_k(h)\Big|\sum_{n\in I} e(n\alpha h)\Big|\Big), \tag{6}$$

where τ_k is the k-th divisor function, $I \subset [N^{1/k}, N^{1/k}+H]$ is a suitable interval and $c_3 = c_3(F) > 0$ is a suitable constant. The contribution of the h with $\tau_k(h) > L^D$ is

$$\ll H^{K-k+1}L^{-D}\sum_{h=1}^{c_3H^{k-1}} \tau_k(h)^2 \ll H^K L^{-D+k^2-1}, \tag{7}$$

by the well known inequality

$$\sum_{n\le x} \tau_k(n)^2 \le x(\log ex)^{k^2-1},$$

see, *e.g.*, [10, p. 120]. The contribution of the h with $\tau_k(h) \le L^D$ is, by [10, Lemma 2.2],

$$\ll H^K\left(\frac{1}{q}+\frac{1}{H}+\frac{q}{H^k}\right)L^{D+1}, \tag{8}$$

and the Lemma follows from (6)-(8).

If $\alpha - \xi \in \overline{\mathfrak{m}}$, choosing $D = B/8 + k^2/2 - 1$, from the Lemma we have that

$$K(F,\alpha-\xi) \ll HL^{-(B-4k^2)/(8K)}, \tag{9}$$

hence from (5) and (9) we get

$$\sum_{\mathfrak{m}} \ll HNL \sup_{\xi\in\mathfrak{m}} \int\limits_{\mathfrak{m}\cap(\xi+\overline{\mathfrak{M}})} |S(\alpha)|^2\, d\alpha + HN^2L^{-(B-4k^2-16K)/(8K)}, \tag{10}$$

where $\xi + \overline{\mathfrak{M}}$ denotes the set $\overline{\mathfrak{M}}$ shifted by ξ.

Since $\overline{\mathfrak{M}}$ is the union of at most $L^{B/2}$ Farey arcs, from (10) we have that

$$\sum_{\mathfrak{m}} \ll HNL^{B/2+1} \sup_{\xi\in\mathfrak{m}} \max_{\substack{\overline{q}\le L^{B/4} \\ (\overline{a},\overline{q})=1}} \int_{\mathfrak{m}\cap(\xi+\overline{\mathfrak{M}}(\overline{q},\overline{a}))} |S(\alpha)|^2\, d\alpha + HN^2L^{-(B-4k^2-16K)/(8K)} . \tag{11}$$

From the definition of $\overline{\mathfrak{M}}(\overline{q},\overline{a})$ we have

$$\max_{\substack{\overline{q}\le L^{B/4} \\ (\overline{a},\overline{q})=1}} |\overline{\mathfrak{M}}(\overline{q},\overline{a})| \le \frac{1}{\overline{Q}} . \tag{12}$$

Since for $a/q \ne a'/q'$ and $q, q' \le Q$ we have

$$\left|\frac{a}{q} - \frac{a'}{q'}\right| \ge \frac{1}{Q^2} = \frac{4}{\overline{Q}} , \tag{13}$$

from (12) and (13) we see that there are at most two punctured arcs $\mathfrak{M}''(q,a)$, where

$$\mathfrak{M}''(q,a) = \begin{cases} \mathfrak{M}(q,a) \setminus \mathfrak{M}'(q,a), & \text{if } q \le P , \\ \mathfrak{M}(q,a) , & \text{if } P < q \le Q , \end{cases}$$

with $q \le Q$ and $(a,q)=1$, which intersect any of the $\xi + \overline{\mathfrak{M}}(\overline{q},\overline{a})$. Hence

$$\sup_{\xi\in\mathfrak{m}} \max_{\substack{\overline{q}\le L^{B/4} \\ (\overline{a},\overline{q})=1}} \int_{\mathfrak{m}\cap(\xi+\overline{\mathfrak{M}}(\overline{q},\overline{a}))} |S(\alpha)|^2\, d\alpha \ll \max_{\substack{q\le Q \\ (a,q)=1}} \int_{\mathfrak{M}''(q,a)} |S(\alpha)|^2\, d\alpha . \tag{14}$$

Writing

$$S\Big(\frac{a}{q} + \eta\Big) = \frac{\mu(q)}{\varphi(q)}\, T(\eta) + R(\eta, q, a) , \tag{15}$$

where

$$T(\eta) = \sum_{n\le c_1 N} e(n\eta) ,$$

$$R(\eta, q, a) = \frac{1}{\varphi(q)} \sum_{\chi} \chi(a)\, \tau(\overline{\chi})\, W(\chi, \eta) + O(L^2),$$

$$W(\chi, \eta) = \sum_{n \le c_1 N} \Lambda(n)\, \chi(n)\, e(n\eta) - \delta_\chi\, T(\eta),$$

$$\delta_\chi = \begin{cases} 1, & \text{if } \chi = \chi_0, \\ 0, & \text{if } \chi \neq \chi_0, \end{cases}$$

and $\tau(\chi)$ is the Gauss sum, we have that

$$\int_{\mathfrak{M}''(q,a)} |S(\alpha)|^2\, d\alpha \ll \frac{1}{\varphi(q)^2} \int_{\xi(q)} |T(\eta)|^2\, d\eta + \int_{-1/(qQ)}^{1/(qQ)} |R(\eta, q, a)|^2\, d\eta, \tag{16}$$

where

$$\xi(q) = \begin{cases} \left(\frac{L^B}{N}, \frac{1}{2}\right), & \text{if } q \le L^B, \\ \left(-\frac{1}{qQ}, \frac{1}{qQ}\right), & \text{if } L^B < q \le Q. \end{cases}$$

Since $T(\eta) \ll \min\{N, 1/\|\eta\|\}$, it is easy to see that

$$\frac{1}{\varphi(q)^2} \int_{\xi(q)} |T(\eta)|^2\, d\eta \ll N L^{-B}. \tag{17}$$

In order to estimate the second integral in the right hand side of (16) we proceed as in [7, Section 5]. We call a character χ good if $L(s, \chi)$ has no zeros in the rectangle

$$1 - \frac{10\,(B/\varepsilon) \log L}{L} \le \sigma \le 1, \qquad |t| \le N, \tag{18}$$

and bad otherwise. By the zero-free region of the Dirichlet L-functions, see Prachar [8, Chapter 8], and Siegel's theorem we have that $L(s, \chi) \neq 0$ in the region

$$\sigma > 1 - \frac{c(\varepsilon')}{\max\{q^{\varepsilon'}, \log^{4/5}(|t| + 1)\}}, \tag{19}$$

where $\varepsilon' > 0$ is arbitrary. Hence the existence of a bad character implies that

$$q \gg_{\varepsilon'} L^{1/(2\varepsilon')}. \tag{20}$$

The density estimate

$$\sum_{\chi(\mathrm{mod}\ q)} N(\sigma, T, \chi) \ll (qT)^{12(1-\sigma)/5} \log^{c_4} qT\,, \tag{21}$$

where $c_4 > 0$ is a suitable constant, see Huxley [2] and Ramachandra [9], implies that the number of bad characters for any modulus $q \le Q$ is

$$\ll L^{25\,B/\varepsilon}\,. \tag{22}$$

Hence from (22) and the estimate $\tau(\chi) \ll q^{1/2}$ we have that

$$\begin{aligned}\int_{-1/(qQ)}^{1/(qQ)} &|R(\eta, q, a)|^2\, d\eta \\ &\ll \frac{q}{\varphi(q)^2} L^{50\,B/\varepsilon} \max_{\chi \text{ bad}} \int_{-1/(qQ)}^{1/(qQ)} |W(\chi, \eta)|^2\, d\eta \\ &\quad + \frac{q}{\varphi(q)} \sum_{\chi \text{ good}} \int_{-1/(qQ)}^{1/(qQ)} |W(\chi, \eta)|^2\, d\eta + \frac{L^4}{qQ}\,.\end{aligned} \tag{23}$$

Choosing $\varepsilon' = \varepsilon/(200\,B)$, from (23) and the Parseval identity we get

$$\begin{aligned}\int_{-1/(qQ)}^{1/(qQ)} &|R(\eta, q, a)|^2\, d\eta \\ &\ll \frac{q}{\varphi(q)} \sum_{\chi \text{ good}} \int_{-1/(qQ)}^{1/(qQ)} |W(\chi, \eta)|^2\, d\eta + NL^{-B}\,.\end{aligned} \tag{24}$$

Now we argue as for [7, (21)-(26)], thus getting from (21) and (24) that

$$\begin{aligned}\int_{-1/(qQ)}^{1/(qQ)} &|R(\eta, q, a)|^2\, d\eta \\ &\ll L^{c_5} \sup_{1/2 \le \sigma \le 1 - \frac{10\frac{B}{\varepsilon}\log L}{L}} N^{2\sigma-1}\, N^{12(5/6-k\varepsilon/2)(1-\sigma)/5} + N\,L^{-B} \\ &\ll NL^{-B}\,,\end{aligned} \tag{25}$$

where $c_5 > 0$ is a suitable constant.

Theorem 1 follows now from (1), (4), (11), (14)-(17) and (25), choosing $B = B(A, F) > 0$ suitably large.

3. Proof of Theorem 2.

We give only a brief sketch of the proof of Theorem 2, since the method is a simpler version of the one we have already used in Theorem 1.

Choose $P = L^B, Q = H^k L^{-B}$ and let B, $\mathfrak{M}(q,a)$, $S(\alpha)$ and $K(F,\alpha)$ be as defined at the beginning of the proof of Theorem 1. Write

$$\mathfrak{M} = \bigcup_{q\le P} \bigcup_{a=1}^{q}{}^{*} \mathfrak{M}(q,a) \qquad \text{and} \qquad \mathfrak{m} = \left[\frac{1}{Q}, 1+\frac{1}{Q}\right] \setminus \mathfrak{M}.$$

Then, by the Parseval identity, from the Lemma with $D = (B+k^2)/2-1$ we obtain that

$$\begin{aligned}(26)\qquad &\sum_{N^{1/k}\le n\le N^{1/k}+H} R(F(n)) \\ &= \int_{\mathfrak{M}} S(\alpha)^2 K(F,-\alpha)\,d\alpha + O(NL \sup_{\alpha\in\mathfrak{m}} |K(F,\alpha)|) \\ &= \int_{\mathfrak{M}} S(\alpha)^2 K(F,-\alpha)\,d\alpha + O(HNL^{-(B-k^2-2K)/(2K)}),\end{aligned}$$

where $K = 2^{k-1}$.

By (15), the Cauchy-Schwarz inequality and the estimate $T(\eta) \ll \min\{N, 1/|\eta|\}$ for $|\eta| \le 1/2$ we have that

$$\begin{aligned}&\int_{\mathfrak{M}} S(\alpha)^2 K(F,-\alpha)\,d\alpha \\ (27)\qquad &= \sum_{N^{1/k}\le n\le N^{1/k}+H} F(n) \sum_{q\le P} \frac{\mu(q)^2}{\varphi(q)^2} c_q(-F(n)) \\ &\qquad + O(H(\textstyle\sum(P) + (NL\sum(P))^{1/2})) + O(HNL^{-A}),\end{aligned}$$

where

$$\textstyle\sum(P) = \displaystyle\sum_{q\le P} \sum_{a=1}^{q}{}^{*} \int_{-1/(qQ)}^{1/(qQ)} |R(\eta,q,a)|^2\,d\eta\,.$$

Arguing as for (3) and (4) we obtain that

$$\begin{aligned}\sum_{N^{1/k}\le n\le N^{1/k}+H} F(n)\sum_{q\le P}\frac{\mu(q)^2}{\varphi(q)^2}\,c_q(-F(n))\\ = \sum_{N^{1/k}\le n\le N^{1/k}+H} F(n)\,\mathfrak{S}(F(n)) + O(HNL^{-B+c_6}),\end{aligned} \tag{28}$$

where $c_6 = c_6(F) > 0$ is a suitable constant.

Since $H \le N^{1/k-\varepsilon}$ and $\mathfrak{S}(F(n)) \ll L$, we have that

$$\sum_{N^{1/k}\le n\le N^{1/k}+H} F(n)\,\mathfrak{S}(F(n)) = a_k N \sum(\mathfrak{S}) + O(HNL^{-A}), \tag{29}$$

where

$$\sum(\mathfrak{S}) = \sum_{N^{1/k}\le n\le N^{1/k}+H} \mathfrak{S}(F(n)),$$

hence from (26)-(29) we get

$$\begin{aligned}\sum_{N^{1/k}\le n\le N^{1/k}+H} R(F(n)) = a_k N \sum(\mathfrak{S})\\ + O(H(\textstyle\sum(P) + (NL\sum(P))^{1/2}))\\ + O(HNL^{-A})\end{aligned} \tag{30}$$

provided $B > 0$ is sufficiently large in terms of A and F.

By the orthogonality of the characters we have that

$$\sum(P) \ll \sum_{q\le P}\frac{q}{\varphi(q)}\sum_{\chi}\int_{-1/(qQ)}^{1/(qQ)} |W(\chi,\eta)|^2\,d\eta + \frac{L^4}{Q}\,. \tag{31}$$

Now we proceed as at the end of the proof of Theorem 1. Since in (29) we have $q \le L^B$, we can use the zero-free region (19) and the density estimate (21) in order to bound the sum over χ, and then we sum trivially over q. In this way we obtain that

$$\sum(P) \ll NL^{-2A-1}, \tag{32}$$

provided $Q \ge N^{1/6+\varepsilon}$, which is satisfied by our choice of H.

In order to treat $\sum(\mathfrak{S})$ we define

$$f(n) = \begin{cases} 1, & \text{if } n=1, \\ 0, & \text{if } 2\mid n, \\ \mu(n)^2 \prod_{p\mid n} \left(\frac{1}{p-2}\right), & \text{if } 2\nmid n, \end{cases}$$

hence f is multiplicative and for $2|F(n)$

$$\mathfrak{S}(F(n)) = \mathfrak{S} \sum_{d|F(n)} f(d).$$

Choosing $c_7 = c_7(F) > 0$ a suitable constant, we thus obtain that

$$\begin{aligned} \sum(\mathfrak{S}) &= \mathfrak{S} \sum_{\substack{N^{1/k}\le n\le N^{1/k}+H \\ 2|F(n)}} \Bigg(\sum_{\substack{d|F(n) \\ (d,2)=1}} f(d)\Bigg) \\ &= \mathfrak{S} \sum_{\substack{d\le c_7 N \\ (d,2)=1}} f(d) \Bigg(\sum_{\substack{N^{1/k}\le n\le N^{1/k}+H \\ F(n)\equiv 0 \,(\mathrm{mod}\ 2d)}} 1\Bigg). \end{aligned} \tag{33}$$

It is easy to see that

$$\sum_{\substack{N^{1/k}\le n\le N^{1/k}+H \\ F(n)\equiv 0 \,(\mathrm{mod}\ 2d)}} 1 = H\,\frac{\varrho_F(2d)}{2d} + O(\varrho_F(2d)), \tag{34}$$

see, *e.g.*, Halberstam-Richert [1]. Moreover, it is well known that ϱ_F is multiplicative and satisfies

$$\varrho_F(m) \ll m^{\varepsilon} \tag{35}$$

for every $\varepsilon > 0$. This follows from Nagell [5, Theorem 54] if F is primitive and its discriminant is different from 0, and the general case is an easy consequence of this special case.

Since

$$f(d) \ll \frac{\log^2 d}{d}, \tag{36}$$

from (33)-(36) we get

$$
\begin{aligned}
\sum(\mathfrak{S}) &= \frac{\varrho_F(2)}{2} H \mathfrak{S} \sum_{\substack{d=1 \\ (d,2)=1}}^{\infty} \frac{f(d)\,\varrho_F(d)}{d} \\
&\quad + O\Big(\sum_{d \le c_7 N} f(d)\,\varrho_F(d) \Big) + O\Big(H \sum_{d > c_7 N} \frac{f(d)\,\varrho_F(d)}{d} \Big) \\
&= \frac{\varrho_F(2)}{2} H \mathfrak{S} \prod_{p>2} \Big(1 + \frac{\varrho_F(p)}{p(p-2)} \Big) + O(N^{\varepsilon}) .
\end{aligned} \tag{37}
$$

Theorem 2 follows now from (30), (32) and (37).

References.

[1] Halberstam, H., Richert, H.-E., *Sieve Methods.* Academic Press, 1974.

[2] Huxley, M. N., Large values of Dirichlet polynomials, III. *Acta Arithm.* **26** (1975), 435-444.

[3] Kaczorowski, J., Perelli, A., Pintz, J., A note on the exceptional set for Goldbach's problem in short intervals. *Mh. für Math.* **116** (1993), 275-282; corrigendum **119** (1995), 215-216.

[4] Mikawa, H., On the exceptional set in Goldbach's problem. *Tsukuba J. Math.* **16** (1992), 513-543.

[5] Nagell, T., *Introduction to Number Theory.* Chelsea Publ., 1964.

[6] Nair, M., Multiplicative functions of polynomial values in short intervals. *Acta Arithm.* **62** (1992), 257-269.

[7] Perelli, A., Pintz, J., On the exceptional set for Goldbach's problem in short intervals. *J. London Math. Soc.* **47** (1993), 41-49.

[8] Prachar, K., *Primzahlverteilung.* Springer Verlag, 1957.

[9] Ramachandra, K., On the number of Goldbach numbers in short intervals. *J. Indian Math. Soc.* **37** (1973), 157-170.

[10] Vaughan, R. C., *The Hardy-Littlewood Method.* Cambridge Univ. Press, 1981.

Recibido: 8 de junio de 1.995

Alberto Perelli*
Dipartimento di Matematica
Universitá di Genova
Via Dodecaneso 35
16146 Genova, ITALY
perelli@dima.unige.it

* Research partially supported by the EEC grant CIPA-CT92-4022 (DG 12 HSMU).

Revista Matemática Iberoamericana
Vol. 12, N.° 2, 1996

Existence and properties of the Green function for a class of degenerate parabolic equations

José C. Fernandes and Bruno Franchi

1. Introduction.

It is known that degenerate parabolic equation exhibit somehow different phenomena when we compare them with their elliptic counterpart: see *e.g.* [CS1], [CS2], [CS3], [GW3], [GW4], [Fe2]. Thus, the problem of existence and properties of the Green function for degenerate parabolic boundary value problems is not completely solved, even after the contributions of [GN] and [GW4], in the sense that the existence problem is still open, even if the a priori estimates proved in [GN] will be crucial in our approach. Roughly speaking, we will consider the following Dirichlet problem

$$(\mathrm{P}_0) \qquad \begin{cases} \partial_t u - \sum_{i,j} \partial_i (a_{ij}\, \partial_j u) = g + \operatorname{div} \vec{f}, & \text{in } \Omega \times (0,T]\,, \\ u = 0\,, \quad \text{on the parabolic boundary of } \Omega \times (0,T]\,, \end{cases}$$

where $a_{ij} = a_{ji}$ are measurable functions such that

$$(1.1) \qquad \nu\, \omega(x)\, |\xi|^2 \le \sum_{i,j} a_{ij}(x,t)\, \xi_i \xi_j \le \frac{1}{\nu}\, \omega(x)\, |\xi|^2\,,$$

almost everywhere in $\Omega\times(0,T]$. Here $\Omega\subset\mathbb{R}^n$ is a bounded domain, the parabolic boundary of $\Omega\times(0,T]$ is the set $\Omega\times\{0\}\cup\partial\Omega\times(0,T)$, $t>0$, $\nu\in(0,1)$, and ω is a weight function belonging to the Muckenhoupt class $A_{1+2/n}$ (see below for precise definitions). We will prove that the problem (P_0) admits a unique Green function $\gamma(x,t,\xi,\tau)$ belonging to the natural function space associated with the differential operator in (P_0). Moreover the Green function γ satisfies the classical properties concerning the adjoint equations and the representation formula for the solution of general Cauchy problems.

Our technique is inspired by Aronson's paper [A], where an analogous result is proved for non degenerate parabolic equations by approaching them with a sequence of parabolic equations with smooth coefficients. In our case, we approximate our degenerate equation by means of a sequence of non degenerate problems, for which Aronson's results are true and precise a priori estimates in terms of our weight function have already been proved in [CS1], [CS2], [CS3], [GN]. To this end, in Section 2, we prove a general approximation theorem for A_p weights ($p\geq 1$) by means of weights which are bounded away from 0 and infinity and whose "A_p-constants" depend only on the "A_p-constant" of ω (see Lemma 2.1). In Section 3, existence and properties of weak solutions of the Cauchy problem for the operator in (P_0) are proved. The crucial point consists of showing that a weak limit (in a suitable function space) of a sequence of solutions of approximate problems is in fact a solution of the original problem (see Theorem 3.14). Finally, in Section 4, our main existence and properties results for the Green functions are proved (Theorems 4.1 and 4.2)

1.1. Notations, definitions and basic inequalities.

We indicate by $(x,t)=(x_1,x_2,\dots,x_n,t)$ the points of $\mathbb{R}^{n+1}$, $n\geq 2$, $B=B(\xi,r)$ the usual Euclidean ball with center in ξ and radius r and by $\langle\cdot,\cdot\rangle$ the usual inner product in $\mathbb{R}^n$. The symbols ∂_i, ∂_t and ∇ indicate the derivatives $\partial/\partial x_i$, $\partial/\partial t$ and the gradient $(\partial/\partial x_1,\dots,\partial/\partial x_n)$. If $f=(f_1,\dots,f_n)$ then $\operatorname{div} f=\sum_{i=1}^n\partial_i f_i$. If E is a measurable set in $\mathbb{R}^n$, we indicate by $|E|$ its Lebesgue measure and if $\omega:\mathbb{R}^n\to\mathbb{R}$ is a non negative, locally integrable function we put $\omega(E)=\int_E\omega(x)\,dx$.

For any $p\geq 1$, $L^p(E)$ is the usual Lebesgue space; $L^p_\omega(E)$ is the Lebesgue space with respect to the measure $\omega(x)\,dx$. If X is a normed space and $f\in X$, $\|f\,;X\|$ indicates the norm of f in X.

We recall that given $1 < p < \infty$, a locally integrable non negative function ω is called an A_p weight if there is a constant $c > 0$ such that for all cubes K in $\mathbb{R}^n$

$$(1.2)\qquad \Big(\frac{1}{|K|}\int_K \omega(x)\,dx\Big)\Big(\frac{1}{|K|}\int_K \omega(x)^{-1/(p-1)}\,dx\Big)^{p-1} \le c\,.$$

The infimum of the set of $c > 0$ such that (1.2) holds will be called the A_p constant of ω. We list now some basic properties of A_p weights (see [GC/RF]).

(1.3) Every A_p weight is doubling, *i.e.* $\omega(2K) \le c\,\omega(K)$ for any cube K in $\mathbb{R}^n$, with the constant c independent of the cube K. Here $2K$ denotes the cube with the same center as K and having twice the side length of K.

(1.4) If $\omega \in A_p$ with A_p constant c_p then there is $p_0 = p_0(n,p,c_p)$, $p_0 < p$ such that if $q \in (p_0, p)$ then $\omega \in A_q$ with A_q constant depending on q, c_p, n and p.

If $p = 1$, we say that a locally integrable non negative function w is a A_1 weight if there exists a constant $c > 0$, such that for all cubes K in $\mathbb{R}^n$,

$$\frac{1}{|K|}\int w(x)\,dx \le c\ \operatorname*{ess\,inf}_K w\,.$$

We now recall results proved in [CS2], namely Sobolev inequality and Sobolev interpolation inequality, that will be used throughout the paper.

Sobolev inequality. *Assume ω is an A_p weight in $\mathbb{R}^n$ for $p \le 2$, and let c_p be its A_p constant. Assume $u : \mathbb{R}^n \to \mathbb{R}$ is a Lipschitz continuous function supported in $B = B(\xi, r)$. Then there exists $c > 0$ depending only on n, p and c_p such that*

$$(1.5)\qquad \Big(\frac{1}{\omega(B)}\int_B |u|^{2k}\,\omega\,dx\Big)^{1/2k} \le c\,r\Big(\frac{1}{\omega(B)}\int_B |\nabla u|^2\,\omega\,dx\Big)^{1/2},$$

for $1 \le k \le n/(n-2/p)$ (see [CS2, Theorem 2.7]).

Sobolev interpolation inequality 1. *Assume ω is a A_2 weight. Then for any cylinder $Q = B(\xi, r) \times I \subset \mathbb{R}^{n+1}$ and for any Lipschitz* *continuous function $f : Q \to \mathbb{R}$ which is compactly supported in $B(\xi, r)$ for any fixed t, we have*

$$
\left(\frac{1}{\omega(Q)} \iint_Q |f|^{2h} \omega \, dx \, dt\right)^{1/h} \le c\left(\sup_I \frac{1}{|B|} \int_B |f|^2 \, dx + \frac{r^2}{\omega(Q)} \iint_Q |\nabla f|^2 \omega \, dx \, dt\right), \tag{1.6}
$$

where $h > 1$ and $c > 0$ depend only on n and c_2.

This inequality was proved in [CS2, Lemma 2.8].

Sobolev interpolation inequality 2. *Assume ω is a $A_{1+2/n}$ weight. Then for any cylinder $Q = B(\xi, r) \times I \subset \mathbb{R}^{n+1}$ and for any Lipschitz continuous function $f : Q \to \mathbb{R}$ which is compactly supported in $B(\xi, r)$ for any fixed t, we have*

$$
\left(\frac{1}{|Q|} \iint_Q |f|^{2h'} \omega \, dx \, dt\right)^{1/h'} \le c\left(\sup_I \frac{1}{|B|} \int_B |f|^2 \, dx + \frac{r^2}{\omega(Q)} \iint_Q |\nabla f|^2 \omega \, dx \, dt\right), \tag{1.7}
$$

where $h' > 1$ and $c > 0$ depend only on n and $c_{1+2/n}$.

This inequality was proved in a more general context in [CS2, Lemma 2.9].

The next remark points out the choice of h' that will be important in the future (see Lemmas 3.7 and 3.11).

REMARK 1.8. If we look at the proof of [CS2, Lemma 2.9] we see that h' is chosen in the following way: let $q > n/2$ be such that $\omega \in A_{1+1/q}$ (see property (1.4)). Then inequality (1.5) holds for $k = n(q+1)/(n(q+1) - 2q)$. Put $\beta = 1/2k$, $\alpha = 1 - \beta(q+1)/q$; then inequality (1.7) holds for $h' = k\beta + \alpha = 1 + (2/n - 1/q)/2$.

2. Approximation of A_p weights.

The first step is to approximate the equation (P_0) by non-degenerate parabolic equations. To this end we have to prove the following lemma.

Lemma 2.1. *Let α, $\beta > 1$ be given and let w belong to some A_p class, $p \geq 1$, with A_p constant $c(w,p)$ and let $a_{ij} = a_{ji}$ be measurable functions satisfying*

$$\nu\, w(x)\, |\xi|^2 \leq \sum_{i,j=1}^{n} a_{ij}(x,t)\, \xi_i \xi_j \leq \frac{1}{\nu}\, w(x)\, |\xi|^2 ,$$

for all $\xi \in \mathbb{R}^n$ and almost every $(x,t) \in \Omega \times (a,b)$. Then there exist $w_{\alpha\beta} \geq 0$ and measurable functions $a_{ij}^{\alpha\beta}(x,t)$ such that

i) *$c_1\, \beta^{-1} \leq w_{\alpha\beta} \leq c_2\, \alpha$ in Ω where c_1, c_2 depend only on w and Ω.*

ii) *$\tilde{w}_1 \leq w_{\alpha\beta} \leq \tilde{w}_2$, where $\tilde{w}_i$ is a fixed A_p weight and $c(\tilde{w}_i, p)$ depends only on $c(w,p)$, for $i = 1, 2$.*

iii) *$w_{\alpha\beta} \in A_p$ with $c(w_{\alpha\beta}, p)$ depending only on $c(w,p)$ uniformly on α and β.*

iv) *there exists a closed set $F_{\alpha\beta}$ such that $a_{ij}^{\alpha\beta} = a_{ij}$ in $F_{\alpha\beta}$, $w_{\alpha\beta} = w$ in $F_{\alpha\beta}$ and $w_{\alpha\beta} \sim \tilde{w}_1 \sim \tilde{w}_2$ in $F_{\alpha\beta}$ with equivalence constants depending on α and β (i.e. $c_{\alpha\beta} \leq w_{\alpha\beta}/\tilde{w}_i \leq C_{\alpha\beta}$ for some positive constants $c_{\alpha\beta}$ and $C_{\alpha\beta}$ and for $i = 1, 2$). Moreover, $F_{\alpha\beta} \subset F_{\alpha'\beta'}$ if $\alpha \leq \alpha'$, $\beta \leq \beta'$ and the complement of $\cup_{\alpha,\beta \geq 1} F_{\alpha\beta}$ has zero measure.*

v) *$w_{\alpha\beta} \longrightarrow w$ almost everywhere in $\mathbb{R}^n$ as α, β tend to infinity.*

vi)

$$\nu\, w_{\alpha\beta}(x)\, |\xi|^2 \leq \sum_{i,j=1}^{n} a_{ij}^{\alpha\beta}(x,t)\, \xi_i \xi_j \leq \frac{1}{\nu}\, w_{\alpha,\beta}(x)\, |\xi|^2 ,$$

for any $\xi \in \mathbb{R}^n$ and almost every $(x,t) \in \Omega \times (a,b)$.

PROOF. Suppose first $w \in A_1$. Since we are interested to approximate in Ω, we may assume, without loss of generality, that $w \in L^1(\mathbb{R}^n)$. Then for each $\alpha > 1$, by Calderon-Zygmund decomposition, there exists a family of non-overlapping cubes $\{Q_j^\alpha\}$ consisting of those maximal dyadic cubes over which the average of w is greater than α. If we put

a) $\cup_{j=1}^\infty Q_j^\alpha = U_\alpha^+$, then

b) $\alpha < \dfrac{1}{|Q_j^\alpha|}\displaystyle\int_{Q_j^\alpha} w(x)\,dx \le 2^n\,\alpha$, for any j;

c) $w(x) \le \alpha$ for any $x \in F_\alpha^+ = (U_\alpha^+)'$;

d) $|U_\alpha^+| \le \dfrac{1}{\alpha}\displaystyle\int_{\mathbb{R}^n} w(x)\,dx$.

Moreover, if $M(w)$ denotes the usual Hardy-Littlewood maximal function, then

$$\{x \in \mathbb{R}^n :\ M(w)(x) > 4^n\,\alpha\} \subset \bigcup_j 3Q_j^\alpha$$

(see, for instance, [GC/RF, Chapter 2, Theorem 1.12]).

We explicitly note that, if $\alpha < \beta$, then $U_\beta^+ \subset U_\alpha^+$. In fact, let x belong to U_β^+; then there exists a (unique) dyadic cube $Q_{j_0}^\beta$ containing x such that b) holds. Let now $\mathcal{I}$ be the set of all indices $j \in \mathbb{N}$ such that $Q_j^\alpha \cap Q_{j_0}^\beta \ne \varnothing$. Note that $\mathcal{I} \ne \varnothing$, since otherwise we would have $Q_{j_0}^\beta \subset F_\alpha^+$ and hence, by c),

$$\beta < \frac{1}{|Q_{j_0}^\beta|}\int_{Q_{j_0}^\beta} w\,dx \le \alpha < \beta\,,$$

a contradiction. Since we are dealing with dyadic cubes, either $\mathcal{I} = \{j_1\}$ and $Q_{j_0}^\beta \subset Q_{j_1}^\alpha$ or $Q_j^\alpha \subset Q_{j_0}^\beta$ for any $j \in \mathcal{I}$. In the first case we are done. Let us show that the second case cannot occur. Indeed, for any j, Q_j^α is a maximal dyadic cube over which the average w is greater than α; on the other hand $Q_{j_0}^\beta$ is a dyadic cube and the average of w over it is greater than $\beta > \alpha$. Thus the assertion is completly proved.

Define

$$w_\alpha(x) = \sum_{k=1}^\infty \frac{1}{|Q_k^\alpha|}\int_{Q_k^\alpha} w(y)\,dy\ \chi_{Q_k^\alpha}(x) + w(x)\chi_{F_\alpha^+}(x)\,.$$

We will show that

I) $w_\alpha \in A_1$ and $c(w_\alpha, 1)$ depends only on $c(w,1)$.

II) $w_\alpha \to w$ almost everywhere in $\mathbb{R}^n$ as $\alpha \to \infty$.

III) $\min\{1, w\} \in A_1$ and $c(\min\{1, w\}, 1)$ depends only on $c(w, 1)$; moreover $\min\{w, 1\} \le w_\alpha \le c\,w$, where c depends only on $c(w, 1)$.

IV) If Q_0 is a fixed cube containing Ω, then

$$2^n\,\alpha \ge w_\alpha(x) \ge c = \frac{1}{|Q_0|}\int_{Q_0} \min\{w, 1\}(y)\,dy\,, \qquad x \in \Omega\,.$$

In order to prove the above statements, first of all we note that, if $v \in L^1$ is a weight function such that for any dyadic cube Q we have $v(3Q) \le c_0\,v(Q)$, then we can restrict ourselves to test the A_1 condition only on dyadic cubes. In fact, denoting by $c^\star(v, 1)$ the "A_1-constant for dyadic cubes", if $f \in L^1_v$ and $\{C^t_j\}$ is the Calderon-Zygmund decomposition for f, we have

$$\begin{aligned} v(\{x:\ M(f)(x) > 4^n t\}) &\le \sum_j v(3C^t_j) \\ &\le c_0 \sum_j v(C^t_j) \\ &\le c_0\, c^\star(v,1) \sum_j |C^t_j| \inf_{C^t_j} v \\ &\le c_0\, c^\star(v,1)\, \frac{1}{t} \sum_j \int_{C^t_j} |f(x)| \inf_{C^t_j} v\; dx \\ &\le c_0\, c^\star(v,1)\, \frac{1}{t} \int |f(x)|\, v(x)\, dx\,, \end{aligned}$$

and hence $v \in A_1$, by [GC/RF, Chapter 4, Theorem 2.1]. We note explicitly that $c(v, 1)$ depends only on c_0 and $c^\star(v, 1)$.

Thus, let us prove first that

$$w_\alpha(3Q) \le c_0\, w_\alpha(Q)\,,$$

for any dyadic cube Q and for any $\alpha > 1$, where c_0 depends only on n and $c(w, 1)$.

First, let us suppose that $|Q \cap U^+_\alpha| < |Q|/2$. We will prove later, in III), that $w_\alpha \le \max\{1, c(w, 1)\}\, w = c_1\, w$; we have:

$$w_\alpha(3Q) \le c_1\, w(3Q)$$

$$
\begin{aligned}
&= c_1\, 3^n\, |Q|\, \frac{1}{|3Q|} \int_{3Q} w(x)\, dx \\
&\le c_1\, c(w,1)\, 3^n\, |Q|\, \inf_{3Q} w \\
&\le c_1\, c(w,1)\, 3^n\, |Q| \inf_{Q\cap F_\alpha^+} w \\
&= c_1\, c(w,1)\, 3^n\, \frac{|Q|}{|Q\cap F_\alpha^+|}\, |Q\cap F_\alpha^+| \inf_{Q\cap F_\alpha^+} w \\
&\le 2\, c_1\, c(w,1)\, 3^n\, |Q\cap F_\alpha^+| \inf_{Q\cap F_\alpha^+} w \\
&\le 2\, c_1\, c(w,1)\, 3^n \int_{Q\cap F_\alpha^+} w_\alpha(x)\, dx
\end{aligned}
$$

since $w_\alpha \equiv w$ on F_α^+ and therefore

$$
w_\alpha(3Q) \le 2\, c_1\, c(w,1)\, 3^n \int_Q w_\alpha(x)\, dx = 2\, c_1\, c(w,1)\, 3^n\, w_\alpha(Q)
$$

and hence, in this case we are done.

Suppose now $|Q\cap U_\alpha^+| \ge |Q|/2$; then either $Q \subset Q_{j_0}^\alpha$ for some j_0 (which in turn is unique), or $Q_j^\alpha \subset Q$ for j belonging to a given set $\mathcal{J}$ of indices. By definition of w_α, it follows from b) and c) that $w_\alpha(x) \le 2^n\alpha$ almost everywhere; hence, if $Q \subset Q_{j_0}^\alpha$ we get

$$
\begin{aligned}
\int_{3Q} w_\alpha(x)\, dx &\le |3Q|\, 2^n\, \alpha \\
&= 3^n\, 2^n\, |Q|\, \alpha \\
&< 3^n\, 2^n\, |Q|\, \frac{1}{|Q_{j_0}^\alpha|} \int_{Q_{j_0}^\alpha} w(x)\, dx \\
&= 6^n \int_Q w_\alpha(x)\, dx\,,
\end{aligned}
$$

since $w_\alpha \equiv |Q_{j_0}^\alpha|^{-1} \int_{Q_{j_0}^\alpha} w(x)\, dx$ on Q. Otherwise

$$
\begin{aligned}
w_\alpha(3Q) &\le 3^n\, |Q|\, 2^n\, \alpha \\
&\le 2\, 6^n \int_{Q\cap U_\alpha^+} \alpha\, dx \\
&= 2\, 6^n \sum_{j\in\mathcal{J}} \int_{Q_j^\alpha} \alpha\, dx \\
&\le 2\, 6^n \sum_{j\in\mathcal{J}} \int_{Q_j^\alpha} w_\alpha(x)\, dx
\end{aligned}
$$

where in the last inequality we used b) and therefore we have

$$w_\alpha(3Q) \le 2\,6^n \int_{\cup Q_j^\alpha} w_\alpha(x)\,dx \le 2\,6^n \int_Q w_\alpha(x)\,dx$$

and the assertion is completely proved.

Now we are ready to prove I). Let Q be a fixed dyadic cube, then one of the three cases can happen

I$_1$) $Q \cap Q_j^\alpha = \varnothing$, for all j,

I$_2$) $Q \subset Q_j^\alpha$, for one and only one j,

I$_3$) $Q_j^\alpha \subset Q$, for some index $j \in J$.

In case I$_1$), $Q \subset F_\alpha^+$ and hence $w_\alpha \equiv w$ in Q and we are done. In case I$_2$),

$$w_\alpha(Q) = |Q|\,\frac{1}{|Q_j^\alpha|}\int_{Q_j^\alpha} w(x)\,dx = \inf_{x\in Q} w_\alpha(x)\,|Q|\,,$$

since $w_\alpha \equiv |Q_j^\alpha|^{-1}\int_{Q_j^\alpha} w(x)\,dx$ over Q_j^α. Finally in case I$_3$)

$$\begin{aligned} w_\alpha(Q) &= \sum_{j\in J}\frac{1}{|Q_j^\alpha|}\int_{Q_j^\alpha} w(x)\,dx\,|Q_j^\alpha| + w(Q\cap F_\alpha^+) \\ &= \sum_{j\in J}\int_{Q_j^\alpha\cap Q} w(x)\,dx + w(Q\cap F_\alpha^+) \\ &\le w(Q) \le c(w,1)\,|Q|\,\inf_{y\in Q} w(y)\,. \end{aligned}$$

On the other hand we note that if $y \in U_\alpha^+$, by definition

$$w_\alpha(y) = \frac{1}{|Q_j^\alpha|}\int_{Q_j^\alpha} \omega(x)\,dx > \alpha\,.$$

Thus, if $y \in Q \cap U_\alpha^+$ and Q_k^α is any cube contained in Q we have

$$\inf_Q w \le \inf_{Q_k^\alpha} w \le \frac{1}{|Q_k^\alpha|}\int_{Q_k^\alpha} w(z)\,dz \le 2^n\,\alpha < 2^n\,w_\alpha(y)\,.$$

In addition, if $y \in Q\cap F_\alpha^+$ then $w_\alpha(y) = w(y) \ge \inf_Q w$ so that $\inf_Q w \le 2^n \inf_Q w_\alpha$ and hence I) is completely proved.

To prove II), we note that $w_\alpha \equiv w$ in F_α^+ and that F_α^+ increases as α tends to infinity. Moreover, $|\cap (F_\alpha^+)'| = 0$.

Finally, to show III) we know that, for any cube Q, either $\inf_Q w \geq 1$ or $\inf_Q w < 1$. In the first case

$$\frac{1}{|Q|}\int_Q \min\{w,1\}(x)\,dx = \frac{1}{|Q|}\int_Q 1\,dx = 1 \leq \min\{w(y),1\}\,,$$

for any $y \in Q$, whereas if $\inf_Q w \leq 1$ then

$$\frac{1}{|Q|}\int_Q \min\{w,1\}(x)\,dx \leq \frac{1}{|Q|}\int_Q w(x)\,dx \leq c(w,1)\inf_Q w\,.$$

Put $\lambda = \inf_Q w < 1$ and assume by contradiction that $\inf_Q \min\{w,1\} < \lambda$; then there exists $E \subset Q$, $|E| > 0$ such that $\min\{w,1\} < \lambda' < \lambda$ in E and hence, since $\lambda < 1$, $w < \lambda'$ in E, which is a contradiction. Thus we have proved the first part of III). To prove the second part we note that if $x \in F_\alpha$ then $w_\alpha(x) = w(x) \geq \min\{1,w\}(x)$; if $x \in Q_j^\alpha$, for some j, then

$$w_\alpha(x) = \frac{1}{|Q_j^\alpha|}\int_{Q_j^\alpha} w(y)\,dy > \alpha \geq \min\{1,w\}(x)\,.$$

Analogously, if $x \in Q_j^\alpha$ for some j, then

$$w_\alpha(x) \leq c(w,1)\inf_{Q_j^\alpha} w \leq c(w,1)\,w(x)\,.$$

Finally, assertion IV) follows straightforwardly from III) by using a) and c).

Suppose now $w \in A_p$, for $p > 1$. Then by Peter Jones' factorization theorem ([GC/RF, Theorem 5.2 and Corollary 5.3, Chapter 4]) there exist w_0, $w_1 \in A_1$ such that $w = w_0\, w_1^{1-p}$. In addition $c(w_i,1)$ depends only on $c(w,p)$, $i = 0,1$. Choose $\alpha, \beta > 1$ and define

$$w_{\alpha\beta} = (w_0)_\alpha\left((w_1)_{\beta^{1/(p-1)}}\right)^{1-p}.$$

We need to show that $w_{\alpha\beta}$ satisfies properties i)-v). Obviously i) follows from IV); ii) from III) and [GC/RF, Theorem 5.2, Corollary 5.3] with $\tilde{w}_1 = \min\{w_0,1\}w_1^{1-p}$, $\tilde{w}_2 = w_0(\min\{w_1,1\})^{1-p}$; iii) from I) and [GC/RF, Theorem 5.2, Corollary 5.3]; v) from II); for iv), we define $F_{\alpha\beta} = F_\alpha^0 \cap F^1_{\beta^{1/(p-1)}}$, where $F_\alpha^0 = F_\alpha^+$ for the weight w_0 and

$F^1_{\beta^{1/(p-1)}} = (F_{\beta^{1/(p-1)}})^+$ for the weight w_1. By definition, $(w_0)_\alpha \equiv w_0$ and $(w_1)_{\beta^{1/(p-1)}} \equiv w_1$ on $F_{\alpha\beta}$. Hence to prove that $w_{\alpha\beta} \sim \tilde{w}_1$ (for instance) we can replace $w_{\alpha\beta}$ by w. Note now that (with the notation we used above) $\min\{w_0, 1\} \sim w_0$ in F_α. Obviously $\min\{w_0, 1\} \leq w_0$; moreover, if, for some $x \in F_\alpha$, $1 = \min\{w_0, 1\}$, then $w_0(x) \leq \alpha = \alpha \min\{w_0(x), 1\}$. An analogous argument shows that $w_1 \sim \min\{w_1, 1\}$ and hence

$$w_{\alpha\beta} = (w_0)_\alpha \big((w_1)_{\beta^{1/(p-1)}}\big)^{1-p} \sim \min\{w_0, 1\} \,\big((w_1)_{\beta^{1/(p-1)}}\big)^{1-p} = \tilde{w}_1 \,.$$

Finally to prove vi) we define

$$\begin{aligned} a_{ij}^{\alpha\beta}(x,t) = \big((w_1)_{\beta^{1/(p-1)}}\big)^{1-p} & \\ \cdot \Big(\sum_{k=1}^{\infty} \frac{1}{|Q_k^\alpha|} \int_{Q_k^\alpha} a_{ij}(y,t)\, w_1^{p-1}(y)\, dy\, \chi_{Q_k^\alpha}(x) & \\ + a_{ij}(x,t)\, w_1^{p-1}(x)\, \chi_{F_\alpha^+}(x)\Big)\,, & \end{aligned}$$

where $\{Q_k^\alpha\}$ is the Calderon-Zygmund decomposition for w_0. Now, by assumption,

$$a_{ij}^{\alpha\beta}(x,t)\,\xi_i\xi_j \leq \frac{1}{\nu}\,((w_1)_{\beta^{1/(p-1)}})^{1-p}\,(w_0)_\alpha\,|\xi|^2 = \frac{1}{\nu}\,w_{\alpha\beta}(x)\,|\xi|^2 \,.$$

The lower estimate can be carried out in the same way.

REMARK. There is a different approach to the approximation theorem, see [G].

3. The weak solutions.

In this section we prove existence and properties of the weak solutions of the Cauchy-Dirichlet problems for the operator $\partial_t - L$. We will follow [A] and [FJK]. The crucial point is to prove that a weak limit (in a suitable function space) of a sequence of solutions of approximate non degenerate problems is in fact a solution of the original problem. This will be done in Theorem 3.14. We start this section by introducing some normed spaces in order to be able to define what we mean by a weak solution for Cauchy-Dirichlet problems for the operator $\partial_t - L$.

Put $Q = \Omega \times (0,T)$ and let ω belong to A_p for some $p > 1$. We denote by $H^{1,p}_{\omega}(Q)$ the closure of Lipschitz functions under the norm

$$\|u\|^p = \iint_Q |u(x,t)|^p \,\omega(x)\,dx\,dt + \iint_Q |\nabla u(x,t)|^p \,\omega(x)\,dx\,dt$$

and similarly $H^{1,p}_{0,\omega}(Q)$ denotes the closure of Lipschitz functions with compact support in Q under the same norm. We note that, since $\omega \in A_p$, ∇u in the limit sense belongs to L^1_{loc} and it coincides with the distribution gradient of u. Moreover we put

$$H^{-1,p}_{\omega}(Q) = \left\{ \operatorname{div} f : \ \frac{|\vec{f}|}{w} \in L^p_w \right\}.$$

The next theorem characterizes the dual space of $H^{1,p}_{\omega}(Q)$. We will assume that Ω is a smooth regular open subset of $\mathbb{R}^n$. This implies in particular that there exist $\alpha > 0$, $\rho_0 > 0$ such that for any $x_0 \in \partial\Omega$, $\rho < \rho_0$ we have

$$|B(x_0,\rho) \setminus \Omega| \geq \alpha |B(x_0,\rho)|.$$

We note that the smoothness assumption could be strongly relaxed; however we hold it to avoid a number of arguments at some points.

Theorem 3.1. *The space $H^{1,p}_{\omega}(Q)$ is a reflexive Banach space, for $p > 1$. Moreover*

i) $H^{-1,p}_{\omega}(Q) = (H^{1,p'}_{0,\omega}(Q))^{\star}$.

ii) *Let ω_1 be another A_p weight; if $u \in H^{1,p}_{0,\omega_1}(Q)$ with respect to ω_1 and u, $|\nabla u| \in L^p_{\omega}$, then $u \in H^{1,p}_{0,\omega}(Q)$ with respect to ω (see also* [CP/SC])

PROOF. It is easy to see that $H^{1,p}_{\omega}(Q)$ is reflexive since it is isometrical to a closed subspace of the reflexive space $(L^p_{\omega}(Q))^{n+1}$.

Let us now prove ii). First, let $x \in \Omega$ be such that $d(x,\partial\Omega) < r$ for a given (small) $r > 0$, and let $y = y(x) \in \partial\Omega$ be such that $d(x,y) = d(x,\partial\Omega) = d$. If v is a Lipschitz function with compact support in Q, it can be continued by zero outside of $\bar{Q}$ and we have

$$B(y,d) \setminus \Omega \subset \{z \in B(y,2d) : \ v(z,t) = 0\},$$

for any $t \in [0,T]$, so that

$$|\{z \in B(y,2d) : \ v(z,t) = 0\}| \geq \alpha\,|B(y,d)| \geq c\,|B(y,2d)|.$$

Hence, keeping in mind that $x \in B(y, 2d)$, by standard arguments (see, *e.g.*, [KS] or arguing as in [FS, Lemma 4.3]) we obtain for $t \in [0, T]$

$$|v(x,t)| \leq c\, d\, M(|\nabla v(\cdot, t)|\, \chi_{B(y,\theta d)\cap\Omega})(x),$$

where Mf is the usual Hardy-Littlewood maximal function and $\theta > 0$ is an absolute constant. On the other hand, if $z \in B(y, \theta d)$ then $d(z, \partial\Omega) \leq d(z, y) \leq \theta\, d \leq \theta\, r$, so that

$$|v(x,t)| \leq c\, r\, M(|\nabla v(\cdot, t)|\, \chi_{\theta r})(x), \tag{3.2}$$

where $\chi_s(x)$ is the characteristic function of $\{z \in \Omega : \ d(z, \partial\Omega) < s\}$ for $s > 0$. On the other hand the function u is (by assumption) the limit in $H^{1,p}_{\omega_1}(Q)$ (with respect to ω_1) of a sequence $(v_k)_{k\in\mathbb{N}}$ of Lipschitz continuous functions supported in $\bar{Q}$. In particular, $v_k \to u$, as $k \to \infty$, for almost every $(x, t) \in Q$. Moreover (by [Mu])

$$\begin{aligned}
&\|M(|\nabla v_k - \nabla u|\, \chi_{\theta r}); L^p_{\omega_1}(\Omega; L^p([0,T]))\|^p \\
&\quad = \int_0^T \Big(\int_\Omega M(|\nabla v_k - \nabla u|\, \chi_{\theta r})^p(x)\, \omega_1(x)\, dx \Big)\, dt \\
&\quad \leq c \int_0^T \Big(\int_\Omega |\nabla v_k - \nabla u|^p\, \chi_{\theta r}\, \omega_1\, dx \Big)\, dt \\
&\quad \leq c\, \|v_k - u; H^{1,p}_{\omega_1}(Q)\|^p \to 0,
\end{aligned}$$

as $k \to \infty$. Hence

$$M(|\nabla v_k(\,\cdot\,, t) - \nabla u(\,\cdot\,, t)|\, \chi_{\theta r})(x) \to 0,$$

as $k \to \infty$ for almost every $(x, t) \in Q$. By applying (3.2) to $v = v_k$, we get

$$|v_k(x,t)| \leq c\, r\, \big(M(|\nabla u(\,\cdot\,, t)\,|\chi_{\theta r})(x) + M(|\nabla v_k(\,\cdot\,, t) - \nabla u(\,\cdot\,, t)|\, \chi_{\theta r})(x)\big)$$

and hence

$$|u(x,t)| \leq c\, r\, M(|\nabla u(\,\cdot\,, t)|\, \chi_{\theta r})(x), \tag{3.3}$$

for almost everywhere $(x, t) \in \Sigma_r = \{(x,t) \in Q : \ d(x, \partial\Omega) < r\}$.

If $\delta > 0$, let now $\sigma_\delta : [0, \infty) \to [0, 1]$ be a smooth function such that $\sigma_\delta(t) \equiv 0$ if $0 \leq t \leq \delta$, $\sigma_\delta(t) \equiv 1$ if $t \geq 2\delta$ and $|\sigma'_\delta(t)| \leq 2/\delta$ for any $t \geq 0$. Define

$$u = u\, \sigma_\delta(d(\,\cdot\,, \partial\Omega)) + u\, (1 - \sigma_\delta(d(\,\cdot\,, \partial\Omega))) = u_\delta + v_\delta\,.$$

Note that, by [GT, Lemma 14.16], $d(\cdot, \partial\Omega)$ is a smooth function in Ω if δ is small enough. Using our previous notations we have (by (3.3))

$$\begin{aligned} |\nabla v_\delta(x,t)| &\le |\nabla u(x,t)|\,\chi_{2\delta}(x) + \frac{2}{\delta}\,|u(x,t)|\,\chi_{2\delta}(x) \\ &\le |\nabla u(x,t)|\,\chi_{2\delta}(x) + c\,M(|\nabla u(\cdot,t)|\,\chi_{2\theta\delta})(x). \end{aligned}$$

Hence

$$\begin{aligned} \|v_\delta; H^{1,p}_\omega(Q)\|^p &\le \iint_{\Sigma_{2\delta}} \left(|u(x,t)|^p + |\nabla u(x,t)|^p\right)\omega(x)\,dx\,dt \\ &\qquad + c\int_0^T \Big(\int M(|\nabla u(\cdot,t)|\,\chi_{2\theta\delta})^p(x)\,\omega(x)\,dx\Big)\,dt \\ &\le c\iint_{\Sigma_{2\theta\delta}} \left(|u(x,t)|^p + |\nabla u(x,t)|^p\right)\omega(x)\,dx\,dt\,, \end{aligned}$$

by [Mu]. Thus, by the absolute continuity of the integral, we can choose $\delta > 0$ such that

$$\|v_\delta; H^{1,p}_\omega(Q)\| < \varepsilon\,. \tag{3.4}$$

Let now δ be fixed so that (3.4) holds. If r, $\rho > 0$ and p_r, ψ_ρ are usual mollifiers in $\mathbb{R}$ and $\mathbb{R}^n$ respectively, we put

$$u_{\delta,r,\rho} = u_\delta * \psi_\rho * p_r\ ,$$

where we have continued $u_\delta(x,\cdot)$ by zero outside of $[0,T]$. Obviously, $u_{\delta,r,\rho}$ belongs to $C_0^\infty(\Omega\times\mathbb{R})$ if ρ is small enough. Moreover

$$\begin{aligned} \|u_\delta - u_{\delta,r,\rho}; H^{1,p}_\omega(Q)\| &\le \|u_\delta - u_\delta * \psi_\rho; H^{1,p}_\omega(Q)\| \\ &\qquad + \|u_\delta * \psi_\rho - u_\delta * \psi_\rho * p_r; H^{1,p}_\omega(Q)\| \\ &= \mathrm{I}_1 + \mathrm{I}_2 \end{aligned}$$

Now

$$\mathrm{I}_1^p = \int_0^T \|u_\delta - u_\delta * \psi_\rho; H^{1,p}_\omega(\Omega)\|^p\,dt\,.$$

By assumption u_δ, $|\nabla u_\delta|$ belong to $L^p_\omega(\Omega)$ for almost every $t \in [0,T]$ and hence, arguing as in [CP/SC, Proposition 2.5],

$$\|u_\delta - u_\delta * \psi_\rho; H^{1,p}_\omega(\Omega)\| \to 0\,,$$

as $\rho \to 0$ for almost every $t \in [0,T]$. On the other hand

$$\begin{aligned}\|u_\delta * \psi_\rho; H^{1,p}_\omega(\Omega)\| &\le \|u_\delta * \psi_\rho; L^p_\omega(\Omega)\| + \|\,|\nabla u_\delta| * \psi_\rho; L^p_\omega(\Omega)\| \\ &\le \|Mu_\delta(\,\cdot\,,t); L^p_\omega(\Omega)\| + \|M(|\nabla u_\delta(\,\cdot\,,t)|); L^p_\omega(\Omega)\| \\ &\le c\,\|u_\delta(\,\cdot\,,t); H^{1,p}_\omega(\Omega)\|\,,\end{aligned}$$

by [Mu]. Hence we can apply Lebesgue's dominate convergence theorem, since $t \to \|u_\delta(\,\cdot\,,t); H^{1,p}_\omega(\Omega)\|$ belongs to $L^p([0,T])$ and we can conclude that there exists $\rho > 0$ such that $I_1 < \varepsilon$.

Let now ρ be fixed as above; we will now show that there exists $r > 0$ such that $\mathrm{I}_2 < \varepsilon$. We have

$$\begin{aligned}\mathrm{I}_2 = &\int_\Omega \Big(\int_0^T |(u_\delta * \psi_\rho) - (u_\delta * \psi_\rho) * p_r|^p\, dt\Big)\, \omega(x)\, dx \\ &+ \int_\Omega \Big(\int_0^T |(\nabla u_\delta * \psi_\rho) - (\nabla u_\delta * \psi_\rho) * p_r|^p\, dt\Big)\, \omega(x)\, dx\,.\end{aligned}$$

We now denote by $\tilde{u}$ either $u_\delta * \psi_\rho$ or $\nabla u_\delta * \psi_\rho$. By assumption

$$|\tilde{u}(x,t)| \le M\,(|u_\delta(\,\cdot\,,t)| + |\nabla u_\delta(\,\cdot\,,t)|)(x)$$

and hence, again by [Mu], $\tilde{u} \in L^p([0,T]; L^p_\omega(\Omega))$. On the other hand $w(x) \neq 0$ for almost every $x \in \Omega$ since it belongs to A_p and hence

$$\tilde{u}(x,\,\cdot\,) \in L^p([0,T])\,,$$

for almost $x \in \Omega$, so that, by standard results on convolution,

$$\int_0^T |\tilde{u}(x,t) - (\tilde{u}(x,\,\cdot\,) * p_r)(t)|^p\, dt \longrightarrow 0\,,$$

as $r \to 0^+$ for almost everywhere $x \in \Omega$. We can now argue as above by using the maximal function in $\mathbb{R}$ and hence apply the Lebesgue's dominate convergence theorem. Thus we get

$$I_2 < \varepsilon\,, \qquad \text{if } r \text{ is small enough.}$$

Combining these estimates with (3.4) we get

$$\|u - u_{\delta,r,\rho}\,; H^{1,p}_\omega(Q)\| < 3\,\varepsilon\,.$$

Thus we have proved that u can be approximated in $H^{1,p}_{\omega}(Q)$ by functions in $C_0^\infty([0,T]\times\Omega)$. In particular, u belongs to $H^{1,p}_{0,\omega}(Q)$.

Finally, let us prove i). Denote by T the application

$$T : H^{1,p'}_{0,\omega}(Q) \longrightarrow (L^{p'}(Q))^n$$

given by $Tu = w^{1/p'}\nabla u$. Note that $\|Tu\,;(L^{p'}(Q))^n\|$ is equivalent to $\|u\,;H^{1,p'}_{0,\omega}(Q)\|$ because of the Sobolev inequality (1.5) for compactly supported functions, so that the range of T is a closed subspace Y of $(L^{p'}(Q))^n$. Thus, if $F\in(H^{1,p'}_{\omega}(Q))^*$ then the application $\vec{v}\to F(T^{-1}(\vec{v}))$ is a linear continuous functional in Y which can be continued as an element of $((L^{p'}(Q))^n)^* = (L^p(Q))^n$ by Hahn-Banach theorem. Thus there exists $\vec{g}=(g_1,\dots,g_n)\in(L^p(Q))^n$ such that for any $u\in H^{1,p'}_{\omega}(Q)$ we can write

$$\begin{aligned}F(u) &= F(T^{-1}(\omega^{1/p'}\nabla u))\\ &= \sum_j \iint_Q \partial_j u(x,t)\, g_j(x,t)\, \omega^{1/p'}(x)\, dx\, dt\\ &= \iint_Q \langle \nabla u(x,t), \vec{f}(x,t)\rangle\, dx\, dt\end{aligned}$$

where $\vec{f}(x,t) = \vec{g}(x,t)\omega^{1/p'}(x)$. Note that

$$\begin{aligned}\iint_Q \Big(\frac{|\vec{f}(x,t)|}{\omega}\Big)^p \omega\, dx\, dt &= \iint_Q |\vec{g}(x,t)|^p\,(\omega(x))^{p(1/p'-1)}\,\omega(x)\, dx\, dt\\ &= \iint_Q |\vec{g}(x,t)|^p\, dx\, dt < \infty,\end{aligned}$$

and the assertion is proved.

REMARK 3.5. If f_0 is such that $f_0/\omega\in L^p_\omega(Q)$, then f_0 can be identified with a linear functional on $H^{1,p'}_{0,\omega}(Q)$ (with respect to the duality between $\mathcal{D}$ and $\mathcal{D}'$) and hence we can write $f_0 = \operatorname{div}\vec{f}$ for some vector field $\vec{f}$ such that $|\vec{f}|/\omega\in L^p_\omega(Q)$.

If $Q=\Omega\times(0,T]$ is the parabolic cylinder, we will denote by Σ its lateral boundary, *i.e.* $\Sigma=\partial\Omega\times(0,T]$. We write

$$Lu = \sum_{i,j}\frac{\partial}{\partial x_i}\Big(a_{ij}\frac{\partial u}{\partial x_j}\Big).$$

Definition 3.6. *Let $S \in H^{-1,2}(Q)$ and $u_0 \in L^2(\Omega)$ be given. We say that $u \in H_0^{1,2}(Q)$ is a weak solution of the problem*

$$\begin{cases} \partial_t u - Lu = -S\,, \\ u(x,0) = u_0(x)\,, \\ u(x,t) = 0 \quad on\ \Sigma\,. \end{cases}$$

if

$$\iint_Q \left(a_{ij}(x,t)\,\frac{\partial u}{\partial x_j}\,\frac{\partial \varphi}{\partial x_i} - u\,\frac{\partial \varphi}{\partial t}\right) dx\,dt = -\langle S,\varphi\rangle_{\mathcal{D}',\mathcal{D}} + \int_\Omega u_0(x)\,\varphi(x,0)\,dx\,,$$

for any $\varphi \in W = \{\varphi \in H_0^{1,2}(Q) : \ \partial\varphi/\partial t \in H^{-1,2}(Q)\}$ with $\varphi(T) = 0$.

For more details about space W see [CS3], where, it is proved, for instance, that $W \subset C([0,T], L^2(\Omega))$. Before we prove the main result of this section, Theorem 3.14, we need two preliminary results stated in Lemmas 3.7 and 3.11.

Lemma 3.7. *Let w be an A_2 weight, $S = \sum_i \partial_i fi \in H_w^{-1,2}(Q)$, $g/w \in L_w^p(Q)$ for $p > h/(h-1)$, where $h > 1$ is an index for which inequality (1.6) holds and let $u_0 \in L^2(\Omega)$. If $u \in H_w^{1,2}(Q)$ is a weak solution of the problem*

$$\begin{cases} \partial_t u - L_1 u = g - S & in\ Q\,, \\ u(x,0) = u_0(x) & in\ \Omega\,, \\ u(x,t) = 0 & in\ \Sigma\,, \end{cases}$$

where $L_1 = \sum \partial_i(b_{ij}\partial_j)$ and $b_{ji} = b_{ij}$ are measurable functions satisfying

$$\sum_{i,j=1}^n b_{ij}\,\xi_i\xi_j \sim w(x)\,|\xi|^2\,, \tag{3.8}$$

then

$$\sup_{t\in[0,T]} \int_\Omega u(x,t)^2\,dx + \iint_Q |\nabla u(x,t)|^2\,w\,dx\,dt \le c\left(\|u_0\,; L^2(\Omega)\|^2 + \left\|\frac{g}{w}\,; L_w^{2p/(p+1)}(Q)\right\|^2 + \sum_{j=1}^n \left\|\frac{f_j}{w}\,; L_w^2(Q)\right\|^2\right), \tag{3.9}$$

where the constant c *depends only on* $c(w,2)$, *the equivalence constant in* (3.8), T *and* Ω.

PROOF. For any $\tau \in (0,T)$, we put $\tilde{Q}_\tau = \Omega \times (0,\tau)$; arguing as in [CS3, proof of Theorem 2.4], we get

$$\iint_{\tilde{Q}_\tau} \sum_{i,j=1}^{n} b_{ij}\, \partial_i u\, \partial_j u\, dx\, dt + \frac{1}{2}\int_\Omega u(x,\tau)^2\, dx = \frac{1}{2}\int_\Omega u(x,0)^2\, dx + \iint_{\tilde{Q}_\tau} g\, u\, dx\, dt + \iint_{\tilde{Q}_\tau} \sum_{i=1}^{n} \partial_i u\, f_i\, dx\, dt\,.$$

By standard arguments, keeping in mind (3.8), we can reduce ourselves to estimate the last two terms above by the right hand side of (3.9). For the last term it is quite easy. On the other hand, if we put $s = 2p/(p+1)$, for any $\varepsilon > 0$ we get

$$\begin{aligned}\iint_{\tilde{Q}_\tau} |g\, u|\, dx\, dt &= \iint_{\tilde{Q}_\tau} \frac{|g|}{\omega}\, \omega^{1/s}\, |u|\, \omega^{1/s'}\, dx\, dt \\ &\le \Big(\iint_{\tilde{Q}_\tau} \Big(\frac{|g|}{\omega}\Big)^s \omega\, dx\, dt\Big)^{1/s} \Big(\iint_{\tilde{Q}_\tau} |u|^{s'} \omega\, dx\, dt\Big)^{1/s'} \\ &\le \frac{1}{2\varepsilon^2}\Big(\iint_{\tilde{Q}_\tau} \Big(\frac{|g|}{\omega}\Big)^s \omega\, dx\, dt\Big)^{2/s} \\ &\quad + \frac{\varepsilon^2}{2}\Big(\iint_{\tilde{Q}_\tau} |u|^{s'}\, \omega\, dx\, dt\Big)^{2/s'}.\end{aligned}$$

Since $s' < 2h$, by (1.6) we have (let r = diameter of Ω)

$$\begin{aligned}\iint_{\tilde{Q}_\tau} |g\, u|\, dx\, dt \le \frac{1}{2\varepsilon^2}\Big(\iint_Q \Big(\frac{|g|}{\omega}\Big)^s \omega dx\, dt\Big)^{2/s} & \\ + \frac{\varepsilon^2}{2}\, \omega(Q)^{2/s'}\Big(\sup_{t\in[0,T]} \frac{1}{|\Omega|}\int_\Omega |u|^2\, dx & \\ + r^2\, \omega(Q) \iint_Q |\nabla u|^2\, \omega dx\, dt\Big)&,\end{aligned}$$

and the assertion follows with a convenient choice of ε.

In order to establish the notation for the proof of Theorem 3.11 we will state, in a simpler context, a lemma proved in [CS2].

Lemma 3.10 (Chiarenza and Serapioni). *Assume ω is a $A_{1+2/n}$ weight. Then for any $(\xi,\tau) \in Q$ there is $R = R(\xi) > 0$ and a function $h : \Omega \times [0,T] \longrightarrow \mathbb{R}^+$, such that*

i) *$h(\xi,\cdot)$ is continuous, strictly increasing and $h(\xi,0) = 0$. (We will also denote $h(\xi,t)$ by $h_\xi(t)$),*

ii) *the set $Q_r(\xi,\tau) = \{(x,t) \in \mathbb{R}^{n+1} : x \in B(\xi,r),\ \tau - h(\xi,r) < t < \tau\}$ is contained in Q, for $r < R(\xi)$, (we will call $Q_r(\xi,\tau)$ a "standard cylinder"),*

iii) $\displaystyle\int_{Q_r(\xi,\tau)} \omega(x)\,dx\,dt \simeq r^{n+2}$,

iv) *there is a constant $\sigma_0 > 1$ such that $h(\xi;2r) < \sigma_0\, h(\xi,r)$, where σ_0 depends on $c(\omega,1+2/n)$ and n, but it is independent of ξ, τ and r.*

v) *There is $\sigma \in (0,1)$ such that $h(\xi;\sigma r) \le h(\xi,r)/4$. Here σ depends only on $c(\omega,1+2/n)$ and n.*

It is also useful to define the following sets:

$$Q_r^+(\xi,\tau) = \left\{(x,t) :\ |x-\xi| < \frac{r}{2},\ \tau - \frac{h(\xi,r)}{4} < t < \tau\right\},$$

$$Q_r^-(\xi,\tau) = \left\{(x,t) :\ |x-\xi| < \frac{r}{2},\ \tau - \frac{7}{8}\,h(\xi,r) < t < \tau - \frac{5}{8}\,h(\xi,r)\right\}.$$

Theorem 3.11 (Chiarenza and Serapioni). *Assume* (1.1) *holds with $\omega \in A_2$. For any $S \in H^{-1,2}(Q)$ there exists a unique $u = G(S) \in H_0^{1,2}(Q)$ which is a weak solution of the problem*

$$\begin{cases} \partial_t u - Lu = -S\,, \\ u(x,0) = u_0(x)\,. \end{cases}$$

Moreover, if $p > 2l/(l-1)$, where l is an index for which (1.5) *and* (1.6) *hold, $S = \operatorname{div}\vec{f} \in H^{-1,p}(Q)$, then*

$$\operatorname*{ess\,sup}_Q |u(x,t)| \le C \left\| \frac{\vec{f}}{\omega}, L^p_\omega(Q) \right\| + \operatorname*{ess\,sup}_\Omega |u_0(x)|\,, \tag{3.12}$$

where C depends only on Ω, T, $c(\omega,2)$ and p. Finally, if $\omega \in A_{1+2/n}$ and l is chosen such that (1.7) *also holds then the solution u is Hölder* *continuous uniformly on the compact subsets of* Q *and, if* $Q_r(\bar{x},\bar{t}) \subset Q_R(\bar{x},\bar{t}) \subset\subset Q$, *then*

$$\underset{Q_r(\bar{x},\bar{t})}{\operatorname{osc}} u \le C_R \left(\frac{r}{R}\right)^{\alpha} \left(\underset{Q_R(\bar{x},\bar{t})}{\operatorname{osc}} u + \left\| \frac{\vec{f}}{\omega}, L^p_\omega(Q) \right\| \right), \tag{3.13}$$

for some $\alpha \in (0,1)$ *and* $C_R > 0$ *depending only on* n, ν, $c(\omega, 1+2/n)$ *and* p.

PROOF. The existence and (3.12) are proved explicitly in [CS3, Theorems 2.3 and 2.4]. The last assertion is stated in [CS2, Theorem 3.7], but we prefer to give an explicit proof which stresses the dependence of the constants. Let $Q_r(\bar{x},\bar{t}) \subset\subset Q$ be a standard cylinder; by Lemma 3.10.v), there exists a constant $\sigma \in (0,1)$ such that $Q_{\sigma r}(\bar{x},\bar{t}) \subset Q_r^+(\bar{x},\bar{t})$. Denote now by $v \in H_0^{1,2}(Q_r(\bar{x},\bar{t}))$ the weak solution of $(\partial_t - L)v = S$ in $Q_r(\bar{x},\bar{t})$ and $v \equiv 0$ on the parabolic boundary of the standard cylinder and put $u = v + \tilde{v}$, so that $L\tilde{v} = 0$. Hence, by [CS2, Theorems 3.3 and 3.6],

$$\begin{aligned}
\underset{Q_r(\bar{x},\bar{t})}{\operatorname{osc}} u &\le \underset{Q_r(\bar{x},\bar{t})}{\operatorname{osc}} \tilde{v} + 2 \sup_{Q_r(\bar{x},\bar{t})} |v| \\
&\le (1-\varepsilon) \underset{Q_{r/\sigma}(\bar{x},\bar{t})}{\operatorname{osc}} \tilde{v} + 2 \sup_{Q_r(\bar{x},\bar{t})} |v| \\
&\le (1-\varepsilon) \underset{Q_{r/\sigma}(\bar{x},\bar{t})}{\operatorname{osc}} u + 4 \sup_{Q_{r/\sigma}(\bar{x},\bar{t})} |v| \\
&\le (1-\varepsilon) \underset{Q_{r/\sigma}(\bar{x},\bar{t})}{\operatorname{osc}} u \\
&\qquad + c \left[\left(\frac{1}{|Q_r(\bar{x},\bar{t})|} \int_{Q_r(\bar{x},\bar{t})} |v|^2 \, dx \, dt \right)^{1/2} \right. \\
&\qquad\quad + \left(\frac{1}{\omega(Q_r(\bar{x},\bar{t}))} \int_{Q_r(\bar{x},\bar{t})} |v|^2 \, \omega \, dx \, dt \right)^{1/2} \\
&\qquad\quad \left. + r \left(\frac{1}{\omega(Q_r(\bar{x},\bar{t}))} \int_{Q_r(\bar{x},\bar{t})} \left(\frac{|f|}{\omega} \right)^p \omega \, dx \, dt \right)^{1/p} \right],
\end{aligned}$$

where ε depends only on $c(\omega, 1+2/n, \nu, n)$. Let us consider now the terms between curly brakets. To estimate the last term, note that, if $r < R$ then

$$r\,\omega(Q_r(\bar{x},\bar{t}))^{-1/p} \simeq r^{1-(n+2)/p},$$

by Lemma 3.10.iii), with equivalence constants depending only on n, ν and $c(\omega, 1+2/n)$. But $p > 2l/(l-1) \ge 2h'/(h'-1)$ (where h' has been defined in Remark 1.8) and hence $1-(n+2)/p>0$, so that it can be estimated by $(r/R)^\theta$ (for some positive θ) times the average for $r=R$. Consider now the first term: it is estimated by

$$\begin{aligned}&\Big(\frac{1}{|B_r(\bar{x})|}\sup_{\{t:\,\bar{t}-h_r(\bar{x},\bar{t})<t<\bar{t}\}}\int_{B_r(\bar{x})}|v(x,t)|^2\,dx\Big)^{1/2}\\ &\qquad\le c\Big(\frac{1}{|B_r(\bar{x})|}\iint_{Q_r(\bar{x},\bar{t})}\Big(\frac{|f|}{\omega}\Big)^2\omega\,dx\,dt\Big)^{1/2}\\ &\qquad\le c\,r^{-n/2}\,\omega(Q_r(\bar{x},\bar{t}))^{(1-2/p)/2}\Big(\iint_{Q_r(\bar{x},\bar{t})}\Big(\frac{|f|}{\omega}\Big)^p\omega\,dx\,dt\Big)^{1/p}\\ &\qquad\le c\,r^{-n/2+(1-2/p)(n+2)/2}\Big(\iint_{Q_R(\bar{x},\bar{t})}\Big(\frac{|f|}{\omega}\Big)^p\omega\,dx\,dt\Big)^{1/p},\end{aligned}$$

where the first inequality follows from (3.9) and, if we look to the proof of Lemma 3.7, we see that the dependence of the constant c on the hight of Q appears only to bound the term in g that in the present case is zero. By remark 1.8, it is easy to see that we can choose $-n+(1-2/p)(n+2)>0$ and therefore we can repeat our previous arguments. Finally, by Sobolev inequality (1.5) and by the a priori estimate (3.9), the second term can be estimated by

$$\begin{aligned}&c\,r\Big(\frac{1}{\omega(Q_r(\bar{x},\bar{t}))}\int_{Q_r(\bar{x},\bar{t})}|\nabla v|^2\,\omega\,dx\,dt\Big)^{1/2}\\ &\qquad\le c\Big(\frac{r^2}{\omega(Q_r(\bar{x},\bar{t}))}\iint_{Q_r(\bar{x},\bar{t})}\Big(\frac{|f|}{\omega}\Big)^2\omega\,dx\,dt\Big)^{1/2}\\ &\qquad\le c\,r^{-n/2}\,\omega(Q_r(\bar{x},\bar{t}))^{(1-2/p)/2}\Big(\iint_{Q_r(\bar{x},\bar{t})}\Big(\frac{|f|}{\omega}\Big)^p.\end{aligned}$$

Note that all constants depend only on n, and $c(\omega,1+2/n)$. Thus we have proved that

$$\begin{aligned}\operatorname*{osc}_{Q_r(\bar{x},\bar{t})}u&\le(1-\varepsilon)\operatorname*{osc}_{Q_{r\sigma}(\bar{x},\bar{t})}u\\ &\quad+c\Big(\frac{r}{R}\Big)^\theta\Big(\frac{1}{\omega(Q_R(\bar{x},\bar{t}))}\int_{Q_R(\bar{x},\bar{t})}\Big(\frac{|f|}{\omega}\Big)^p\omega\,dx\,dt\Big)^{1/p}\end{aligned}$$

and the assertion follows.

Theorem 3.14. *Assume* $\omega\in A_2$, *(1.1) holds,* $S=\sum_i\partial_i f_i\in H_\omega^{-1,2}(Q)$, $g/\omega\in L^p_\omega$ *for* $p>2l/(l-1)$ *and* $u_0\in L^2(\Omega)$, *where* $l>1$ *is an index for* *which inequalities* (1.5) *and* (1.6) *hold. Then the solution* $u \in H^{1,2}_{\omega}(Q)$ *of the problem*

$$\text{(P)} \qquad \begin{cases} (\partial_t - L)u = g + S & \text{in } Q\,, \\ u(x,0) = u_0 & \text{in } \Omega\,, \\ u(x,t) = 0 & \text{on } \Sigma\,, \end{cases}$$

is the weak limit in $H^{1,2}_{\tilde{\omega}_1}(Q)$ *of a sequence of solutions* $u_m \in H^{1,2}_{\omega_m}(Q)$ *of the problems*

$$(\text{P}_m) \qquad \begin{cases} (\partial_t - L_m)u_m = g_m + S_m & \text{in } Q\,, \\ u_m(x,0) = u_0 & \text{in } \Omega\,, \\ u_m(x,t) = 0 & \text{on } \Sigma\,, \end{cases}$$

where $\omega_m = \omega_{mm}$ *and* $\tilde{\omega}_1$ *are as in Lemma* 2.1, $L_m = \sum_i \partial_i (a^{mm}_{ij} \partial_j)$, $g_m = g(\omega/\omega_m)^{(1-p)/p}$, $S_m = \sum_i \partial_i f_{mi}$, $f_{mi} = f_i(\omega/\omega_m)^{-1/2}$. *Moreover, if* $\omega \in A_{1+2/n}$ *and* l *is chosen such that* (1.7) *also holds then* u *is the uniform limit of* $(u_m)_{m\in\mathbb{N}}$ *in any compact subset of* Q.

PROOF. Obviously, $g_m/\omega_m \in L^p_{\omega_m}(Q)$ and $S_m \in H^{-1,p}_{\omega_m}(Q)$. Since u_m is a solution of (P_m), by (3.9),

$$\sup_{t\in[0,T]} \int_\Omega u_m(x,t)^2\,dx + \iint_Q |\nabla u_m(x,t)|^2\,\omega_m\,dx\,dt$$
$$\leq c\left(\|u_0; L^2(\Omega)\|^2 + \left\|\frac{g_m}{\omega_m}; L^{2p/(p+1)}_{\omega_m}(Q)\right\|^2 + \sum_{j=1}^n \left\|\frac{f_{mj}}{\omega_m}; L^2_{\omega_m}(Q)\right\|^2\right),$$

where the constant c depends only on $c(\omega, 2)$, (since $c(\omega_m, 2)$ depends only on $c(\omega, 2)$), ν, T and Ω.

Note that

$$\left\|\frac{f_{mj}}{\omega_m}; L^2_{\omega_m}(Q)\right\| = \left\|\frac{f_j}{\omega}; L^2_\omega(Q)\right\|,$$

$$\left\|\frac{g_m}{\omega_m}; L^{2p/(p+1)}_{\omega_m}(Q)\right\| \leq (\omega_m(\Omega)\,T)^{(p-1)/2p} \left\|\frac{g_m}{\omega_m}; L^p_{\omega_m}(Q)\right\|$$
$$\leq (\tilde{\omega}_2(\Omega)\,T)^{(p-1)/2p} \left\|\frac{g}{\omega}; L^p_\omega(Q)\right\|,$$

where $\tilde{\omega}_2$ was defined in Lemma 2.1.ii). Thus

$$\begin{aligned}(3.15)\qquad &\sup_{t\in[0,T]}\int_\Omega u_m(x,t)^2\,dx + \|\,|\nabla u_m|;L^2_{\omega_m}(Q)\|^2 \\ &\leq c\left(\|u_0;L^2(\Omega)\|^2 + (\tilde{\omega}_2(\Omega)\,T)^{(p-1)/p}\left\|\frac{g}{\omega};L^p_\omega(Q_T)\right\|^2\right. \\ &\qquad \left.+\sum_{j=1}^n\left\|\frac{f_j}{\omega};L^2_\omega(Q)\right\|^2\right) = C_1\,.\end{aligned}$$

In particular, since $\omega_m \geq \tilde{\omega}_1$ (see Lemma 2.1.ii)), we have

$$\sup_{t\in[0,T]}\|u_m(\,\cdot\,,t);L^2(\Omega)\|^2 + \|\,|\nabla u_m|;L^2_{\tilde{\omega}_1}(Q)\|^2 \leq C_1\,.$$

By (1.5), in view of the weak compactness of bounded sets in $H^{1,2}_{0,\tilde{\omega}_1}(Q)$ there exists a subsequence of $(u_m)_{m\in\mathbb{N}}$, again denoted by $(u_m)_{m\in\mathbb{N}}$, which converges weakly to an element u in $H^{1,2}_{0,\tilde{\omega}_1}(Q)$.

First we note that $u\in H^{1,2}_{0,\omega}(Q)$. In fact, fix $F_k = F_{k,k}$ (see Lemma 2.1). We know that $\partial u_m/\partial x_i \to \partial u/\partial x_i$ weakly in $L^2_{\tilde{\omega}_1}(Q)$ and that if $\varphi\in L^2_\omega(Q)$ then $\varphi\,\chi_{F_k}\in L^2_{\tilde{\omega}_1}$ (since on F_k, $\omega\simeq\tilde{\omega}_1$). Thus,

$$\iint \frac{\partial u_m}{\partial x_i}\,\chi_{F_k}\,\varphi\,\tilde{\omega}_1\,dx\,dt \longrightarrow \iint \frac{\partial u}{\partial x_i}\,\chi_{F_k}\,\varphi\,\tilde{\omega}_1\,dx\,dt\,,$$

for any $\varphi\in L^2_\omega(Q)$ and hence $(\chi_{F_k}\,\partial u_m/\partial x_i)_{m\in\mathbb{N}}$ converges weakly to a function in $L^2_\omega(Q)$, again since $\omega\simeq\tilde{\omega}_1$ on F_k. Therefore

$$\|\nabla u;L^2_\omega(F_k)\| \leq \limsup_{m\to\infty}\|\nabla u_m;L^2_\omega(F_k)\|$$

and since for $m\geq k$, $\omega=\omega_m$ over F_k, it follows that

$$\|\nabla u;L^2_\omega(F_k)\| \leq \limsup_{m\to\infty}\|\nabla u_m;L^2_{\omega_m}(F_k)\| \leq C_1\,,$$

for any $k\in\mathbb{N}$. By monotone convergence theorem

$$\|\nabla u;L^2_\omega(Q)\| \leq C_1\,.$$

On the other hand the same argument shows that $u\in L^2_\omega(Q)$ and hence u belongs to $H^{1,2}_{0,\omega}(Q)$ by Theorem 3.1.

Next we have to show that u is a weak solution to the problem (P). Suppose $\varphi \in W$, $\varphi(T) = 0$. By a density argument (see [CS3, proof of Theorem 2.3]) we can suppose $\varphi \in C^\infty(\bar{Q})$ and $\varphi(\,\cdot\,,t)$ compactly supported in Ω for any t. We have

$$\iint_Q \big(\langle A\,\nabla u, \nabla\varphi\rangle - u\,\varphi_t\big)\,dx\,dt = \lim_{k\to\infty}\iint_{F_k} \big(\langle A\,\nabla u, \nabla\varphi\rangle - u\,\varphi_t\big)\,dx\,dt\,.$$

It is easy to see that the linear functional

$$u \longrightarrow \iint_Q \big(\langle A\,\nabla u, \nabla\varphi\rangle - u\,\varphi_t\big)\,\chi_{F_k}(x)\,dx\,dt$$

is continuous in $H^{1,2}_{0,\tilde{\omega}_1}(Q)$ (since $\tilde{\omega}_1 \simeq \omega$ in F_k by Lemma 2.1.iv). Since $u_m \to u$ weakly in $H^{1,2}_{0,\tilde{\omega}_1}(Q)$, $A^m = A$ in F_k for any $m \geq k$ and u_m is a solution of (P_m), it follows that

$$\begin{aligned}
\int_0^T \int_{F_k} &\big(\langle A\,\nabla u, \nabla\varphi\rangle - u\,\varphi_t\big)\,dx\,dt \\
&= \lim_{m\to\infty}\int_0^T\int_{F_k}\big(\langle A^m\,\nabla u_m, \nabla\varphi\rangle - u_m\,\varphi_t\big)\,dx\,dt \\
&= \lim_{m\to\infty}\Big(\iint_Q \big(\langle A^m\,\nabla u_m, \nabla\varphi\rangle - u_m\,\varphi_t\big)\,dx\,dt \\
&\qquad\qquad - \int_0^T\int_{F'_k}\big(\langle A^m\,\nabla u_m, \nabla\varphi\rangle - u_m\,\varphi_t\big)\,dx\,dt\Big) \\
&= \lim_{m\to\infty}\Big(\int_\Omega u_0(x)\,\varphi(x,0)\,dx \\
&\qquad\qquad + \sum_{j=1}^n \iint_Q f_{mj}(x,t)\,\frac{\partial\varphi}{\partial x_j}(x,t)\,dx\,dt \\
&\qquad\qquad - \iint_Q g_m(x,t)\,\varphi(x,t)\,dx\,dt \\
&\qquad\qquad - \int_0^T\int_{F'_k}\big(\langle A^m\,\nabla u_m, \nabla\varphi\rangle - u_m\,\varphi_t\big)\,dx\,dt\Big)\,.
\end{aligned}$$

Note now that both integrals over Q converge as $m \to \infty$. Indeed $f_{mj}(x,t) \to f(x,t)$ almost everywhere by Lemma 2.1.v) and

$$\iint_Q f_{mj}(x,t)\,\frac{\partial\varphi}{\partial x_j}(x,t)\,dx\,dt = \iint_Q \frac{f_j(x,t)}{\omega^{1/2}(x)}\,\frac{\partial\varphi}{\partial x_j}(x,t)\,\omega_m^{1/2}(x)\,dx\,dt,$$

where $f_j(x,t)\,\omega^{-1/2}(x) \in L^2(Q_T)$ (by hypothesis) and

$$\left|\frac{\partial\varphi}{\partial x_j}(x,t)\right| \omega_m^{1/2}(x) \le c_\varphi\, \tilde\omega_2^{1/2} \in L^2(\Omega)$$

(by Lemma 2.1.ii)), since $\tilde\omega_2 \in A_2$. Thus, the conclusion follows by Lebesgue dominated convergence theorem. Analogously, to prove the corresponding assertion for the integral containing g_m, we note that $g(x,t)\,\omega^{1/p-1}(x) \in L^p(Q)$ and $\omega_m^{1-1/p} \le \tilde\omega_2^{1-1/p} \in L^{p'}(\Omega)$, $1/p + 1/p' = 1$, since $\tilde\omega_2 \in A_2$. By difference, also the integrals over F'_k converge as $m \to \infty$ to $l(k) \in \mathbb{R}$, so that

$$\begin{aligned}\int_0^T \int_{F_k} \big(\langle A\,\nabla u, \nabla\varphi\rangle - u\,\varphi_t\big)\,dx\,dt = \int_\Omega u_0(x)\,\varphi(x,0)\,dx \\ + \sum_{j=1}^n \iint_Q f_j(x,t)\,\frac{\partial\varphi}{\partial x_j}(x,t)\,dx\,dt \\ - \iint_Q g(x,t)\,\varphi(x,t)\,dx\,dt + l(k)\,.\end{aligned}$$

But, by Cauchy-Schwarz inequality, Lemma 2.1.vi) and (3.15), we have

$$\begin{aligned}&\left|\int_0^T \int_{F'_k} \big(\langle A^m\,\nabla u_m, \nabla\varphi\rangle - u_m\,\varphi_t\big)\,dx\,dt\right| \\ &\quad\le c_\varphi\sqrt{T}\,\Big(\frac{1}{\nu}\Big(\iint_Q |\nabla u_m|^2\,\omega_m\,dx\,dt\Big)^{1/2}\,\omega_m(\Omega\cap F'_k)^{1/2} \\ &\qquad + \Big(\iint_Q u_m^2\,\omega_m\,dx\,dt\Big)^{1/2}\Big(\int_{F'_k\cap\Omega}\frac{1}{\omega_m}\,dx\Big)^{1/2}\Big) \\ &\quad\le c(\varphi,T,\nu,\text{data})\,\Big(\big(\omega_m(\Omega\cap F'_k)\big)^{1/2} + \Big(\int_{F'_k\cap\Omega}\frac{1}{\omega_m}\,dx\Big)^{1/2}\Big)\,.\end{aligned}$$

Note now that $\omega_m \le \tilde\omega_2$ and $1/\omega_m \le 1/\tilde\omega_1$ and that both $\tilde\omega_2$ and $1/\tilde\omega_1$ belong to A_∞. Hence (considering, for instance, the first case) there exists $\delta > 0$ such that, if Q_0 is a cube containing $\bar\Omega$, then ([GC/RF, Theorem 2.9, Chapter IV])

$$\omega_m(\Omega\cap F'_k) \le \tilde\omega_2(\Omega\cap F'_k) \le \tilde\omega_2(Q_0)\,\Big(\frac{|F'_k|}{|Q_0|}\Big)^\delta\,,$$

which is independent of m and tends to zero as $k \to \infty$, by Lemma 2.1.iv). Thus, taking the limit as $k \to \infty$, we conclude that u is a weak solution (which in turn is unique: and hence it is the limit of the original sequence, since our argument can be carried out in the same way starting from any subsequence of the original one).

Now, by (3.13) (uniform Hölder continuity of the solutions), keeping in mind that the norms of g_m and S_m are uniformly bounded with respect to m (as we showed at the beginning of the present proof), we obtain that the sequence $(u_m)_{m\in\mathbb{N}}$ is locally equicontinuous. In addition, it is locally uniformly bounded, since we can combine the local boundedness of the solutions ([CS2, Theorem 3.3]) with the a priori estimate arguing as in the proof of Hölder continuity (*i.e.* noting that the L^2_ω norm of u can be estimated by the analogous norm of its gradient, by Sobolev inequality (1.5)). Hence, we can apply Arzelà-Ascoli theorem and conclude that u_m converges to u uniformly on compact subsets (note that all converging subsequences converge to u).

4. The Green function.

In this section we will prove that there exists a weak Green function $\gamma(x,t;\xi,\tau)$ for $\partial_t u - Lu = 0$, where $Lu = \sum_{i,j}\partial_i(a_{ij}\partial_j u)$ satisfies condition (1.1), for any bounded cylinder $Q = \Omega \times (0,T)$. In particular, γ belongs to $H^{1,p'}_{0,\omega}(Q)$ and it gives a representation formula for the solutions of the Cauchy-Dirichlet problem. We will also derive some additional properties of γ which are analogous to the corresponding ones in the non-degenerate problems (see [A, Theorem 9]). The derivation is based on the fact that γ can be approximated, in convenient spaces, by Green functions in the sense of Aronson (see [A]). In the sequel, we will use the following notations: if τ is an arbitrary point in $[0,T)$ we set $Q_\tau = \Omega \times (\tau, T]$, while if t is an arbitrary point in $(0,T]$ we set $\tilde{Q}_t = \Omega \times [0,t)$.

Next, we prove the existence of a Green function for $\partial_t - L$ in Q, by following the abstract argument given by Aronson.

Theorem 4.1. *Suppose $\omega \in A_{1+2/n}$ and $p > 2l/(l-1)$, where l is an index for which* (1.5), (1.6) *and* (1.7) *hold. Then there exists a function $\gamma = \gamma(x,t;\xi,\tau)$ such that:*

i) $\gamma(x,t;\,\cdot\,,\,\cdot\,) \in H^{1,p'}_{0,\omega}(Q)$ *for* $(x,t) \in Q$ *and*

$$\|\gamma(x,t;\,\cdot\,,\,\cdot\,); H^{1,p'}_{0,\omega}(Q)\| \le C\,,$$

where C depends only on Ω, T, ν, $c(\omega,2)$ and p;

ii) *If u is a solution of the problem*

$$\begin{cases} (\partial_t - L)u = f_0 - \operatorname{div} \vec{f} \quad \text{in } Q\,, \\ u \equiv 0 \quad \text{on the parabolic boundary of } Q\,, \end{cases}$$

for some f_0, $\vec{f}$ such that f_0/ω, $|\vec{f}|/\omega \in L^p_\omega(Q)$, then

$$u(x,t) = \sum_{j=1}^{n} \iint_Q \partial_{\xi_j} \gamma(x,t;\xi,\tau)\, f_j(\xi,\tau)\, d\xi\, d\tau + \iint_Q \gamma(x,t;\xi,\tau)\, f_0(\xi,\tau)\, d\xi d\tau\,.$$

We will say that γ is the weak Green function for $\partial_t - L$.

PROOF. Let $\vec{f} = (f_1, \ldots, f_n)$ be such that $|\vec{f}|/\omega \in L^p_\omega$ for $p > 2l/(l-1)$. By Theorem 3.11, if $(x,t) \in Q$ and u is a solution of the problem stated there with $u_0 \equiv 0$ and $S = \operatorname{div} \vec{f}$, then the linear functional

$$\vec{f} \longmapsto u(x,t)$$

is well defined and is continuous on $H^{-1,p}_\omega(Q)$ by (3.13). By Theorem 3.1, there exists a function $\gamma(x,t;\,\cdot\,,\,\cdot\,)$ in $H^{1,p'}_0(Q)$ such that

$$u(x,t) = \iint_Q \sum_{j=1}^{n} \partial_{\xi_j} \gamma(x,t;\xi,\tau)\, f_j(\xi,\tau)\, d\xi\, d\tau\,,$$

and

$$\|\gamma(x,t;\,\cdot\,,\,\cdot\,); H^{1,p'}_0(Q)\| \le C\,,$$

where C is the constant of (3.12) and hence it depends only on Ω, T, ν, $c(\omega,2)$ and p. Then the assertion follows from the Remark 3.5.

Theorem 4.2. *Suppose $\omega \in A_{1+2/n}$, condition (1.1) holds, $u_0 \in L^2(\Omega)$ and G such that $G/\omega \in L^p_\omega(Q)$, $p > 2l/(l-1)$, where l is an index for which (1.5) and (1.6) hold. Then the weak Green function $\gamma(x,t;\xi,\tau)$ of $\partial_t - Lu$ in Q has the following properties:*

i) *$\gamma(x,t;\xi,\tau) = \tilde{\gamma}(\xi,\tau;x,t)$ in $Q \times Q$ for $t > \tau$, where $\tilde{\gamma}$ is the weak Green function for the adjoint problem in Q,*

ii) *For fixed* $(\xi, \tau) \in \tilde{Q}_T$ *let* Z *denote an arbitrary open domain such that* $\bar{Z} \subset \Omega \backslash \{\xi\}$. *Then the function* $\gamma(\cdot, \cdot; \xi, \tau)$ *is a weak solution of* $\partial_t - Lu = 0$ *in* $Z \times (\tau, T)$ *with initial value zero on* $t = \tau$ *and vanishing on the lateral boundary. For fixed* $(x, t) \in Q_0$ *let* Z *denote an arbitrary open domain such that* $\bar{Z} \subset \Omega \backslash \{x\}$. *Then the function* $\gamma(x, t; \cdot, \cdot)$ *is a weak solution of the adjoint problem in* $Z \times (0, t)$ *with initial value zero on* $\tau = t$ *and vanishing on the lateral boundary.*

iii) *The weak solution of the boundary value problem*

$$\text{(P)} \qquad \begin{cases} (\partial_t - L)u = G(x,t) & \text{in } Q\,, \\ u(x,0) = u_0(x) & \text{for } x \text{ in } \Omega\,, \\ u(x,t) = 0 & \text{for } (x,t) \in \Sigma\,, \end{cases}$$

is given by

$$u(x,t) = \int_\Omega \gamma(x,t;\xi,0)\, u_0(\xi)\, d\xi + \iint_Q \gamma(x,t;\xi,\tau)\, G(\xi,\tau)\, d\xi\, d\tau\,.$$

iv) *The weak solution of the adjoint boundary value problem*

$$(\tilde{\text{P}}) \qquad \begin{cases} (-\partial_t - L)v = G(x,t) & \text{in } Q\,, \\ v(\xi,T) = u_0(\xi) & \text{for } \xi \text{ in } \Omega\,, \\ v(\xi,t) = 0 & \text{for } (x,t) \in \Sigma\,, \end{cases}$$

is given by

$$v(\xi,\tau) = \int_\Omega \gamma(x,T;\xi,\tau)\, u_0(x)\, dx + \iint_Q \gamma(x,t;\xi,\tau)\, G(x,t)\, dx\, dt\,.$$

PROOF. We denote by γ_m the Green function of $\partial_t - L^m$ (with the notations of Theorem 3.14). First, we assert that

$$\text{(4.3)} \qquad \gamma_m(x,t;\cdot,\cdot)\,\omega_m^{1/p'} \longrightarrow \gamma(x,t;\cdot,\cdot)\,\omega^{1/p'}$$

and

$$\text{(4.4)} \qquad \gamma_m(\cdot,\cdot;\xi,\tau)\,\omega_m^{1/p'} \longrightarrow \tilde{\gamma}(\xi,\tau;\cdot,\cdot)\,\omega^{1/p'}$$

weakly in $L^{p'}(Q)$, for any p' whose Hölder conjugate satisfies $p > 2l/(l-1)$ (*i.e*, $p' < 2l/(l+1)$). First, we prove (4.3). Fix $f_0 \in L^p(Q)$ and put $g_0 = \omega^{1/p'} f_0$, so that $g_0/\omega \in L^p_\omega(Q)$. Then, as in Theorem 3.14, consider the solution u_m of the Dirichlet problem

$$\begin{cases} (\partial_t - L^m)u_m = (g_0)_m \ \text{ in } Q\,, \\ u_m \equiv 0 \ \text{ on the parabolic boundary of } Q\,. \end{cases}$$

By Theorem 3.14, we know that $\{u_m\}_{m\in\mathbb{N}}$ converges uniformly on compact subsets of Q to the solution u of the problem

$$\begin{cases} (\partial_t - L)u = g_0 \ \text{ in } Q\,, \\ u \equiv 0 \ \text{ on the parabolic boundary of } Q\,. \end{cases}$$

So, by definition of the Green function, Theorem 4.1, and [A, Theorem 9],

$$\iint_Q \gamma_m(x,t;\xi,\tau)\,(g_0)_m(\xi,\tau)\,d\xi\,d\tau \longrightarrow \iint_Q \gamma(x,t;\xi,\tau)\,g_0(\xi,\tau)\,d\xi\,d\tau\,.$$

Recalling that $(g_0)_m = g_0(\omega_m/\omega)^{1/p'} = f_0\,\omega_m^{1/p'}$ we are done. In addition, (4.4) follows by applying the same argument to the adjoint equation.

Now, let $(\xi,\tau) \in \tilde{Q}_T$ be fixed and let $\delta > 0$ be such that $\tau < T - \delta < T$; from [A, Theorem 9.v)], we know that $\gamma_m(\,\cdot\,,\,\cdot\,;\xi,\tau)$ is a solution of the Dirichlet problem

$$\begin{cases} (\partial_t - L^m)v = 0 \ \text{ in } Q_{\tau+\delta}\,, \\ v(x,\tau+\delta) = \gamma_m(x,\tau+\delta;\xi,\tau) \ \text{ in } \Omega\,, \\ v \equiv 0 \ \text{ on } \partial\Omega\times]\tau+\delta,T]\,. \end{cases}$$

By Lemma 3.7 we have

$$\sup_{\tau+\delta\le t\le T} \|\gamma_m(\,\cdot\,,t;\xi,\tau);L^2(\Omega)\|^2 + \|\nabla_x\gamma_m(\,\cdot\,,\,\cdot\,;\xi,\tau);L^2_{\omega_m}(Q_{\tau+\delta})\|^2 \le c\,\|\gamma_m(\,\cdot\,,\tau+\delta;\xi,\tau);L^2(\Omega)\|^2\,,$$

where c does not depend on m. Now, by [GN] we have (note that Green function γ_m satisfies the assumptions of [GN] and, in addition, a fundamental solution Γ_m exists by [A], since L^m is an usual elliptic operator)

$$\gamma_m(x,\tau+\delta;\xi,\tau) \le \Gamma_m(x,\tau+\delta;\xi,\tau) \le c\left(\frac{1}{(h_x)_m^{-1}(\delta)} + \frac{1}{(h_\xi)_m^{-1}(\delta)}\right),$$

where the constant c is independent of m and the function $(h_x)_m$ corresponds to h_x at the step m. By [CS1, Proposition 1.1] there exist positive numbers l_1 and l_2 (both depending only on $c(\omega, 1+2/n)$) and there are two constants c_1 and c_2 (depending only on $c(\omega, 1+2/n)$, l_1, l_2 and Ω) such that

$$c_1\, r^{l_1} \leq (h_{x_0})_m(r) \leq c_2\, r^{l_2}\,.$$

In particular,

$$c_2'\, t^{1/l_2} \leq ((h_{x_0})_m)^{-1}(t) \leq c_1'\, t^{1/l_1}\,, \tag{4.5}$$

where c_1' and c_2' depend only on $c(\omega, 1+2/n)$ and therefore

$$\gamma_m(x, \tau+\delta; \xi, \tau) \leq C\, \delta^{-1/l_2}$$

and

$$\|\gamma_m(\,\cdot\,, \tau+\delta; \xi, \tau); L^2(\Omega)\|^2 \leq C\, \delta^{-2/l_2}\,.$$

Thus we have

$$\sup_{\tau+\delta\leq t\leq T} \|\gamma_m(\,\cdot\,, t; \xi, \tau); L^2(\Omega)\|^2 + \|\nabla_x\gamma_m(\,\cdot\,,\,\cdot\,; \xi, \tau); L^2_{\omega_m}(Q_{\tau+\delta})\|^2 \leq C\, \delta^{-2/l_2}$$

and since $\omega_m \geq \tilde{\omega}_1$,

$$\sup_{\tau+\delta\leq t\leq T} \|\gamma_m(\,\cdot\,, t; \xi, \tau); L^2(\Omega)\|^2 + \|\nabla_x\gamma_m(\,\cdot\,,\,\cdot\,; \xi, \tau); L^2_{\tilde{\omega}_1}(Q_{\tau+\delta})\|^2 \leq C\, \delta^{-2/l_2}\,.$$

Hence (keeping in mind the Sobolev inequality for compact support functions in space variables), there exists a subsequence of $\gamma_m(\,\cdot\,,\,\cdot\,; \xi, \tau)$ which converges weakly to a limit function γ^* in $H^{1,2}_{0,\tilde{\omega}_1}(Q_{\tau+\delta})$. Next we show that γ^* is $\tilde{\gamma}(\xi, \tau, \,\cdot\,, \,\cdot\,)$. By (4.4),

$$\gamma_m(\,\cdot\,,\,\cdot\,; \xi, \tau)\, \omega_m^{1/p'} \longrightarrow \tilde{\gamma}(\xi, \tau; \,\cdot\,, \,\cdot\,)\, \omega^{1/p'}$$

weakly in $L^{p'}(Q_{\tau+\delta})$, $p' < 2l/(l+1)$. On the other hand, $\omega_m \geq \tilde{\omega}_1$ (by Lemma 2.1.iv)) and hence, for any $g \in L^p(Q_{\tau+\delta})$, the functions $(\tilde{\omega}_1/\omega_m)^{1/p'} g$ are uniformly bounded by a constant times g for any m and converge to $(\tilde{\omega}_1/\omega)^{1/p'} g$ as m tends to infinity by Lemma 2.1.v).

Then, by Lebesgue Theorem, $(\tilde{\omega}_1/\omega_m)^{1/p'}g$ converges to $(\tilde{\omega}_1/\omega)^{1/p'}g$ in $L^p(Q_{\tau+\delta})$ and hence

$$\iint_{Q_{\tau+\delta}} \gamma_m(x,t,\xi,\tau)\,\tilde{\omega}_1^{1/p'}(x)\,g(x,t)\,dx\,dt = \iint_{Q_{\tau+\delta}} \gamma_m(x,t,\xi,\tau)\,\omega_m^{1/p'}(x)\left(\frac{\tilde{\omega}_1}{\omega_m}\right)^{1/p'}(x)\,g(x,t)\,dx\,dt$$

tends to

$$\iint_{Q_{\tau+\delta}} \tilde{\gamma}(x,t,\xi,\tau)\,\omega^{1/p'}(x)\left(\frac{\tilde{\omega}_1}{\omega}\right)^{1/p'}(x)\,g(x,t)\,dx\,dt$$

as $m \to \infty$, *i.e.*

$$\gamma_m(\,\cdot\,,\,\cdot\,;\xi,\tau)\,\tilde{\omega}_1^{1/p'} \longrightarrow \tilde{\gamma}(\xi,\tau;\,\cdot\,,\cdot\,)\,\tilde{\omega_1}^{1/p'}$$

weakly in $L^{p'}(Q_{\tau+\delta})$, $p' < 2l/(l+1)$. Or equivalently

$$\gamma_m(\,\cdot\,,\,\cdot\,;\xi,\tau) \longrightarrow \tilde{\gamma}(\xi,\tau;\,\cdot\,,\,\cdot\,)$$

weakly in $L^{p'}_{\tilde{\omega}_1}(Q_{\tau+\delta})$, $p' < 2l/(l+1)$. But

$$\gamma_m(\,\cdot\,,\,\cdot\,;\xi,\tau) \longrightarrow \gamma^*(\xi,\tau;\,\cdot\,,\,,\cdot\,)$$

weakly in $L^2_{\tilde{\omega}_1}(Q_{\tau+\delta})$ and therefore $\tilde{\gamma} = \gamma^*$ almost everywhere. The same argument can be carried out starting from any subsequence $\{\gamma_{m_k}\}_{k\in\mathbb{N}}$ and hence $\gamma_m(\,\cdot\,,\,\cdot\,;\xi,\tau)$ converges weakly in $H^{1,2}_{0,\tilde{\omega}_1}$ to $\tilde{\gamma}(\xi,\tau;\,\cdot\,,\,\cdot\,)$. As in the proof of Theorem 3.14, $\tilde{\gamma}(\xi,\tau;\,\cdot\,,\,\cdot\,)$ is a weak solution of $Lu = 0$ for $(x,t) \in Q_\tau$. If we hold $(x,t) \in Q_0$ fixed and apply the same argument to γ_m considered as function of (ξ,τ), we find that $\gamma(x,t;\cdot\,,\,\cdot\,)$ is a weak solution of $\tilde{L}v = 0$ for $(\xi,\tau) \in \tilde{Q}_t$.

On the other hand, for fixed $(\xi,\tau) \in \tilde{Q}_T$ it follows from [GN, (1.3)], (4.5) and (3.13) that the sequence $\{\gamma_m(\,\cdot\,,\,\cdot\,;\xi,\tau)\}_{m\in\mathbb{N}}$ is uniformly bounded and equicontinuous for (x,t) in any compact subset of Q_τ. Thus,

$$\gamma_m(\,\cdot\,,\,\cdot\,;\xi,\tau) \longrightarrow \tilde{\gamma}(\xi,\tau;\,\cdot\,,\,\cdot\,)$$

uniformly in any compact subset of Q_τ. Similarly, for each $(x,t) \in Q_0$,

$$\gamma_m(x,t;\,\cdot\,,\,\cdot\,) \longrightarrow \gamma(x,t;\,\cdot\,,\,\cdot\,)$$

uniformly in any compact subset of $\tilde{Q}_t$. Thus, in particular, i) holds. From i) and what we proved before ii) also holds.

According to Theorem 3.14, if u is a solution of the problem (P) then for $(x,t) \in Q$ we have $u_m(x,t) \to u(x,t)$ where u_m is the solution of the approximate problem (P_m). Here we are using the notation introduced in Theorem 3.14. If γ_m is the Green function associated with problem (P_m) then we have

$$u_m(x,t) = \int_\Omega \gamma_m(x,t;\xi,0)\, u_0(\xi)\, d\xi + \iint_Q \gamma_m(x,t;\xi,\tau)\, G_m(\xi,\tau)\, d\xi\, d\tau\,.$$

Now let Γ_m denotes the fundamental solution associated with problem (P_m). Then

$$\gamma_m(x,t;\xi,\tau) \le \Gamma_m(x,t;\xi,\tau)\,,$$

and by [GN, (1.3)] it follows that

$$\gamma_m(x,t,\xi,0) \le C\left(\frac{1}{[(h_x)_m^{-1}(t)]^n} + \frac{1}{[(h_\xi)_m^{-1}(t)]^n}\right)$$

where C depends only on ν, n, $c(\omega, 1{+}1/2n)$. By (4.5) $\{\gamma_m(x,t;\cdot,0)\}_{m\in\mathrm{N}}$ is uniformly bounded in $L^2(\Omega)$ and therefore

$$\int_\Omega \gamma_m(x,t;\xi,0)\, u_0(\xi)\, d\xi \longrightarrow \int_\Omega \gamma(x,t;\xi,0)\, u_0(\xi)\, d\xi\,,$$

as m tends to infinity. Moreover, $\{\gamma_m(x,t;\,\cdot\,,\,\cdot\,)\,\omega_m^{1/p'}\}_{m\in\mathrm{N}}$ is uniformly bounded in $L^{p'}(Q)$ for $p' < 2l/(l+1)$ (by (4.3)). Since

$$\iint_Q \gamma_m(x,t;\xi,\tau)\, G_m(\xi,\tau)\, d\xi\, d\tau$$

$$= \iint_Q \gamma_m(x,t;\xi,\tau)\, \omega_m^{1/p'}\, \frac{G}{\omega}\, \omega^{1/p}\, d\xi\, d\tau\,,$$

and $(G/\omega)\,\omega^{1/p} \in L^p$ we have that

$$\iint_Q \gamma_m(x,t;\xi,\tau)\, G_m(\xi,\tau)\, d\xi\, d\tau \longrightarrow \iint_Q \gamma(x,t;\xi,\tau)\, G(\xi,\tau)\, d\xi\, d\tau\,,$$

as m tends to infinity. This proves iii), and the proof of iv) is similar.

REMARK 4.6, If $\omega \equiv 1$ then the lower bound of the set of p such that Theorem 3.14 and Theorem 4.2 hold does not coincide with the analogous bound established in [A, Theorem 9 and Theorem 1], since we have followed the results proved in [CS1], [CS2] and [CS3]. On the other hand, in the degenerate case, the optimal value of l and hence of p is rather implicit, since it depends on the lower bound of the set of q such that $\omega \in A_{1+q}$ (see for instance [W] and [F/SC]).

References.

[A] Aronson, D., Non-negative solutions of linear parabolic equations. *Ann. Scuola Norm. Sup. Pisa* **22** (1968), 607-693.

[AS] Aronson, D., Serrin, J., Local behavior of solutions of quasilinear parabolic equations. *Arch. Rational Mech. Anal.* **25** (1967), 81-122.

[CP/SC] Chiadò Piat, V., Serra Cassano, F., Relaxation of degenerate variational integrals. *Nonlinear Analysis* **22** (1994), 409-424.

[CW1] Chanillo, S., Wheeden, R., Harnack's inequality and mean value inequalities for solutions of degenerate elliptic equations. *Comm. Partial Diff. Equations* **11** (1986), 1111-1134.

[CW2] Chanillo, S., Wheeden, R., Weighted Poincaré and Sobolev inequalities for the Peano maximal function. *Amer. J. Math.* **107** (1985), 1191-1226.

[CS1] Chiarenza, F., Serapioni, R., A Harnack inequality for degenerate parabolic equations. *Comm. Partial Diff. Equations* **9** (1984), 719-749.

[CS2] Chiarenza, F., Serapioni, R., Pointwise estimates for degenerate parabolic equations. *Appl. Analysis* **23** (1987), 287-299.

[CS3] Chiarenza, F., Serapioni, R., Degenerate parabolic equations and Harnack inequality. *Ann. Mat. Pura ed Appl.* **137** (1984), 139-162.

[CF] Coifman, R. R., Fefferman, Ch., Weighted norm inequalities for maximal functions and singular integrals. *Studia Math.* **51** (1974), 241-250.

[F] Friedman, A., *Partial differential equations of parabolic type.* Prentice-Hall, 1964.

[Fe1] Fernandes, J., Mean value and Harnack inequalities for a class of degenerate parabolic equations. PhD. thesis, Rutgers University, USA.

[Fe2] Fernandes, J., Mean value and Harnack inequalities for a certain class of degenerate parabolic equations. *Revista Mat. Iberoamericana* **7** (1991), 241-286.

[FJK] Fabes, E., Jerison, D., Kenig, C., The Wiener test for degenerate elliptic equations. *Ann. Inst. Fourier* **32** (1982), 151-182.

[FS] Franchi, B., Serapioni, R., Pointwise estimates for a class of strongly degenerate elliptic operators:a geometrical approach. *Ann. Scuola Norm. Sup. Pisa* **14** (1987), 527-568.

[F/SC] Franchi, B., Serra Cassano, F., Gehring's Lemma for metrics and higher integrability of the gradient for minimizers of noncoercive variational functionals. To appear in *Studia Math.*

[FSt] Fabes, E., Stroock, D., The L^p-integrability of Green's functions and Fundamental solutions for elliptic and parabolic equations. *Duke Math. J.* **51** (1984), 997-1016.

[G] Greco, L., An approximation theorem for the Γ^--convergence of degenerate quadratic functionals. *Riv. Mat. Pura Appl.* **7** (1990), 53-80.

GC/RF] Garcia-Cuerva, J., Rubio de Francia, J. L., *Weighted norm inequalities and related topics.* North-Holland Math. Studies **116**, 1985.

[GW1] Gutiérrez, C., Wheeden, R., Sobolev interpolation inequalities with weights. *Trans. Amer. Math. Soc.* **323** (1991), 263-281.

[GW2] Gutiérrez, C., Wheeden, R., Mean value and Harnack inequalities for degenerate parabolic equations. *Colloquium Math.* (Volume dedicated to A. Zygmund), **60/61** (1990), 157-194.

[GW3] Gutiérrez, C., Wheeden, R., Harnack's inequality for degenerate parabolic equations. *Comm. Partial Diff. Equations* **16** (1991), 745-770.

[GW4] Gutiérrez, C., Wheeden, R., Bounds for the fundamental solution of degenerate parabolic equations. *Comm Partial Diff Equations* **17** (1992), 1287-1307.

[GN] Gutiérrez, C., Nelson, F., Bounds for the fundamental solution of degenerate parabolic equations. *Comm. Partial Diff. Equations* **13** (1988), 635-649.

[GT] Gilbarg, D., Trudinger, N. S., *Elliptic partial equations of second order.* Second edition, Springer-Verlag, 1977.

[KS] Kinderlehrer, D., Stampacchia, G., *An introduction to variational inequalities.* Academic Press, 1983.

[LSU] Ladyzenskaja, O., Solonnikov, V., Uralceva, N., *Linear and quasilinear equations of parabolic type.* Amer. Math. Soc., 1968.

[M] Moser, J., A Harnack inequality for parabolic differential equations. *Comm. Pure Appl. Math.* **17** (1964), 101-134.

[Mu] Muckenhoupt, B., Weighted norm inequalities for the Hardy maximal function. *Trans. Amer. Math. Soc.* **165** (1972), 207-226.

[S] Stein, E., *Singular integrals and differentiability of functions.* Princeton Univ. Press, 1970.

[SW] Sawyer, E., Wheeden, R., Weighted inequalities for fractional integrals in Euclidean and homogeneous spaces. *Amer. J. Math.* **114** (1992), 813-874.

[T] Treves, F., *Basic linear partial differential equations.* Academic Press, 1975.

[W] Wik, I., On Moukenhoupt's classes of weight functions. *Studia Math.* **94** (1989), 245-255.

Recibido: 5 de septiembre de 1.995

José C. Fernandes*
Departamento de Matemática
Universidade de São Paulo
Caixa Postal 20570, J. Paulistano
01452-990, São Paulo, SP, BRASIL
diniz@ime.usp.br

Bruno Franchi†
Dipartimento di Matematica
Universitá di Bologna
Piazza di Porta San Donato, 5
40127, Bologna, ITALIA
franchib@dm.unibo.it

* Supported by M.U.R.S.T and C.N.R. (Italy), CNPq and Fapesp (Brazil). The first author wants to also express his gratitude to the Department of Mathematics of the University of Bologna for the kindness and hospitality. The authors are also indebted to R. Serapioni for many helpful discussions.

† Supported by M.U.R.S.T., Italy (40 and 60 %) and G.N.A.F.A of C.N.R., Italy.

Revista Matemática Iberoamericana
Vol. 12, N.° 2, 1996

A new technique to estimate the regularity of refinable functions

Albert Cohen and Ingrid Daubechies

Abstract. We study the regularity of refinable functions by analyzing the spectral properties of special operators associated to the refinement equation; in particular, we use the Fredholm determinant theory to derive numerical estimates for the spectral radius of these operators in certain spaces. This new technique is particularly useful for estimating the regularity in the cases where the refinement equation has an infinite number of nonzero coefficients and in the multidimensional cases.

1. Introduction.

Refinable functions are functions that satisfy a refinement equation, *i.e.*

$$\varphi(x) = 2 \sum_n c_n \, \varphi(2x - n) . \tag{1.1}$$

The coefficients c_n are often, but not always, chosen finite in number. Such functions appear in different settings, most notably in subdivision schemes for computer aided design, where they are tools for the fast generation of smooth curves and surfaces (see Cavaretta, Dahmen and Micchelli (1991) and Dyn (1992) for reviews), and in the construction of wavelet bases and multiresolution analysis (see Mallat (1989), Daubechies (1988) and Meyer (1990)).

One of the earliest examples of refinable functions are the B-splines with equally spaced simple knots (see de Boor (1978) for a general review on splines), where

$$c_n = \binom{N}{n} 2^{-N+1},$$

and the corresponding φ_N is a C^{N-2} function, piecewise-polynomial of order $N-1$. Among refinable functions, the B-spline case is exceptional in that $\varphi(x)$ is given by an explicit analytical expression; in many other cases of interest, φ is defined by fixing an appropriate choice for the c_n in (1.1), and it is not immediately clear how smooth φ is. Over the years, several techniques have been developed to determine the regularity of refinable functions. In this paper, we present a new technique for this purpose.

The regularity of a function φ can be measured in different ways; we shall restrict ourselves to Hölder and Sobolev exponents. If φ is in C^n but not in C^{n+1}, then its Hölder exponent is given by $\mu = n + \nu$ with

$$\nu = \inf_x \left(\liminf_{|t|\to 0} \frac{\log |\varphi^{(n)}(x+t) - \varphi^{(n)}(x)|}{\log |t|} \right), \tag{1.2}$$

where $\varphi^{(n)}$ is the n-th derivative of φ.

The Sobolev exponent s is defined by

$$s = \sup \left\{ \gamma : \int |\hat{\varphi}(\omega)|^2 (1 + |\omega|^2)^\gamma \, d\omega < +\infty \right\}, \tag{1.3}$$

where $\hat{\varphi}(\omega) = \int \varphi(x) e^{-i\omega x} dx$ is the Fourier transform of φ. One can generalize this to L^p-Sobolev exponents s_p which, following Hervé (1995), we define by

$$s_p = \sup \left\{ \gamma : \int |\hat{\varphi}(\omega)|^p (1 + |\omega|^p)^\gamma \, d\omega < +\infty \right\}, \tag{1.4}$$

These different regularity indices are related to each other by $s = s_2$, $\mu \geq s_1$ and, by Hölder's inequality, $s_r + r^{-1} \leq s_p + p^{-1}$, for $0 \leq p \leq r$.

Most of the techniques developed to estimate the regularity of a refinable function concentrate on the case where only finitely many h_n are non-zero, which was, until recently, the only case of interest for applications: in subdivision schemes it corresponds to finite masks; in wavelet constructions, to compactly supported scaling functions and wavelets. If there are only $N+1$ nonzero c_n, then Micchelli and Prautzsch (1989) and Daubechies and Lagarias (1991, 1992) showed how to find, at least in principle, the Hölder exponent of φ by computing bounds on the norms of $N \times N$ matrices; in practice, this method becomes quickly impractical if N is not small. Still for the case of finitely many nonzero c_n, a technique that can handle larger N was proposed by Rioul (1992) and Dyn and Levin (1991); it is still the best available technique for finding the Hölder exponent when N is finite. In general it is easier to compute the Sobolev exponents; these can moreover be used to find a bound on the Hölder exponent, since $\mu \geq s_1 \geq s_2 + 1/2$. In the case where $m(\omega) = \sum_n h_n e^{-in\omega}$ is a nonnegative trigonometric polynomial, one even has $\mu = s_1$, which was exploited in one of the first computations of the regularity of a refinable function in Dubuc (1986) for the special case $h_0 = 1$, $h_{\pm 1} = 9/16$, $h_{\pm 3} = -1/16$, all other $h_n = 0$, related to Lagrangian interpolation and later generalized in Dubuc and Deslauriers (1989). When $m(\omega)$ is not restricted to be nonnegative, most of the first developments were concentrated on the computation of $s_2 = s$; see Conze (1989), Eirola (1992) and the appendix in Daubechies (1988). Extensions to the computation of s_p (including $p = 1, 2$) have appeared in Gripenberg (1992), Villemoes (1992) and Hervé (1992, 1995). With the exception of Hervé (1992, 1995), all the papers above apply only to N finite. Most of them are also hard to generalize to the multidimensional case where (1.1) is replaced by

$$\varphi(x) = |D| \sum_n h_n \, \varphi(Dx - n)\,, \tag{1.5}$$

where $n \in \mathbb{Z}^d$, $x \in \mathbb{R}^d$ and D is a $d \times d$ matrix with integer entries and all eigenvalues strictly superior to 1 in absolute value; $|D|$ is the determinant of D. Examples of the type (1.5) occur in *e.g.* wavelet bases corresponding to quincunx subsampling in two dimensions proposed for image processing in Vetterli (1984), Feauveau (1989) and Kovačević and Vetterli (1990), with explicit orthonormal wavelet bases in Feauveau (1990), Kovačević and Vetterli (1992) and Cohen and Daubechies (1993). As illustrated by the trickiness of the estimates in Cohen and Daubechies (1993) and especially in Villemoes (1993), it is not easy to find the regularity in the multidimensional case by generalizing the approach referred to above, even when only finitely many c_n are nonzero.

In this paper, we present a different technique for computing s_p for a refinable function. This technique is independent of whether the h_n are finite in number or not (like in Hervé (1992, 1995)) and it generalizes easily to the multidimensional case. Like most of the other approaches, our results hinge on the computation of the spectral radius of a particular operator (see sections 3 and 4 below). We introduce a different space on which this operator acts however, and we use a computation of the Fredholm determinant borrowed from Ruelle (1976) to compute its spectral radius.

This paper is organized as follows. In Section 2, we recall basic facts on trace-class operators and Fredholm determinants. In Section 3, we specialize these results to the particular case of transfer operators. In Section 4 we show how the spectral radius of such a transfer operator can be used to compute s_p. Although the connection between the spectral radius of a transfer operator and the Sobolev regularity index s_p is not new (essentially all the computations of s_2 or s_p to which we referred above, rely on such a connection), we present an essentially self-contained discussion in Section 4, for the sake of convenience, and also because we require some results tailored to our different framework. Next, we need to compute the spectral radius of the transfer operator, for which different techniques can be used. We propose three different procedures. The first two are similar (but not identical) to the formula in Hervé (1992, 1995), and have the same type of convergence: one can prove that the estimate ρ_n at the n-th step approaches the true ρ exponentially fast, $|\rho - \rho_n| \leq Ce^{-\alpha n}$, but C and α are unknown. The third procedure, on which we concentrate almost exclusively, requires more computational work (*i.e.* a longer program), but at every step n we have an explicit upper bound on $|\rho - \rho_n|$. In this procedure the spectral radius is found by determining the zero with the smallest absolute value of the corresponding Fredholm determinant. This determinant is an entire function which in practice needs to be truncated to its first order terms to derive numerical estimates. Section 5 shows how this is done, how the rest term can be controlled and how this translates into error estimates on our computation of s_p. In Section 6, we present many examples, in one as well as two dimensions, with finitely many as well as infinitely many nonzero h_n. In particular we apply our method to the easily implementable orthonormal wavelet filters recently introduced in Herley and Vetterli (1992) (which have an infinite number of nonvanishing h_n). Finally, in Section 7, we discuss whether and how the technique proposed here can be extended to a direct computation of the Hölder exponent μ.

There is clearly some similarity of our results with those of Hervé (1992, 1995), the only other work in the literature so far that can deal with infinitely many nonvanishing h_n. Most of this work was carried out in the summer of 1992. At the time we were not aware yet of the results in Hervé (1992, 1995). Nevertheless, Hervé (1992, 1995) clearly has priority, since he proved his results earlier. Moreover, we initially developed our results only for s_2; after reading a preprint of Hervé (1995), we saw the interest of computing also s_p for $p \neq 2$, and we extended our results to this case. If $p \neq 2$ then our method only works if we impose an extra technical condition which Hervé (1995) does not require; in this sense our results are weaker than Hervé's. On the other hand, our method has the new feature that it leads to explicit control over the error in the algorithm, as mentioned above, and as will be discussed in detail in Section 5.

2. Fredholm determinants of trace-class operators.

The most general treatment of Fredholm determinants is within the framework of Banach spaces, as in *e.g.* Grothendieck (1956). For the present paper, it suffices to work in an appropriate Hilbert space setting, where everything can be formulated and proved more simply. This section presents the results that will be needed in the sequel of the paper.

In this section we denote our (generic) Hilbert space by $\mathcal{H}$; we assume that $\mathcal{H}$ is separable. Recall that any bounded operator A on $\mathcal{H}$ can be written as $A = U(A^*A)^{1/2} = U|A|$, where U is a partial isometry with $\|Ux\| = \|x\|$ if $x \in \overline{\mathrm{Ran}\,|A|}$, $Ux = 0$ if $x \perp \mathrm{Ran}\,|A|$. If A is a compact operator, then so is $|A|$; the spectrum of $|A|$ then consists of a decreasing sequence of nonnegative eigenvalues. The strictly positive eigenvalues λ_m of $|A|$ are called the singular values of A. If φ_m is a corresponding orthonormal system of eigenvectors of $|A|$, and we define another orthonormal system ψ_m by $\psi_m = U\varphi_m$, then we have the following representation for A:

$$Ax = \sum_m \lambda_m \langle x, \varphi_m \rangle \psi_m \,. \tag{2.1}$$

Another useful representation of compact operators is obtained as follows. The spectrum of A itself also consists of a sequence of discrete eigenvalues α_n, which accumulate only at 0. For any $\alpha_n \neq 0$ we define its (algebraic) multiplicity d_n by $d_n = \max_{k\geq 1} \dim(\operatorname{Ker}(A - \alpha_n \operatorname{Id})^k)$, which is necessarily finite because A is compact. On $\operatorname{Ker}(A - \alpha_n \operatorname{Id})^k$ we can then choose a suitable basis in which A is upper triangular (one can, for instance, reduce it to its Jordan normal form on this subspace). The union of all the different and independent bases constructed in this way for the distinct eigenvalues spans a closed subspace $\mathcal{H}_1$, which is invariant for A. By Gram-Schmidt orthogonalizing this basis we obtain an orthonormal basis $\{u_k\}$ for $\mathcal{H}_1$ in which $A|_{\mathcal{H}_1}$ is triangular:

$$\langle Au_k, u_\ell\rangle = 0\,, \qquad \text{if } \ell > k\,, \tag{2.2}$$

$$\langle Au_k, u_k\rangle = \alpha_k\,, \tag{2.3}$$

where the α_k now occur with their multiplicity. It is immediately clear that the same basis also gives a triangular representation for all the powers $A^m|_{\mathcal{H}_1}$, $m \geq 1$, with diagonal elements

$$\langle A^m u_k, u_k\rangle = \alpha_k^m\,. \tag{2.4}$$

Let us now assume that our compact operator is in fact trace class. This means that A satisfies one of the following two equivalent statements:

$$\sum_m \lambda_m < \infty$$

or

$$\sum_n |\langle Ae_n, e_n\rangle| < \infty\,, \qquad \text{for all orthonormal systems } \{e_n\}\,.$$

If A is trace-class, then the sum of the series $\sum_n \langle Ae_n, e_n\rangle$ is independent of the choice of the $\{e_n\}$, and is called the trace of A:

$$\operatorname{Tr} A = \sum_n \langle Ae_n, e_n\rangle\,.$$

In particular, using the representation (2.1) above, one has

$$\operatorname{Tr}|A| = \sum_m \langle |A|\varphi_m, \varphi_m\rangle = \sum_m \lambda_m\,.$$

In Lidskii (1959) it is proved that for any trace-class operator A in a Hilbert space $\mathcal{H}$ one has

$$\operatorname{Tr} A = \sum_n \alpha_n \,, \tag{2.6}$$

where, as before, the α_n are the non-zero eigenvalues of A, taken with their algebraic multiplicity. This theorem, trivial in finite-dimensional spaces, is far less so in infinite dimensions. It is immediately clear from (2.2), (2.3) that the trace of $A|_{\mathcal{H}_1}$ is exactly $\sum_n \alpha_n$; the problem in proving (2.6) is to show that on the orthogonal complement $\mathcal{H}_0$ of $\mathcal{H}_1$, the operator $\operatorname{Proj}_{\mathcal{H}_0} A \operatorname{Proj}_{\mathcal{H}_0}$, which is of course trace-class too, and has spectrum consisting of only the point zero, has trace equal to zero. See *e.g.* Simon (1979) or Gohberg and Krein (1969) for other proofs of (2.6) than Lidskii's original proof.

The trace class operators form an ideal in the algebra of bounded operators: the product of a trace class operator and a bounded operator is again trace class. Consequently all the powers A^n, $n \geq 1$, of a trace class operator are trace class as well. Since $\operatorname{Proj}_{\mathcal{H}_0} A^n \operatorname{Proj}_{\mathcal{H}_0} = (\operatorname{Proj}_{\mathcal{H}_0} A \operatorname{Proj}_{\mathcal{H}_0})^n$ also has trace zero, we then see from (2.4) that

$$\operatorname{Tr} A^n = \sum_k \alpha_k^n \,. \tag{2.7}$$

(This can of course also be derived directly from Lidskii's theorem (2.6) and spectrum$(A^n) = \{\lambda^n : \lambda \in \text{spectrum } (A)\}$.) Finally, note that

$$\begin{aligned} \sum_k |\alpha_k| &= \sum_k |\langle A u_k, u_k\rangle| \\ &\leq \sum_{k,m} \lambda_m \, |\langle u_k, \varphi_m\rangle| \, |\langle \psi_m, u_k\rangle| \\ &\leq \sum_m \lambda_m \Big(\sum_k |\langle u_k, \varphi_m\rangle|^2\Big)^{1/2} \Big(\sum_k |\langle \psi_m, u_k\rangle|^2\Big)^{1/2} \\ &\leq \sum_m \lambda_m \, \|\varphi_m\| \, \|\psi_m\| = \sum_m \lambda_m = \operatorname{Tr} |A| \,. \end{aligned} \tag{2.8}$$

The Fredholm determinant of A is defined as

$$D_A(z) = \det(I - zA) = \prod_{n=1}^{+\infty} (1 - z\,\alpha_n) \,, \tag{2.9}$$

where the α_n occur with their multiplicity. Because $\sum_n |\alpha_n| < +\infty$, $D_A(z)$ is an entire function with its zeros exactly at the α_n^{-1}, with the same multiplicity. In particular, the spectral radius ρ_A of A is given by

$$\rho_A = (\min\{|z_0| : \ D_A(z_0) = 0\})^{-1} . \tag{2.10}$$

For sufficiently small values of z (*e.g.* for $|z| < \rho_A^{-1}$), one can rewrite $D_A(z)$ as follows:

$$\begin{aligned} D_A(z) &= \exp\Big(\sum_{n=1}^{\infty} \log(1 - z\alpha_n)\Big) \\ &= \exp\Big(\sum_{n=1}^{\infty}\sum_{m=1}^{\infty} -\frac{1}{m}(z\alpha_n)^m\Big) \\ &= \exp\Big(-\sum_{m=1}^{\infty} \frac{1}{m} z^m \operatorname{Tr} A^m\Big), \end{aligned} \tag{2.11}$$

a formula which already shows that D_A is completely determined by the traces $\operatorname{Tr} A^m$. Expanding the infinite product in (2.9) leads to a different formula for $D_A(z)$,

$$D_A(z) = 1 + \sum_{k=1}^{+\infty} z^k \gamma_k , \tag{2.12}$$

with

$$\gamma_k = \sum_{l_1 < l_2 < \cdots < l_k} \alpha_{l_1} \cdots \alpha_{l_k} . \tag{2.13}$$

We then have

$$|\gamma_k| \le \frac{1}{k!}\Big(\sum_l |\alpha_l|\Big)^k \le \frac{1}{k!}(\operatorname{Tr}|A|)^k . \tag{2.13}$$

It follows that we can therefore always write

$$D_A(z) = D_A^N(z) + R_A^N(z), \tag{2.14}$$

where $D_A^N(z)$ is the Taylor series for D_A truncated after the term in z^N, and

$$|R_A^N(z)| \le \sum_{k=N+1}^{+\infty} \frac{1}{k!}(\operatorname{Tr}|A|)^k |z|^k . \tag{2.15}$$

We shall use this estimate to find the smallest zero of D_A: since D_A^N is a polynomial, its smallest zero can be found by a host of different numerical methods, and the control we have via (2.15) on the rest term R_A^N will tell us that the smallest zero of D_A itself cannot be far from that of D_A^N if N is sufficiently large (see Section 5).

In order to identify the zeros of D_A^N, we need again a different representation; in particular, we are interested in a way of computing the Taylor coefficients of D_A which does not require knowledge of the eigenvalues α_n. To do this, let us start by restricting ourselves to the disk $B\left(0, \rho_A^{-1}\right)$. On this disk, we can write (using a trick going back to Newton)

$$\begin{aligned}\sum_{k=0}^{\infty}(k+1)\,\gamma_{k+1}\,z^k = D'_A(z) &= \sum_{n=1}^{\infty}(-\alpha_n)\prod_{\substack{m=1\\ m\neq n}}^{\infty}(1-\alpha_m z)\\ &= -D_A(z)\sum_{n=1}^{\infty}\frac{\alpha_n}{1-\alpha_n z}\\ &= -D_A(z)\sum_{n=1}^{\infty}\sum_{k=0}^{\infty}\alpha_n^{k+1} z^k\\ &= -D_A(z)\sum_{k=0}^{\infty}\operatorname{Tr}A^{k+1}\,z^k\\ &= -\sum_{r=0}^{\infty}z^r\sum_{m=0}^{r}\gamma_m\operatorname{Tr}A^{r-m+1}\,,\end{aligned}$$

where the reordering of the sums is allowed because the series converges absolutely for $|z| < (\rho_A)^{-1}$, and where we have introduced $\gamma_0 = 1$. It follows that

$$\gamma_{k+1} = -\frac{1}{k+1}\sum_{m=0}^{k}\gamma_m\operatorname{Tr}A^{k+1-m}\,. \tag{2.16}$$

(This derivation is in fact the standard way of relating the elementary symmetric functions $\sum_{i_1<i_2<\cdots<i_r}\alpha_{i_1}\cdots\alpha_{i_r}$ with the power sums $\sum_i \alpha_i^r$; see Macdonald (1979), p. 12-16.)

We now see the outline of a program that will be used below: given the $\operatorname{Tr}A^m$ for a trace-class operator A, we now know how to find, via the smallest zero of the Fredholm determinant, the spectral radius of A and how to control the error. In the next section, we introduce specific operators A in specific Hilbert spaces, we verify that they are trace-class, and we show how to compute $\operatorname{Tr} A^m$.

3. A special case: transfer operators.

The operators to which we shall apply the results in the previous section act on 2π-periodic functions $f(\omega)$ and are defined by

$$(\mathcal{L}_w f)(\omega) = w\left(\frac{\omega}{2}\right) f\left(\frac{\omega}{2}\right) + w\left(\frac{\omega}{2}+\pi\right) f\left(\frac{\omega}{2}+\pi\right), \tag{3.1}$$

where w is a 2π-periodic weight function, for which several concrete choices will be proposed in Section 4. We shall say that $\mathcal{L}_w$ is the transfer operator associated with the function w; these operators are also called Perron-Frobenius operators or transition operators in the literature. We shall always assume that the Fourier coefficients of w decay exponentially; that is,

$$w(\omega) = \sum_n w_n \, e^{-in\omega} \tag{3.2}$$

and

$$|w_n| \le C \, e^{-\gamma |n|}, \tag{3.3}$$

for some $C, \gamma > 0$. In terms of the Fourier coefficients f_n of $f(\omega) = \sum_n f_n \, e^{-in\omega}$, (3.1) can also be rewritten as

$$(\mathcal{L}_w f)_n = \frac{1}{2\pi} \int_0^{2\pi} (\mathcal{L}_w f)(\omega) \, e^{in\omega} \, d\omega = 2 \sum_k w_{2n-k} \, f_k \, . \tag{3.4}$$

When no confusion is possible, we shall often drop the subscript w on $\mathcal{L}_w$.

Operators of the type (3.1) can be studied in many different function spaces. They have been linked with the study of refinable functions before; see *e.g.* Conze (1989), Eirola (1992), Villemoes (1992), Hervé (1995). They are special cases of the operators in Ruelle (1976, 1990). In this section we discuss their properties on some Hilbert spaces of analytic functions; in Section 7 we shall come back to their action on other, larger function spaces.

As candidates for the space $\mathcal{H}$ we define

$$E_\alpha = \Big\{ f \ 2\pi\text{-periodic} : f(\omega) = \sum_n f_n \, e^{-in\xi} \ \text{ and } \quad \|f\|_\alpha^2 = \sum_n |f_n|^2 \, e^{2|n|\alpha} < \infty \Big\} . \tag{3.5}$$

(Note that these are different from the spaces E^α in Hervé (1995).) The E_α are Hilbert spaces of analytic functions ($f \in E_\alpha$ can be extended to complex $\omega = \omega_1 + i\omega_2$ and is then analytic for $|\omega_2| = |\mathrm{Im}\,\omega| < \alpha$); their inner product is given by

$$\langle f, g \rangle_\alpha = \sum_n f_n \, \overline{g}_n \, e^{2|n|\alpha} . \tag{3.6}$$

Note that for each α, the constant function 1 is in E_α; moreover, for $f \in E_\alpha$,

$$\langle f, 1 \rangle_\alpha = f_0 = \frac{1}{2\pi} \int_0^{2\pi} f(\omega) \, d\omega . \tag{3.7}$$

The functions

$$e_{n,\alpha}(\omega) = e^{-|n|\alpha} \, e^{-in\omega} \tag{3.8}$$

constitute an orthonormal basis in E_α.

In order to be able to apply Section 2 to $\mathcal{L}$ and E_α, we need to verify: 1) that $\mathcal{L}$ is a bounded operator on E_α, 2) that $\mathcal{L}$ is trace class on E_α. We start by computing $\|\mathcal{L}f\|_\alpha$, using (3.4):

$$\begin{aligned}
\|\mathcal{L}f\|_\alpha^2 &= \sum_n |(\mathcal{L}f)_n|^2 \, e^{2|n|\alpha} \\
&= 4 \sum_n \Big| \sum_k w_{2n-k} \, f_k \Big|^2 \, e^{2|n|\alpha} \\
&\leq 4 \, \|f\|_\alpha^2 \sum_{n,k} |w_{2n-k}|^2 \, e^{-2|k|\alpha} \, e^{2|n|\alpha} \\
&\leq 4 \, C^2 \, \|f\|_\alpha^2 \sum_{n,k} e^{-2(\alpha-\gamma)|k|} \, e^{-2(2\gamma-\alpha)|n|} ,
\end{aligned}$$

where the last inequality used (3.3) and $|2n - k| \geq 2|n| - |k|$. It follows that $\mathcal{L}$ is a bounded operator on E_α if $\gamma < \alpha < 2\gamma$. We can us the same estimate to bound the matrix elements of $\mathcal{L}$ with respect to the orthonormal basis (3.8):

$$
\begin{aligned}
|\langle \mathcal{L}e_{k,\alpha}, e_{n,\alpha}\rangle_\alpha| &= |2\, w_{2n-k}\, e^{|n|\alpha}\, e^{-|k|\alpha}| \\
&\leq 2\,C\, e^{-(\alpha-\gamma)|k|}\, e^{-(2\gamma-\alpha)|n|}\,.
\end{aligned} \tag{3.9}
$$

This implies that $\mathcal{L}$ is trace class on E_α if $\gamma < \alpha < 2\gamma$. By a simple application of Cauchy-Schwarz, we have indeed, for any orthonormal system u_m in E_α,

$$
\begin{aligned}
\sum_m |\langle \mathcal{L}u_m, u_m\rangle_\alpha| &\leq \sum_{m,k,n} |\langle u_m, e_{k,\alpha}\rangle_\alpha|\, |\langle \mathcal{L}e_{k,\alpha}, e_{n,\alpha}\rangle_\alpha|\, |\langle e_{n,\alpha}, u_m\rangle_\alpha| \\
&\leq \sum_{k,n} |\langle \mathcal{L}e_{k,\alpha}, e_{n,\alpha}\rangle_\alpha| < \infty\,.
\end{aligned} \tag{3.10}
$$

Since $\mathcal{L}$ is trace class, it has a representation of type (2.1) with $\sum_m \lambda_m < \infty$; in fact, by the same argument as in (3.10), we have

$$
\operatorname{Tr}|\mathcal{L}| = \sum_m \lambda_m = \sum_m \langle \mathcal{L}\varphi_m, \psi_m\rangle_\alpha \leq \sum_{k,n} |\langle \mathcal{L}e_{k,\alpha}, e_{n,\alpha}\rangle_\alpha|\,.
$$

This leads to a bound on the sum of the absolute values $|\alpha_n|$ of the eigenvalues of $\mathcal{L}$:

$$
\sum_n |\alpha_n| \leq \operatorname{Tr}|\mathcal{L}| \leq 2\,C\Big(\sum_{k\in\mathbb{Z}} e^{-(\alpha-\gamma)|k|}\Big)\Big(\sum_{n\in\mathbb{Z}} e^{-(2\gamma-\alpha)|n|}\Big)\,. \tag{3.11}
$$

We shall need this bound in order to control the rest term when we try to locate the smallest zero of the Fredholm determinant after truncation (see (2.15) and Section 5).

We are therefore in a position to apply the results of Section 2 to $\mathcal{L}$ on E_α; in the next section we shall see how this will then help to determine s_p. In order to be of practical use however, we need to be able to compute the Taylor coefficients of the Fredholm determinant $D_{\mathcal{L}}$ explicitly, and for this, we need the traces $\operatorname{Tr}\mathcal{L}^m$ (see (2.15)). Let us therefore now concentrate on their computation. As a warmup, let us compute $\operatorname{Tr}\mathcal{L}$ itself. We have

$$
\operatorname{Tr}\mathcal{L} = \sum_n \langle \mathcal{L}e_{n,\alpha}, e_{n,\alpha}\rangle_\alpha
$$

$$
\begin{aligned}
&= \sum_n (\mathcal{L}e_{n,\alpha})_n \, e^{|n|\alpha} \\
&= \sum_n (\mathcal{L}e^{-in\,\cdot})_n \qquad (3.12)\\
&= \frac{1}{2\pi} \sum_n \int_{-\pi}^{\pi} e^{in\omega} (\mathcal{L}e^{-in\,\cdot})(\omega)\, d\omega \,.
\end{aligned}
$$

To compute integrals of this type, we shall use the following standard lemma, which crops up in any study using these operators (see *e.g.* Cohen (1990), Eirola (1992), Villemoes (1992), Gripenberg (1992)). For the sake of completeness, we include its short proof.

Lemma 3.1. *Let w be a 2π-periodic function, and let $\mathcal{L}$ be defined as in (3.1). Then, for any $k > 0$ and any f, g 2π-periodic functions, we have*

$$
\begin{aligned}
\int_{-\pi}^{\pi} f(\omega)\, (\mathcal{L}^k g)(\omega)\, d\omega &= \int_{-2^k\pi}^{2^k\pi} f(\omega) \Big(\prod_{\ell=1}^{k} w(2^{-\ell}\omega) \Big)\, g(2^{-k}\omega)\, d\omega \\
&= 2^k \int_{-\pi}^{\pi} f(2^k\omega) \Big(\prod_{m=0}^{k-1} w(2^m\omega) \Big)\, g(\omega)\, d\omega \,. \qquad (3.13)
\end{aligned}
$$

PROOF. By induction. For $k = 1$ we have

$$
\begin{aligned}
\int_{-\pi}^{\pi} f(\omega)\, (\mathcal{L}g)(\omega)\, d\omega &= \int_{-\pi}^{\pi} f(\omega) \Big(w\big(\frac{\omega}{2}\big)\, g\big(\frac{\omega}{2}\big) + w\big(\frac{\omega}{2} + \pi\big)\, g\big(\frac{\omega}{2} + \pi\big) \Big)\, d\omega \\
&= 2 \int_{-\pi/2}^{\pi/2} f(2\omega)\, (w(\omega)\, g(\omega) + w(\omega + \pi)\, g(\omega + \pi))\, d\omega \\
&= 2 \int_{-\pi/2}^{3\pi/2} f(2\omega)\, w(\omega)\, g(\omega)\, d\omega \\
&= 2 \int_{-\pi}^{\pi} f(2\omega)\, w(\omega)\, g(\omega)\, d\omega \\
&= \int_{-2\pi}^{2\pi} f(\omega)\, w\big(\frac{\omega}{2}\big)\, g\big(\frac{\omega}{2}\big)\, d\omega \,.
\end{aligned}
$$

If we now assume that (3.13) holds for k, then

$$
\int_{-\pi}^{\pi} f(\omega)\, (\mathcal{L}^{k+1} g)(\omega)\, d\omega
$$

$$= 2^k \int_{-\pi}^{\pi} f(2^k\omega) \Big(\prod_{m=0}^{k-1} w(2^m\omega) \Big) \Big(w(\frac{\omega}{2})\, g(\frac{\omega}{2}) + w(\frac{\omega}{2}+\pi)\, g(\frac{\omega}{2}+\pi) \Big)\, d\omega$$

$$= 2^{k+1} \int_{-\pi/2}^{\pi/2} f(2^{k+1}\omega) \Big(\prod_{m=1}^{k} w(2^m\omega) \Big) \Big(w(\omega)\, g(\omega) + w(\omega+\pi)\, g(\omega+\pi) \Big)\, d\omega$$

$$= 2^{k+1} \int_{-\pi}^{\pi} f(2^{k+1}\omega) \Big(\prod_{m=1}^{k} w(2^m\omega) \Big)\, g(\omega)\, d\omega\,.$$

Applying this to (3.12), we find

$$\operatorname{Tr} \mathcal{L} = \frac{1}{2\pi} \sum_n 2 \int_{-\pi}^{\pi} e^{2in\omega}\, w(\omega)\, e^{-in\omega}\, d\omega = 2 \sum_n w_n = 2\, w(0)\,.$$

Similarly, we can compute, for any $k \geq 1$,

$$\begin{aligned} \operatorname{Tr} \mathcal{L}^k &= \sum_n (\mathcal{L}^k e^{-in\,\cdot})_n \\ &= \frac{1}{2\pi} \sum_n \int_{-\pi}^{\pi} e^{in\omega}\, (\mathcal{L}^k e^{-in\,\cdot})(\omega)\, d\omega \\ &= 2^k \sum_n \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(2^k-1)n\omega} \Big(\prod_{m=0}^{k-1} w(2^m\omega) \Big)\, d\omega \\ &= 2^k \sum_n (W_k)_{(2^k-1)n}\,, \end{aligned} \tag{3.14}$$

where the 2π-periodic function W_k is defined by

$$W_k(\omega) = \prod_{m=0}^{k-1} w(2^m\omega)\,,$$

and where its Fourier coefficients are denoted by $(W_k)_\ell$, as usual. Sums of the type (3.14) can be computed by means of the following lemma, which is essentially a version of the Poisson summation formula.

Lemma 3.2. *Let f be a 2π-periodic function, and f_n its Fourier coefficients. Assume that $\sum_n |f_n| < \infty$. Then, for any $\ell \geq 1$,*

$$\sum_{m=0}^{\ell-1} f\Big(\frac{2\pi m}{\ell}\Big) = \ell \sum_{n\in\mathbb{Z}} f_{\ell n}\,. \tag{3.15}$$

PROOF.

$$\sum_{m=0}^{\ell-1} f(\frac{2\pi m}{\ell}) = \sum_{m=0}^{\ell-1} \sum_{n\in\mathbb{Z}} f_n\, e^{-2\pi i m n/\ell}. \tag{3.16}$$

If $n \in \ell\mathbb{Z}$, then

$$\sum_{m=0}^{\ell-1} e^{-2\pi i m n/\ell} = \sum_{m=0}^{\ell-1} 1 = \ell. \tag{3.17}$$

If $n \in \ell\mathbb{Z} + k$, with $0 < k < \ell$, then

$$\sum_{m=0}^{\ell-1} e^{-2\pi i m n/\ell} = \sum_{m=0}^{\ell-1} e^{-2\pi i k m/\ell} = \frac{\left(e^{-2\pi i k/\ell}\right)^{\ell} - 1}{e^{-2\pi i k/\ell} - 1} = 0. \tag{3.18}$$

since (3.16) is absolutely summable, we may change the order of summations, and (3.15) then follows from (3.17) and (3.18).

This can then be used to give an explicit formula for $\operatorname{Tr} \mathcal{L}^k$. The following theorem summarizes the findings of this section so far.

Theorem 3.3. *Let $w(\omega)$ be a 2π-periodic function with Fourier coefficients satisfying (3.3). Define $\mathcal{L}$ to be the corresponding transfer operator, as in (3.1), and let E_α be the Hilbert spaces defined by (3.5). If $\gamma < \alpha < 2\gamma$, then $\mathcal{L}$ is a trace class operator on E_α. The spectrum of $\mathcal{L}$ does not depend on the choice of α in $]\gamma, 2\gamma[$, and for any $k \geq 1$, the traces $\operatorname{Tr} \mathcal{L}^k$ are given by the explicit formula*

$$\operatorname{Tr} \mathcal{L}^k = \frac{2^k}{2^k - 1} \sum_{m=0}^{2^{k-2}} \Big(\prod_{\ell=0}^{k-1} w(2^\ell \frac{2\pi m}{2^k - 1}) \Big). \tag{3.19}$$

PROOF. Most of the assertions were proved in our discussion above. Formula (3.19) is a direct consequence of applying Lemma 3.2 to (3.14). Since the Taylor coefficients of the Fredholm determinant $D_{\mathcal{L}}(z)$ are completely determined by the traces $\operatorname{Tr} \mathcal{L}^k$ (see Section 2), and the zeros of $D_{\mathcal{L}}$ are the inverses of the eigenvalues of $\mathcal{L}$, the fact that (3.19) does not depend on α immediately implies that the spectrum of $\mathcal{L}$ doesn't either.

REMARK. Another way of obtaining (3.19) is the following. The operator $\mathcal{L}_w$ can also be viewed as an integral operator,

$$(\mathcal{L}_w f)(w) = 2\int_{-\pi}^{\pi} w(\omega')\,\delta(\tau\omega' - \omega)\,f(\omega')\,d\omega'\,,$$

where τ on $]-\pi,\pi]$ is multiplication by 2, modulo 2π. The trace of $\mathcal{L}_w$ can then be obtained by integrating the kernel $\mathcal{K}_w(\omega,\omega') = w(\omega')\,\delta(\tau\omega' - \omega)$ along the diagonal $\omega = \omega'$,

$$\operatorname{Tr}\mathcal{L}_w = 2\int_{-\pi}^{\pi} w(\omega)\,\delta(\tau\omega - \omega)\,d\omega = 2\,w(0)\,.$$

The integral kernel of $(\mathcal{L}_w)^k$ is given by

$$\int\cdots\int \mathcal{K}_w(\omega,\omega_1)\,\mathcal{K}_w(\omega_1,\omega_2)\cdots\mathcal{K}_w(\omega_{k-1},\omega')\,d\omega_1\cdots d\omega_{k-1}\;;$$

restricting this to $\omega = \omega'$ and integrating over ω leads to a delta-function $\delta(\tau^k\omega - \omega)$. This results in a sum of different contributions in the fixed points of τ^k (*i.e.* the points $2\pi m/(2^k-1)$), as in (3.19); the denominator $(2^k - 1)$ multiplying these contributions results from the Jacobian of $\tau^k - \mathrm{Id}$.

In the next section, we shall see how, for a judicious choice of w, the operator $\mathcal{L}$ and in particular its spectral radius on E_α can be used to compute the Sobolev exponent s_p of a refinable function. We shall be interested in multidimensional refinable functions as well. We will then need a slight generalization of the constructions above. Instead of (3.1), we have then, for $\omega \in [-\pi,\pi]^d$,

$$(\mathcal{L}f)(\omega) = \sum_{j=0}^{|\det D|-1} w(D^{-1}\omega + \xi_j)\,f(D^{-1}\omega + \xi_j)\,, \tag{3.20}$$

where D is a $d\times d$ matrix with integer entries and all its eigenvalues strictly larger than 1 in absolute value, and where f, w are functions in d variables, 2π-periodic in each (*i.e.* they are functions on the torus $\mathbb{T}^d$); the ξ_j are defined by $\xi_j = D^{-1}\zeta_j$, where the ζ_j are distinct elements in $2\pi\mathbb{Z}^d/D\mathbb{Z}^d$, so that $D^{-1}\omega + \xi_j$ are exactly the $|\det D|$ distinct pre-images of ω under the map D. In the case where $d = 1$ and D is multiplication by 2, (3.20) obviously reduces to (3.1). (3.20) can also be rewritten as

$$(\mathcal{L}f)_n = |\det D| \sum_k w_{Dn-k}\, f_k \,, \tag{3.21}$$

where f_k again denotes the k-th Fourier coefficient of f; the summation index k now ranges over $\mathbb{Z}^d$. As before we shall assume

$$|w_n| \leq C\, e^{-\gamma|n|}\,, \tag{3.22}$$

with $|n| = (n_1^2 + \cdots + n_d^2)^{1/2}$; the space E_α is then defined by

$$E_\alpha = \Big\{ f \text{ function on } [0,2\pi]^d : \ \|f\|_\alpha^2 = \sum_{n\in\mathbb{Z}^d} |f_n|^2\, e^{2|n|\alpha} < \infty \Big\}. \tag{3.23}$$

Repeating the same arguments as before, one finds then the following generalization of Lemma 3.2:

Theorem 3.4. *Let D be a $d\times d$ matrix with integer entries and with all its eigenvalues strictly larger than 1 in absolute value. Define the spaces E_α and the operator $\mathcal{L}$ as in (3.21)-(3.23). Then $\mathcal{L}$ is trace class on E_α if $\gamma < \alpha < r_D\gamma$, where $r_D = \min\{|\lambda| :\ \lambda$ is eigenvalue of $D\} > 1$, and the spectrum of $\mathcal{L}$ on E_α does not depend on α.*

Define now, for any $k \geq 1$, the set F_k by

$$F_k = \{\eta \in\,]-\pi,\pi[^d\colon\ D^k\eta - \eta \in 2\pi\mathbb{Z}^d\}\,; \tag{3.24}$$

this set has exactly $|\det(D^k - \mathrm{Id})|$ elements, which are the fixed points in $\mathbb{T}^d$ of D^k. Then the traces $\mathrm{Tr}\,\mathcal{L}^k$ are given by the following explicit formula:

$$\mathrm{Tr}\,\mathcal{L}^k = \frac{|\det D|^k}{|\det(D^k - \mathrm{Id})|} \sum_{\eta\in F_k} \Big(\prod_{m=0}^{k-1} w(D^m\eta) \Big)\,. \tag{3.25}$$

The proof is exactly along the same lines as in the one-dimensional case, and we leave it to the reader to fill in the details. In order to obtain the explicit formula (3.25) one needs the following higher dimensional generalization of Lemma 3.2:

Lemma 3.5. *Let L be an integer $d \times d$ matrix, and define the set R by*

$$R = \{\zeta \in]-\pi, \pi]^d : \ L\zeta \in 2\pi\mathbb{Z}^d\}.$$

Then for any function f on $\mathbb{T}^d$ such that $\sum_{n\in\mathbb{Z}^d} |f_n|$ is finite,

$$\sum_{n\in\mathbb{Z}^d} f_{L^t n} = \frac{1}{|\det L|} \sum_{\zeta\in R} f(\zeta). \tag{3.26}$$

PROOF. As for Lemma 3.2, the proof hinges on the computation of $\sum_{\zeta\in R} e^{-in\cdot\zeta}$, where $n \in \mathbb{Z}^d$ and $n \cdot \zeta = n_1\zeta_1 + \cdots + n_d\zeta_d$. This can be done by a standard argument on character sums. The set R is an additive group isomorphic with $\mathbb{Z}^d/L\mathbb{Z}^d$; it follows that

$$\Big(\sum_{\zeta\in R} e^{-in\cdot\zeta}\Big)^2 = \sum_{\zeta_1,\zeta_2\in R} e^{-in\cdot(\zeta_1-\zeta_2)} = (\#R)\sum_{\zeta\in R} e^{-in\cdot\zeta}.$$

Consequently we have either $\sum_{\zeta\in R} e^{-in\cdot\zeta} = \#R = |\det L|$ or $\sum_{\zeta\in R} e^{-in\cdot\zeta} = 0$. The former is possible only if each of the terms $e^{-in\cdot\zeta}$ equals 1, which is equivalent to the requirement $n = L^t k$ for some $k \in \mathbb{Z}^d$. Hence

$$\sum_{\zeta\in R} e^{-in\cdot\zeta} = \begin{cases} |\det L|, & \text{if } n \in L^t\mathbb{Z}^d, \\ 0, & \text{if } n \in \mathbb{Z}^d\backslash L^t\mathbb{Z}^d. \end{cases}$$

(3.26) then follows easily.

The following examples show how the explicit formulas for $\operatorname{Tr}\mathcal{L}^k$ can be used to determine the spectrum of $\mathcal{L}$ completely and explicitly in some simple cases.

EXAMPLE 3.6. $w(\omega) = ((1 + e^{-i\omega})/2)^N$; this corresponds to the choice $w(\omega) = m(\omega)$ for the case where the refinable function is a B-spline (see Section 1). Then $w(2^\ell 2\pi m/(2^k - 1)) = 1$, for all ℓ, if $m = 0$, and for $m \neq 0$ we have

$$\prod_{\ell=0}^{k-1} w(2^\ell \frac{2\pi m}{2^k-1}) = \exp(i\pi N m) \left(\prod_{\ell=0}^{k-1} \frac{\sin\left(2^\ell 2\pi \frac{m}{2^k-1}\right)}{2\sin\left(2^\ell \pi \frac{m}{2^k-1}\right)} \right)^N = 2^{-Nk}.$$

Consequently

$$\begin{aligned}\operatorname{Tr}\mathcal{L}^k &= \frac{1}{1-2^{-k}} + \frac{2^k-2}{1-2^{-k}}\,2^{-Nk} \\ &= \sum_{\ell=0}^{N-2} 2^{-k\ell} + 2\,2^{-k(N-1)}\,.\end{aligned} \tag{3.27}$$

By (2.11) it then follows that

$$\begin{aligned}D_{\mathcal{L}}(z) &= \left(\prod_{\ell=0}^{N-2}\exp\left(-\sum_{m=1}^{\infty}\frac{1}{m}(2^{-\ell}z)^m\right)\right)\exp\left(-2\sum_{m=1}^{\infty}\frac{1}{m}(2^{-N+1}z)^m\right) \\ &= (1-z)\left(1-\frac{z}{2}\right)\cdots\left(1-\frac{z}{2^{N-2}}\right)\left(1-\frac{z}{2^{N-1}}\right)^2,\end{aligned}$$

implying that the nonzero eigenvalues of $\mathcal{L}$ are $1, 1/2, \ldots, 1/2^{N-2}$ (with multiplicity 1) and $1/2^{N-1}$ (with multiplicity 2); this can in fact also be read off from (3.27).

EXAMPLE 3.7. $w(\omega) = ((1+e^{-i\omega})/2)^N\, w_1(\omega)$; this corresponds to a factorization which we almost always impose (see Section 4); we assume $w_1(0) = 1$. Then the same computations as in the previous example give

$$\prod_{\ell=0}^{k-1} w\left(2^\ell \frac{2\pi m}{2^k-1}\right) = \begin{cases} 1, & \text{if } m = 0\,, \\ 2^{-kN}\prod_{\ell=0}^{k-1} w_1\left(2^\ell\frac{2\pi m}{2^k-1}\right), & \text{if } m \neq 0\,. \end{cases}$$

This leads to

$$\operatorname{Tr}\mathcal{L}^k = \frac{1-2^{-kN}}{1-2^{-k}} + 2^{-kN}\operatorname{Tr}\mathcal{L}_1^k\,,$$

where $\mathcal{L}_1$ is the operator (3.1) with w replaced by w_1. By (2.11) we have therefore

$$D_{\mathcal{L}}(z) = (1-z)\left(1-\frac{z}{2}\right)\cdots\left(1-\frac{z}{2^{N-1}}\right)D_{\mathcal{L}_1}(2^{-N}z)\,.$$

The spectrum of $\mathcal{L}$ consists now of two parts: the eigenvalues $1, 1/2, \ldots,$ $1/2^{N-1}$, together with 2^{-N} spectrum $\mathcal{L}_1$.

The spectra for these simple $\mathcal{L}$ had been analyzed by other means before (see *e.g.* Daubechies and Lagarias (1992)), but their eigenvalues are recovered here in a particularly simple way. Let us consider a simple two-dimensional example next.

EXAMPLE 3.8. Take $d = 2$, $D = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$, and

$$w(\omega) = \left(\frac{a(D\omega)}{2\,a(\omega)}\right)^N w_1(\omega)\,,$$

where $a(\omega) \sim \omega_1^2 + \omega_2^2$ as $|\omega| \to 0$, $w_1(0) = 1$, and where we assume that a and w_1 both satisfy (3.22), with moreover $\min\{|a(\omega)| : \omega \in [-\pi,\pi]^2\} > 0$. Then w also satisfies a bound of the type (3.22). We have now, for $\eta \in F_k$ (defined by (3.24))

$$\prod_{m=0}^{k-1} w(D^k\eta) = \begin{cases} 1\,, & \text{if } \eta = 0\,, \\ 2^{-kN} \displaystyle\prod_{m=0}^{k-1} w_1(D^k\eta)\,, & \text{if } \eta \neq 0\,. \end{cases}$$

One easily checks that $|\det D| = 2$, $|\det(D^{2k} - \mathrm{Id})| = (2^k - 1)^2$ and $|\det(D^{2k+1} - \mathrm{Id})| = 2^{2k+1} - 1$. Consequently

$$\begin{aligned} D_{\mathcal{L}}(z) &= \exp\Big(-\sum_{m=1}^{\infty} \frac{z^{2m}}{2m} \frac{2^{2m}(1 - 2^{-2mN})}{(2^m - 1)^2}\Big) \\ &\quad \cdot \exp\Big(-\sum_{m=1}^{\infty} \frac{z^{2m+1}}{2m+1} \frac{2^{2m+1}(1 - e^{-(2m+1)N)})}{2^{2m+1} - 1}\Big) D_{\mathcal{L}_1}(2^{-N}z) \\ &= \exp\Big(-\sum_{m=1}^{\infty}\sum_{\ell=0}^{N}\sum_{n=0}^{\infty} \frac{z^{2m}}{2m}(1 + 2^{-mN})\,2^{-m(\ell+n)}\Big) \\ &\quad \cdot \exp\Big(-\sum_{m=1}^{\infty}\sum_{\ell=0}^{N} \frac{z^{2m+1}}{2m+1} 2^{-(2m+1)\ell}\Big) D_{\mathcal{L}_1}(2^{-N}z) \\ &= \Big(\prod_{\ell=0}^{N-1}(1 - 2^{-\ell}z)\Big)\Big(\prod_{\ell=0}^{2N-1}\prod_{\substack{n=0 \\ n\neq\ell}}^{\infty}(1 - 2^{-(\ell+n)}z^2)^{1/2}\Big) D_{\mathcal{L}_1}(2^{-N}z) \\ &= \Big(\prod_{\ell=0}^{N-1}(1 - 2^{-\ell}z)\Big)\Big(\prod_{k=0}^{\infty}(1 - 2^{-k}z^2)^{n(k)}\Big) D_{\mathcal{L}_1}(2^{-N}z)\,, \end{aligned}$$

where $n(k) = [(k+1)/2]$ if $k \leq 2N - 1$, $n(k) = N$ if $k \geq 2N - 1$. This shows that in addition to the eigenvalues $1, 1/2, \ldots, 1/2^{N-1}$, which were also found in Cohen and Daubechies (1993), we have infinitely many eigenvalues $\pm 2^{-k/2}$, as well as of course the spectrum of $\mathcal{L}_1$ multiplied by 2^{-N}.

4. Regularity estimates using the spectral radius of $\mathcal{L}$.

In this section, we discuss the relations between the Sobolev or Hölder (global) regularity of a refinable function and the spectral radius, in the previously described function spaces, of certain transfer operators that are associated to this function. More precisely, the study of these operators leads to an exact estimate of the L^p-Sobolev exponents s_p defined in the introduction.

Let $\varphi(x)$ be an L^1 solution of a refinement equation

$$\varphi(x) = 2 \sum_{n \in \mathbb{Z}} h_n \, \varphi(2x - n) \,. \tag{4.1}$$

We assume that the coefficients h_n are summable, and that φ is normalized in the sense that $\int \varphi = 1$. By integrating on both sides of (4.1), we obtain

$$\sum_n h_n = 1 \,. \tag{4.2}$$

Define the continuous function $m(\omega) = \sum_n h_n \, e^{-in\omega}$. In all that follows, we shall assume that $m(\omega)$ can be put in the factorized form

$$m(\omega) = \cos^N(\omega/2) \, q(\omega) \,, \tag{4.3}$$

where N is a strictly positive integer and $q(\omega)$ is a 2π-periodic function whose Fourier coefficients c_n satisfy a geometric decay estimate,

$$|c_n| \leq C \, e^{-\beta n} \,. \tag{4.4}$$

We shall often also impose that $q(\omega)$ does not vanish on $[0, 2\pi]$.

Consequently, $m(\omega)$ is a smooth 2π-periodic function and (4.2) indicates that $m(0) = 1$. By applying the Fourier transform to (4.1), one obtains

$$\hat{\varphi}(\omega) = m(\omega/2) \, \hat{\varphi}(\omega/2) \tag{4.5}$$

and by iteration $\hat{\varphi}(\omega)$ can be written as the pointwise convergent infinite product

$$\hat{\varphi}(\omega) = \prod_{k=1}^{+\infty} m(2^{-k}\omega).$$

Before proceeding further, we would like to make some comments about the assumptions made on the functions $m(\omega)$ and $q(\omega)$ and their Fourier coefficients.

The infinite product formula (4.6) indicates that the function φ is the limit of a stationary subdivision scheme, *i.e.* successive refinements of a Dirac sequence by means of interpolation: the sequence obtained at a given scale is filled with zeros at the mid-points and then convolved with the sequence h_n and multiplied by 2. It is clear that a necessary condition for the existence of a non-trivial limit is $m(0) = 1$.

Formula (4.6) only indicates, however, that this scheme converges "weakly", *i.e.* in the sense of tempered distributions. The limit φ itself may be a tempered distribution without any regularity. Clearly, for numerical applications, one is more interested in uniform (or strong) convergence of the subdivision scheme to a continuous or more regular limit function. It has been proved by Dyn and Levin (1990) that the limit function can be continuous only if $m(\pi)=0$. This justifies the factorization of $m(\omega)$ expressed in (4.3). In addition, strong convergence of the subdivision scheme follows if $\varphi \in C^0 \cap L^2$ and if there exists a compact set K congruent to $[-\pi, \pi]$ modulo 2π (*i.e.* $|K| = 2\pi$, and for any $\omega \in [-\pi, \pi]$ there exists $\omega' \in K$ so that $\omega - \omega' \in 2\pi\mathbb{Z}$), containing a neighborhood of 0, such that $\inf_{j \geq 1, \omega \in K} |m(2^{-j}\omega)| > 0$. (See Cohen and Ryan (1995).) When this holds, we shall say that m is of type C.

The next two lemmas give the decay of the Fourier coefficients of $|q(\omega)|^p$. Let us look first at the case $p = 2\ell$, $\ell \in \mathbb{N}$, $\ell > 0$.

Lemma 4.1. *The Fourier coefficients $\{c_{n,2\ell}\}_{n\in\mathbb{Z}}$ of $|q(\omega)|^{2\ell}$, with $\ell \in \mathbb{N}$, satisfy the estimate*

$$|c_{n,2\ell}| \leq C_{2\ell}\, e^{-\beta|n|}. \tag{4.7}$$

PROOF. We can write $q(\omega) = Q(e^{i\omega})$ where $Q(z) = \sum_{n\in\mathbb{Z}} c_n z^n$ is analytic on the ring $R_\beta = \{e^{-\beta} < |z| < e^{\beta}\}$. We can define the function $Q_2(z) = Q(z)\,Q(z^{-1})$ that coincides with $|q(\omega)|^2$ on the unit circle $z = e^{i\omega}$. It is also analytic on R_β, as are all its integer powers, $(Q_2(z))^\ell$. The exponential decay (4.7) then follows immediately.

For $p \neq 2\ell$ we need an extra nonvanishing condition on $q(\omega)$.

Lemma 4.2. *Suppose that $q(\omega)$ does not vanish on $[0, 2\pi]$. Then there exists γ in $]0, \beta]$ such that for all $p > 0$, the Fourier coefficients $\{c_{n,p}\}_{n\in\mathbb{Z}}$ of $|q(\omega)|^p$ satisfy the estimate*

$$|c_{n,p}| \leq C_p \, e^{-\gamma n} . \tag{4.8}$$

PROOF. Since $q(\omega)$ does not vanish, there exists γ in $]0, \beta]$ such that $Q_2(z)$ does not vanish on $R_\gamma = \{e^{-\gamma} < |z| < e^{\gamma}\}$ (with the same notations as in the proof of Lemma 4.1 above). On this narrower ring, it is possible to define a set of analytic functions by

$$Q_p(z) = \exp\left(\frac{p}{2} \log Q_2(z)\right) = \sum_{n\in\mathbb{Z}} c_{n,p} \, z^n . \tag{4.9}$$

These functions are equal to $|q(\omega)|^p$ on the unit circle; their analyticity on R_γ implies the estimate on the coefficients $c_{n,p}$.

From the results of the previous section, we thus know that the transfer operators associated to the functions $|q(\omega)|^p$ are trace class on E_α, for any $\alpha \in \,]\gamma, 2\gamma[$. By Lemma 4.1, the transfer operators associated to the functions $|m(\omega)|^p$ will be trace class on E_α for $p \in 2\mathbb{N}$, but not for general p: because of the Hölder singularity at $\omega = \pi$ of $|m(\omega)|^p$ for $p \notin 2\mathbb{N}$, the Fourier coefficients of these functions do not decay exponentially. We have however the following result:

Lemma 4.3. *Let $\mathcal{L}_p$ (respectively $\mathcal{L}'_p$) be the transfer operators associated with $|q(\omega)|^p$ (respectively $|m(\omega)|^p$). For any $\alpha \in \,]\gamma, 2\gamma[$, $\mathcal{L}'_p$ acts as a trace class operator on the space E'_α composed of the functions $g(\omega) = |\sin(\omega/2)|^{Np} f(\omega)$ with $f \in E_\alpha$, the norm of g in E'_α being identified to the norm of f in E_α. Moreover, if $f(\omega)$ is a continuous eigenfunction of $\mathcal{L}_p$ with eigenvalue λ, then $g(\omega) = |\sin(\omega/2)|^{Np} f(\omega)$ is a continuous eigenfunction of $\mathcal{L}'_p$ with eigenvalue $2^{-2Np}\lambda$.*

PROOF. It suffices to note that

$$\begin{aligned}\mathcal{L}'_p g(2\omega) &= |m(\omega)|^p\, g(\omega) + |m(\omega+\pi)|^p\, g(\omega+\pi)\\ &= |\sin(\omega/2)\cos(\omega/2)|^{Np}\,(|q(\omega)|^p\, f(\omega) + |q(\omega+\pi)|^p\, f(\omega+\pi))\\ &= 2^{-2Np}\,|\sin(\omega)|^{Np}\,\mathcal{L}_p f(2\omega).\end{aligned}$$

The operators $\mathcal{L}_p, \mathcal{L}'_p$ will be used to estimate the regularity of φ. In our proofs we shall use that $\mathcal{L}_p$ is a positive operator, in the sense that $(\mathcal{L}_p f)(\omega) \geq 0$ for all $\omega \in [-\pi,\pi]$ if $f(\omega) \geq 0$ for all $\omega \in [-\pi,\pi]$. Such operators have special spectral properties; see *e.g.* Schaefer (1966) or Schaefer (1974). To see how the general theorems on positive operators apply here, we first need to establish some facts about E_α. Define $E_\alpha^+ = \{f \in E_\alpha : \ f(\omega) \geq 0 \text{ for } \omega \in [-\pi,\pi]\}$. This is a cone in E_α, which contains in particular all the positive trigonometric polynomials. It follows that the closed linear span of E_α^+ equals E_α, or, in the terminology of Schaefer (1966), E_α is an ordered Banach space with total positive cone. It then already follows from the Krein-Rutman theorem (see *e.g.* Schaefer (1966), p. 265) that

Lemma 4.4. *The spectral radius r_p of $\mathcal{L}_p$ in E_α is an eigenvalue for $\mathcal{L}_p$ and there exists a positive eigenfunction for this eigenvalue.*

(The statement of the Krein-Rutman theorem in Schaefer (1965) is for real ordered Banach spaces, but since E_α can easily be seen to be the complexification of $\{f \in E_\alpha : \ f(\omega) \in \mathbb{R} \text{ for all } \omega \in [-\pi,\pi]\}$, the theorem still applies.)

More restrictions on the spectrum of $\mathcal{L}_p$ can be derived if q satisfies extra conditions. We shall need the following lemma:

Lemma 4.5. *Let $w(\omega)$ be a 2π-periodic function satisfying* (3.3). *Assume furthermore that $w(\omega) \geq 0$ for all $\omega \in [-\pi,\pi]$, $w(0) = 1$, $w(\pi) \neq 0$, and that w is of type C. Then, for all $f \in E_\alpha$ (with $\gamma < \alpha < 2\gamma$) with $f \geq 0$ and for all $\omega \in [-\pi,\pi]$, there exists $n \geq 1$ such that $(\mathcal{L}_w^n f)(\omega) > 0$.*

PROOF. 1) Since w is of type C, we can find a compact set K, congruent with $[-\pi,\pi]$ mod 2π, and a constant $C > 0$ so that, for all $\omega \in \mathbb{R}$, and all $j \geq 1$,

$$w(2^{-j}\omega) \geq C\,\chi_K(\omega).$$

2) Assume now $\omega \neq 0$. Then

$$(\mathcal{L}_w^n f)(\omega) = \sum_{m=-2^{n-1}+1}^{2^{n-1}} \Big(\prod_{j=1}^{n} w(2^{-j}(\omega + 2m\pi))\Big) f(2^{-n}(\omega + 2m\pi))$$

$$\geq C^n \sum_{m=-2^{n-1}+1}^{2^{n-1}} \chi_K(\omega + 2m\pi)\, f(2^{-n}(\omega + 2m\pi))\,.$$

There exists $\tilde{\omega} \in K$, $m_1 \in \mathbb{Z}$, so that $\omega + 2m_1\pi = \tilde{\omega}$. Therefore, if n is large enough, so that $2^{n-1} > |m_1|$, we have

$$(\mathcal{L}_w^n f)(\omega) \geq C^n\, f(2^{-n}\tilde{\omega})\,.$$

Since $\tilde{\omega} \neq 0$, and since $f \in E_\alpha$ is analytic, f cannot vanish on all the $2^{-n}\tilde{\omega}$, implying that $(\mathcal{L}_w^n f)(\omega) > 0$ for some $n \geq 1$.

3) If $\omega = 0$, then

$$(\mathcal{L}_w^n f)(0) = \sum_{m=-2^{n-1}+1}^{2^{n-1}} \Big(\prod_{j=1}^{n} w(2^{-j}2m\pi)\Big)\, f(2^{-n}2m\pi)$$

$$\geq \sum_{\ell=-2^{n-2}+1}^{2^{n-2}-1} \Big(\prod_{j=1}^{n} w(2^{-j+1}(2\ell+1)\pi)\Big)\, f(2^{-n+1}(2\ell+1)\pi)$$

$$= w(\pi) \sum_{\ell=-2^{n-2}+1}^{2^{n-2}-1} \Big(\prod_{j=1}^{n-1} w(2^{-j}(\pi + 2\ell\pi))\Big)\, f(2^{-n+1}(\pi + 2\ell\pi))\,.$$

Again, there exists $\tilde{\omega} \in K$ and $\ell_1 \in \mathbb{Z}$ so that $\pi + 2\ell_1\pi = \tilde{\omega}$. If $2^{n-2} > |\ell_1|$, then it follows that

$$(\mathcal{L}_w^n f)(0) \geq w(\pi)\, C^{n-1}\, f(2^{-n+1}\tilde{\omega})\,.$$

Since $\tilde{\omega} \neq 0$, the conclusion then follows as in point 2) above.

Lemma 4.6. *Let m, q be as in (4.3), (4.4), with $q(0) = 1$, $q(\pi) \neq 0$. Assume moreover that one of the following two sets of conditions holds:*

1) *$p > 0$ and $q(\omega)$ does not vanish on $[-\pi, \pi]$, or*

2) *$p \in 2\mathbb{N}$, $p \geq 2$, and m is of type C.*

Then r_p, the spectral radius of $\pounds_p$ on E_α, $\gamma < \alpha < 2\gamma$, is an eigenvalue of algebraic multiplicity 1, *and the corresponding eigenfunction is strictly positive. Moreover $r_p > 1$.*

PROOF. If q does not vanish on $[-\pi,\pi]$, then $|q|^p$ is obviously of type C. On the other hand, if m is of type C, then so is q, hence $|q|^p$. In both cases we can therefore apply Lemma 4.5, and we find, for $\lambda > r_p$, and for any $f \geq 0$, any $\omega \in [-\pi,\pi]$,

$$\sum_{n=1}^{\infty} \lambda^{-n} (\pounds_p^n f)(\omega) > 0 .$$

In the terminology of Schaefer (1966), this means that $\pounds_p$ is irreducible. It then follows from Theorem 3.3 in the Appendix of Schaefer (1966) that the algebraic multiplicity of r_p is 1 and that the associated eigenfunction is strictly positive; let us call this eigenfunction F. The inequality $r_p > 1$ follows from

$$\begin{aligned} r_p^n F(0) &= (\pounds_p^n F)(0) \\ &= F(0) + \sum_{\substack{m=-2^{n-1}+1 \\ m \neq 0}}^{2^{n-1}} \prod_{j=1}^{n} |q(2^{-j} 2m\pi)|^p \, F(2^{-n} 2m\pi) . \end{aligned}$$

The argument in point 3) of the proof of Lemma 4.5 shows that this second term must be strictly positive for some n, implying $r_p > 1$.

We prove one additional lemma, which we shall use in the next section, although we do not need it for Theorem 4.8 below. The argument in the proof is borrowed from Hervé (1995).

Lemma 4.7. *Let m, q be as in Lemma* 4.6. *Then r_p is the only eigenvalue of $\pounds_p$ in the peripheral spectrum, i.e. all the other eigenvalues λ satisfy $|\lambda| < r_p$.*

PROOF. Let F be the strictly positive eigenfunction of $\pounds_p$ in Lemma 4.6, and define the function

$$v(\omega) = \frac{1}{r_p \, F(2\omega)} \, F(\omega) \, |q(\omega)|^p .$$

Then v is a continuous function, and it satisfies $v(\omega) + v(\omega + \pi) = 1$. It is then a consequence of results proved in Keane (1972) that, for any continuous 2π-periodic function g, $(\mathcal{L}_v^n g)(\omega)$ converges uniformly to some constant C_g. Consequently, for any $f \in E_\alpha$,

$$r_p^{-n}\,(\mathcal{L}_p^n f)(\omega) = F(\omega)\,(\mathcal{L}_v^n(f/F))(\omega) \to F(\omega)\,C_{f/F}\,.$$

If now f was an eigenvector of $\mathcal{L}_p$ with eigenvalue λ, with $\lambda = r_p\,\tilde{\lambda}$, $\tilde{\lambda} \neq 1$, $|\tilde{\lambda}| = 1$, then this would imply

$$\tilde{\lambda}^n f(\omega) \underset{n\to\infty}{\longrightarrow} C_{f/F}\,F(\omega)\,.$$

This is impossible (just take any ω such that $F(\omega) \neq 0 \neq f(\omega)$).

We are now ready to state the result that links the regularity of φ with the spectral properties of transfer operators.

Theorem 4.8. 1) *Assume that $m(\omega)$, $q(\omega)$ satisfy the same conditions as in Lemma 4.6. Let $\mathcal{L}_p$ be the transfer operator associated to the function $|q(\omega)|^p$ and let r_p be the spectral radius of this operator on E_α, for any $\alpha \in]\gamma, 2\gamma[$. Then the L^p-Sobolev exponent s_p of φ satisfies*

$$s_p = N - \frac{1}{p}\,\log_2(r_p)\,. \tag{4.10}$$

Furthermore, one always has

$$s_p < N\,. \tag{4.11}$$

2) *If $q(\omega)$ has some zeros in $[-\pi, \pi]$ and is not of type C, then we still have $s_p \geq N - \log_2(r_p)/p$, for $p \in 2\mathbb{N}$.*

PROOF. We start by proving that $s_p \geq N - \log_2(r_p)/p$ for general $q(\omega)$. Combining (4.3) and (4.6), we obtain

$$\hat{\varphi}(\omega) = \prod_{k=1}^{+\infty} \cos^N(2^{-k-1}\omega) \prod_{k=1}^{+\infty} q(2^{-k}\omega) = \left(\frac{2\sin(\omega/2)}{\omega}\right)^N A(\omega)$$

with $A(\omega) = \prod_{k=1}^{+\infty} q(2^{-k}\omega)$.

To exploit Lemma 3.1, we remark that for all $\omega \in [-2^n\pi, 2^n\pi]$, we have

$$|A(\omega)|^p \leq C_p \prod_{k=1}^{n} |q(\omega)|^p \tag{4.12}$$

with $C_p = \max_{\omega\in[-\pi,\pi]} |A(\omega)|^p$. By Lemma 3.1, we can write

$$\begin{aligned}\int_{-2^n\pi}^{2^n\pi} |A(\omega)|^p \, d\omega &\leq C_p \int_{-2^n\pi}^{2^n\pi} \prod_{k=1}^{n} |q(2^{-k}\omega)|^p \, d\omega \\ &= C_p \int_{-\pi}^{\pi} (\pounds_p)^n 1(\omega)\, d\omega \\ &= 2\pi\, C_p \, \langle (\pounds_p)^n 1 | 1 \rangle_\alpha \,,\end{aligned}$$

where we have used (3.7).

It follows that for all $\varepsilon > 0$ and $p > 0$, there exists a constant $C_{p,\varepsilon}$ such that

$$\int_{-2^n\pi}^{2^n\pi} |A(\omega)|^p \, d\omega \leq C_{p,\varepsilon} \, (r_p + \varepsilon)^n \,. \tag{4.13}$$

We now study the convergence of $\int |\hat{\varphi}(\omega)|^p (1+|\omega|^p)^s \, d\omega$ by a dyadic decomposition of the frequency domain:

$$\begin{aligned}\int_{\mathbb{R}} |\hat{\varphi}(\omega)|^p \, (1 + |\omega|^p)^s \, d\omega &= \int_{-\pi}^{\pi} |\hat{\varphi}(\omega)|^p \, (1 + |\omega|^p)^s \, d\omega \\ &\quad + \sum_{j=1}^{+\infty} \int_{|\omega|\in[2^{j-1}\pi, 2^j\pi]} |\hat{\varphi}(\omega)|^p \, (1 + |\omega|^p)^s \, d\omega \\ &\leq C_1 + C_2 \sum_{j=1}^{+\infty} 2^{psj} \, 2^{-Npj} \int_{-2^n\pi}^{2^n\pi} |A(\omega)|^p \, d\omega \\ &\leq C_1 + C_3 \sum_{j=1}^{+\infty} 2^{psj} \, 2^{-Npj} \, (r_p + \varepsilon)^j \,,\end{aligned}$$

where C_1, C_2 and C_3 only depend on the choice of p and ε. Since ε can be chosen arbitrarily small, it is clear that the integral will converge whenever $sp < Np - \log_2(r_p)$. This shows that we have

$$s_p \geq N - \frac{1}{p} \, \log_2(r_p) \,. \tag{4.14}$$

To sharpen this into an equality when $q(\omega)$ satisfies the extra conditions, we shall use the transfer operator $\mathcal{L}'_p$ associated to the function $|m(\omega)|^p$. According to Lemma 4.3, the spectral radius of $\mathcal{L}'_p$ on E'_α is $2^{-Np} r_p$. Moreover, since $F > 0$ is an eigenfunction of $\mathcal{L}_p$ for the eigenvalue r_p, it follows that $2^{-Np} r_p$ is an eigenvalue for $\mathcal{L}'_p$, with a corresponding positive eigenfunction $g(\omega) = |\sin(\omega/2)|^{Np} F(\omega)$.

Define now $S = \sup_\omega |F(\omega)|$, $G = \int_{-\pi}^{\pi} g(\omega)\, d\omega > 0$. Let K be the compact set congruent with $[-\pi, \pi]$ for which $\inf_{j \geq 1,\, \omega \in K} |m(\omega)|^p > 0$. It then follows (see *e.g.* Cohen and Ryan (1995)) that

$$\rho = \inf_{\omega \in [-\pi,\pi]} |\hat{\varphi}(\omega)|^p > 0\,.$$

Define now

$$I_n = \int_{\omega \in 2^n K} |\omega|^{Np}\, |\hat{\varphi}(\omega)|^p\, d\omega\,. \tag{4.15}$$

Using again Lemma 3.1, we obtain, for all $n \geq 1$,

$$\begin{aligned}
I_n &\geq \rho^{-1} \int_{\omega \in 2^n K} |\omega|^{Np} \Big| \prod_{k=1}^{n} m(2^{-k}\omega) \Big|^p d\omega \\
&\geq \rho^{-1}\, 2^{Np(n+1)} \int_{\omega \in 2^n K} |\sin(2^{-n-1}\omega)|^{Np} \Big| \prod_{k=1}^{n} m(2^{-k}\omega) \Big|^p d\omega \\
&= \rho^{-1}\, 2^{Np(n+1)} \int_{-2^n \pi}^{2^n \pi} |\sin(2^{-n-1}\omega)|^{Np} \Big| \prod_{k=1}^{n} m(2^{-k}\omega) \Big|^p d\omega \\
&\geq (S\rho)^{-1}\, 2^{Np(n+1)} \int_{-2^n \pi}^{2^n \pi} g(2^{-n}\omega) \Big| \prod_{k=1}^{n} m(2^{-k}\omega) \Big|^p d\omega \\
&= (S\rho)^{-1}\, 2^{Np(n+1)} \int_{-\pi}^{\pi} (\mathcal{L}'_p)^n g(\omega)\, d\omega \\
&= G(S\rho)^{-1}\, 2^{Np}\, (r_p)^n = C\, (r_p)^n\,.
\end{aligned}$$

Since, for some $L < \infty$, $K \subset \{\omega : \ |\omega| \leq 2^L \pi\}$, it follows that

$$\begin{aligned}
\tilde{I}_n &= \int_{-2^n \pi}^{2^n \pi} |\omega|^{Np}\, |\hat{\varphi}(\omega)|^p\, d\omega \\
&\geq \int_{\omega \in 2^{n-L} K} |\omega|^{Np}\, |\hat{\varphi}(\omega)|^p\, d\omega \\
&= I_{n-L} \geq C'\, (r_p)^n\,.
\end{aligned} \tag{4.16}$$

If we now define

$$J_n = \tilde{I}_n - \tilde{I}_{n-1} = \int_{|\omega|\in[2^{n-1}\pi,2^n\pi]} |\omega|^{Np}\,|\hat{\varphi}(\omega)|^p\,d\omega\,, \tag{4.17}$$

then (4.16) shows that for all $C, \varepsilon > 0$, there is an infinite number of $n \geq 1$ such that $J_n \geq C(r_p)^n\,2^{-\varepsilon n}$ (here we use that, by Lemma 4.6, r_p is strictly larger than 1), or equivalently

$$\int_{|\omega|\in[2^{n-1}\pi,2^n\pi]} |\omega|^{Np-\log_2(r_p)+\varepsilon}\,|\hat{\varphi}(\omega)|^p\,d\omega \geq C\,. \tag{4.18}$$

This last inequality shows that s_p is smaller than (and thus equal to) $N - \log_2(r_p)/p$. Finally, $s_p < N$ follows from $r_p > 1$.

5. Numerical precision.

Combining the results of the previous sections, we immediately obtain

Theorem 5.1. *For a function φ as defined by (4.6), with $m(\omega)$, $q(\omega)$ satisfying the conditions in Lemma 4.6, the L_p-Sobolev exponent of φ can be expressed as $s_p = N - \log_2(r_p)/p$, where $(r_p)^{-1} = x_p$ is the zero of smallest absolute value of the Fredholm determinant $d_p(z)$ of the operator $\mathcal{L}_p$ associated to the weight function $w(\omega) = |q(\omega)|^p$.*

Formulas (2.12), (2.16) and (3.19) give us an explicit expression for the Taylor series of the analytic function $d_p(z)$. In practice, to estimate x_p numerically, we are obliged to truncate this series and work with the polynomials that are obtained from the first order terms. What is the effect of this truncation on the numerical precision of the estimate for x_p? More precisely, can we measure how well x_p is approximated by the smallest zero of the truncated series at a given order?

We first discuss this problem in very general terms. Let $f(z)$ be an analytic function on $\mathbb{C}$ and suppose that we want to estimate the value of z_0, the zero of f with the smallest absolute value. For a given N, let us denote by P_N the polynomial corresponding to the N first order terms in the Taylor development of f around 0 and let $R_N(z) = f(z) - P_N(z)$ be the residual term. Then, for any fixed $A > 0$, and any $\lambda \in]0,1[$, one can find $C > 0$ so that $\sup_{|z|\leq A}\ |R_N(z)| \leq C\lambda^N$: the R_N converge to zero faster than any geometric sequence, uniformly on the disk $|z| \le A$. To exploit this estimate, we shall use a classic result of complex analysis (see, for example, Rudin (1967)) that we recall here:

Rouché's Theorem. *Let $g(z)$ and $h(z)$ be analytic functions on an open set V; let D be an open set such that $\overline{D} \subset V$ and ∂D is a Jordan curve. If $|h(z)| < |g(z)|$ for all $z \in \partial D$, then g and $h+g$ have the same number of zeros inside D.*

This theorem leads to a systematic method for tracking the zeros of f:

- First, find $\varepsilon_0 > 0$, $A > 0$ and $N_0 \in \mathbb{N}$ such that $|P_{N_0}(z)| > \varepsilon_0$ on $S_A = \{|z| = A\}$, $|R_n(z)| < \varepsilon_0/4$ in $B_A = \{|z| \le A\}$ for all $n \ge N_0$ and P_{N_0} has at least one zero in B_A. This is always possible, if f has at least one zero in B_A and does not vanish on S_A. Moreover, Rouché's theorem implies that f and P_{N_0} have the same number M of zeros in B_A.

- For a given zero $z_{0,j}$, $j \in \{1, \dots, M\}$ of P_{N_0} in B_A, consider then the parametrized curve $\Gamma_{0,j}(\theta) = z_{0,j} + u_{0,j}(\theta)\, e^{i\theta}$, $\theta \in [0, 2\pi]$, defined by

$$u_{0,j}(\theta) = \min\{u : \ P_{N_0}(z_{0,j} + u(\theta)\, e^{i\theta}) = \varepsilon_0\}\,. \tag{5.1}$$

  This is clearly a Jordan curve contained in B_A. By Rouché's theorem again, we know that the curve $\Gamma_{0,j}$ embraces the same number of zeros not only for f and P_{N_0} but also for all the P_n for $n \ge N_0$ since, on $\Gamma_{0,j}$, we have $|P_n(z)| \ge |P_{N_0}(z)| + |R_{N_0}(z)| + |R_n(z)| \ge \varepsilon_0/2$.

- We now iterate this process by taking any sequence ε_k starting with ε_0 and such that $0 < \varepsilon_k \le \varepsilon_{k-1}/2$. After k steps, we choose $N_k > N_{k-1}$ such that $|R_n(z)| < \varepsilon_k/4$ in $B_A = \{|z| \le A\}$ for all $n \ge N_k$. The curves $\Gamma_{k,j}$ will be defined around the zeros $z_{k,j}$, $j \in \{1, \dots, M\}$ of P_{N_k} by $\Gamma_{k,j}\ (\theta) = z_{k,j} + u_{k,j}(\theta)\, e^{i\theta}$, $\theta \in [0, 2\pi]$, with

$$u_{k,j}(\theta) = \min\{u : \ P_{N_k}(z_{k,j} + u(\theta)\, e^{i\theta}) = \varepsilon_k\}\,. \tag{5.2}$$

  It is clear that $\Gamma_{k,j}$ lies completely within $\Gamma_{k-1,j}$ and that for fixed k, each $\Gamma_{k,j}$ contains at least one zero z_j of f, $j \in \{1, \dots, M\}$ (if it contains two zeros z_i and z_j, then the curves $\Gamma_{k,i}$ and $\Gamma_{k,j}$ are necessarily identical).

The speed of convergence of this process will be measured by the decay of the diameter of the curves $\Gamma_{k,j}$ as k goes to $+\infty$. We can estimate this diameter by remarking that $|f(z)| < 5\varepsilon_k/4$ within $\Gamma_{k,j}$. Consequently if d is the maximal order of all the zeros z_j of f in B_A, $j \in \{1, \ldots, M\}$, then there exists a constant C such that

$$\begin{aligned} \max_j \operatorname{diam}(\Gamma_{k,j}) &\le \max_{j,r}\{r : |z - z_j| < r \Rightarrow |f(z)| < 5\varepsilon_k/4\} \\ &< C(\varepsilon_k)^{1/d}. \end{aligned} \tag{5.3}$$

In particular, we see that since ε_k has at least exponential decay, the speed of convergence will always be, at least, exponential. Note that for sufficiently large k, all the $\Gamma_{k,j}$ curves ($i = 1, \ldots, M$) are disjoint from each other so that all the zeros are isolated and can be tracked separately.

In the particular case that we are interested in, some additional considerations can be made:

- We are looking for the zero with the smallest absolute value. By Lemma 4.6, this zero is unique and situated in $\mathbb{R}_+$. Furthermore, we know that it is contained in $[0, 1[$. Consequently, we may restrict our tracking process to this interval after the first step, using the fact that, for all k, the zero that we are looking for is necessarily situated between the two extremal intersections of a certain $\Gamma_{k,j}$ with the real axis.
- We have now a specific estimate for the rest $R_{N,p}(z)$ of $d_p(z)$: according to (2.11), for all $|z| \le 1$,

$$|R_{N,p}(z)| \le \sum_{k=N+1}^{+\infty} \frac{1}{k!}(\mathrm{Tr}|\mathcal{L}_p|)^k. \tag{5.4}$$

This estimate indicates that the estimation process that uses $P_k(z)$ at step k should converge at least exponentially fast.

Before we proceed to the examples in the next section, we list a few comments on our procedure.

COMMENTS.

1). According to Lemma 4.7, we can split the eigenvalues α_n of $\mathcal{L}_p$ into $\alpha_0 = r_p$, and all the other α_n, $n \geq 1$, which satisfy $|\alpha_n| < r_p$. We can therefore write the following estimate for $\operatorname{Tr} \mathcal{L}_p^k$:

$$|\operatorname{Tr} \mathcal{L}_p^k - r_p^k| \leq r_p^k \sum_{\ell=1}^{\infty} p_\ell^k N_\ell \,, \tag{5.5}$$

where all the p_ℓ are < 1, where $N_\ell = \#\{\alpha_n : \ |\alpha_n| = p_\ell\}$, and $\sum_{\ell=1}^{\infty} p_\ell N_\ell = \sum_{n=1}^{\infty} |\alpha_n| < \infty$. This leads to two alternative formulas for the computation of r_p, starting from the traces $\operatorname{Tr} \mathcal{L}_p^k$:

$$r_p = \lim_{k\to\infty} |\operatorname{Tr} \mathcal{L}_p^k|^{1/k} \,, \tag{5.6}$$

and

$$r_p = \lim_{k\to\infty} \frac{|\operatorname{Tr} \mathcal{L}_p^{k+1}|}{|\operatorname{Tr} \mathcal{L}_p^k|} \,. \tag{5.7}$$

According to the estimate (5.5), we have indeed

$$|\operatorname{Tr} \mathcal{L}_p^k|^{1/k} = r_p + C_1(k)\, p_1^k \tag{5.8}$$

and

$$\frac{|\operatorname{Tr} \mathcal{L}_p^{k+1}|}{|\operatorname{Tr} \mathcal{L}_p^k|} = r_p + C_2(k)\, p_1^k \,, \tag{5.9}$$

where, for k sufficiently large,

$$|C_1(k)| \leq (1+\epsilon)\, \frac{N_1}{k}\, r_p$$

and

$$|C_2(k)| \leq (1+\epsilon)\,(p_1+1)\, N_1\, r_p \,,$$

which proves (5.6) and (5.7) (since $p_1 < 1$). Both formulas converge exponentially fast, with the same rate, but a slightly better multiplicative constant in (5.6) than in (5.7), according to (5.8), (5.9). Computing r_p via either (5.6) or (5.7) is simpler than the Fredholm determinant method explained above, but although we have exponential convergence in (5.6), (5.7), we have no control over, or no good estimate for the rate of convergence. Since $\mathcal{L}_p$ is not selfadjoint, its eigenvalues can be complex, and it is conceivable that the largest $|\alpha_0|$ may be "masked" in the first sums. This is illustrated by the following example (which is admittedly ad hoc, and not computed as the spectrum of a true $\mathcal{L}_w$). Take

$$\begin{aligned} &\alpha_0 = 1\,, \qquad \alpha_n = (1-\epsilon)\,e^{2\pi i n/K}\,, \quad n = 1, \ldots, K-1\,, \\ &\alpha_K = \gamma\,, \qquad \alpha_n = 0\,, \quad n > K\,. \end{aligned}$$

Then we have

$$\sum_{n=0}^{\infty} \alpha_n^k = \begin{cases} 1 + (K-1)(1-\epsilon)^k + \gamma^k\,, & \text{if } k = 0 \pmod{K}\,, \\ 1 - (1-\epsilon)^k + \gamma^k\,, & \text{if } k \neq 0 \pmod{K}\,. \end{cases}$$

Consequently the first $K-1$ sums $\sum_{n=0}^{\infty} \alpha_n^k$, $1 \le k \le K-1$, may lead to a very misleading picture. Figure 1a plots $\log\left(\sum_{n=0}^{\infty} \alpha_n^k\right)$ as a function of k for $1 \le k \le 14$ for the choices $K = 15$, $\gamma = .9$, $\epsilon = .001$; the graph is virtually indistinguishable from the straight line $k \log \gamma$. The picture changes drastically when we reach $k = 15$, as shown in Figure 1b, which plots the behavior of $\sum_{n=0}^{\infty} \alpha_n^k$ for a much larger range of k. The point of this toy example is that the first $K-1$ sums $\sum_{n=0}^{\infty} \alpha_n^k$ contain no clue whatsoever indicating that the asymptotic regime is far from attained.

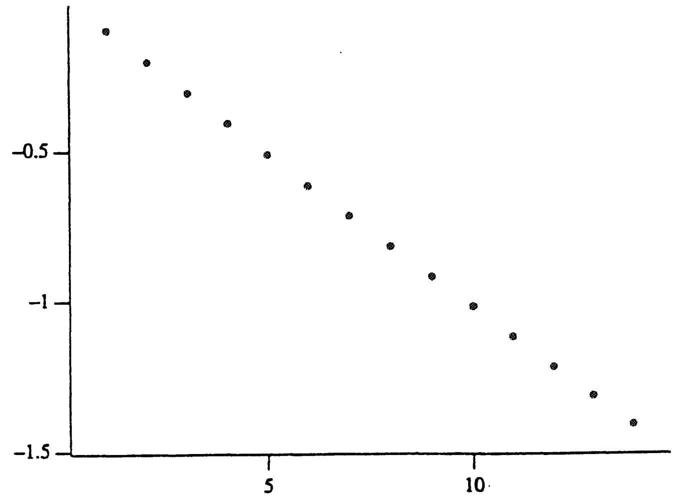


Figure 1a. Plot of $\log(\sum_{n=0}^{\infty} \alpha_n^k)$ for the values $k = 1, \ldots, 14$ in the toy example.

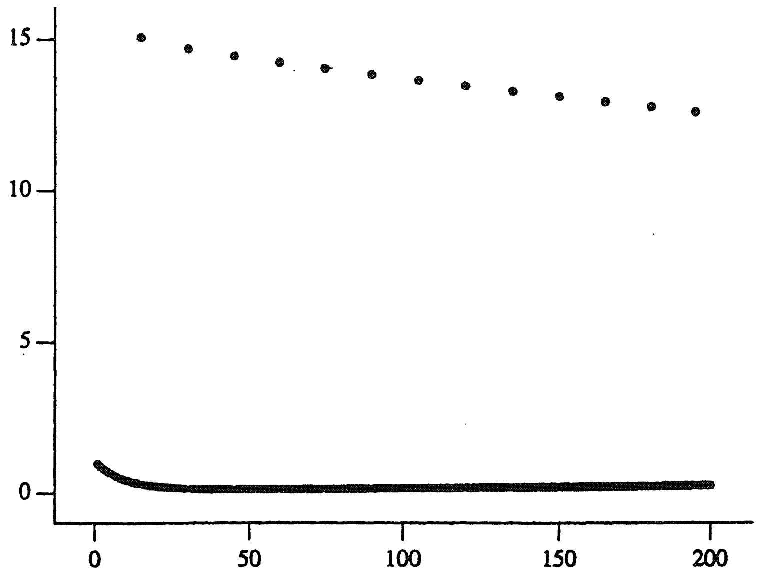


Figure 1b. Plot of $\sum_{n=0}^{\infty} \alpha_n^k$ for $k = 1, \ldots, 200$ in the toy example, showing the much slower exponential decay to the limit value 1.

The zero-tracking method above, in contrast, requires more work than (5.6) or (5.7), but it gives explicit error estimates at every step. An example of a true $\mathcal{L}_w$-spectrum that gives rise to a similar masking effect is given by $w(\omega) = |q(\omega)|^2$, with

$$q(\omega) = 100\,(1 - .9\,e^{3i\xi})\,(1 - .9\,e^{7i\xi})\,.$$

In this case w is a trigonometric polynomial, so that the spectral radius r_2 of $\mathcal{L}_2 = \mathcal{L}_w$ is simply the largest eigenvalue of a 21×21 matrix, which can easily be computed explicitly; one finds $r_2 = 32\,642.525$. For small values of k, the traces of $\mathcal{L}_2^k$ are much smaller than r_2^k; this is due to the fact that q is close to vanishing at the nontrivial fixed points of τ^2 and τ^3, where τ is the doubling operator, modulo 2π, on $[0, 2\pi]$. Figure 2 shows the values of $|\mathrm{Tr}\ \mathcal{L}_2^k|^{1/k}$ for $k = 1$ to 25.

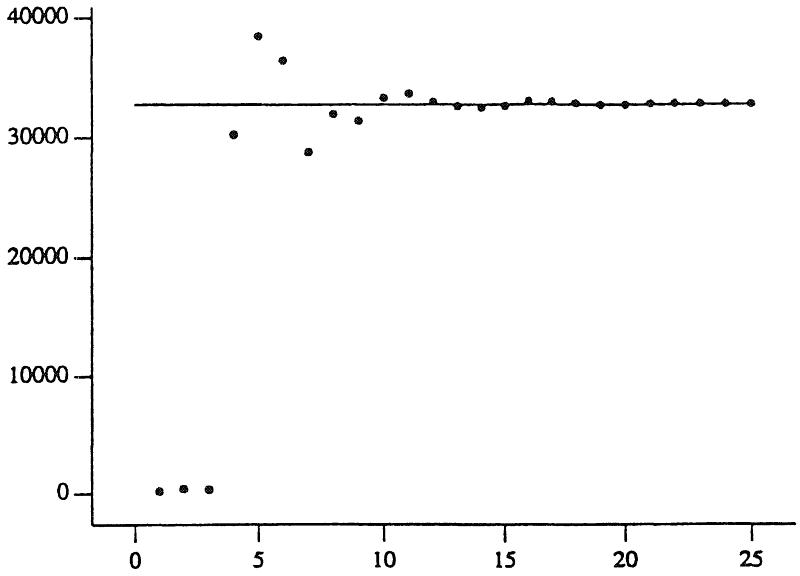


Figure 2. The values of $|\mathrm{Tr}\,\mathcal{L}_2^k|^{1/k}$ for $k = 1, \ldots, 25$. The first few values, for small k, give a misleading idea of what the limit value might be. The horizontal solid line indicates the true value of r_2, to which $|\mathrm{Tr}\,\mathcal{L}_2^k|^{1/k}$ is seen to converge.

The misleadingly small values (when compared with r_2) for $k = 2, 3$ can also be understood in terms of the eigenvalues of $\mathcal{L}_2$, illustrated in Figure 3, which do indeed fan out over different angles in the complex plane. The effect is not as pronounced here as in the ad hoc example above; it occurs only for $k = 2, 3$, and the first values of $\log|\mathrm{Tr}\,\mathcal{L}_2^k|$ do not line up along a line with misleading slope. In other, more complicated operators $\mathcal{L}_w$ the masking effect could well be more pronounced, and possibly be as strong as in the toy example. The Fredholm determinant method would of course not succeed any better in extracting r_2 from the values of $\mathrm{Tr}\,\mathcal{L}_2^k$ for small k; the "error bar estimate", computed from the Fourier coefficients of $w(\omega)$, would automatically tell us, however, that we have to look at larger k in order to conclude something sensible.

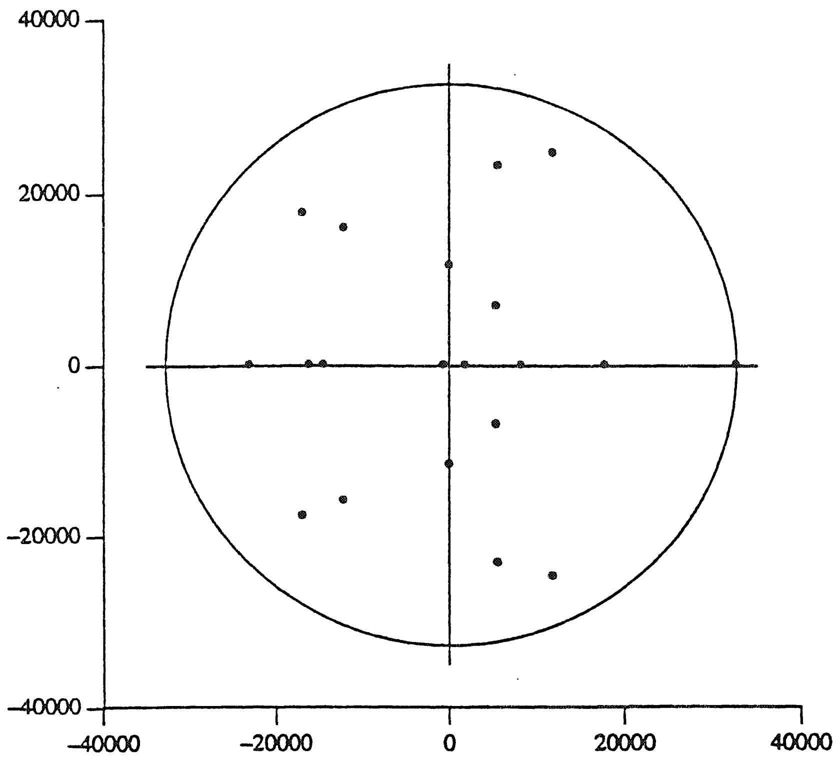


Figure 3. Position of the different eigenvalues of $\mathcal{L}_2$ for $q(\omega) = 100\,(1-.9\,e^{3i\omega})\,(1-.9\,e^{7i\omega})$.

2) If we replace w in (3.19) by $|q|^k$, and substitute this into (5.7), then the resulting formula is very similar to formulas found in Hervé (1995) (see *e.g.* Theorem 6.2 in Hervé (1995) -beware of changes of notation). The only difference is that we have a denominator $2^k - 1$ in the arguments of $|q(\cdot)|^k$, where Hervé has 2^k, because we sum over fixed points of D^k, where Hervé's approach sums over preimages of 0 under D^k (where D is multiplication by 2 on $[-\pi, \pi]$, $\mathrm{mod}\, 2\pi$). The convergence of Hervé's formula is similar to that of (5.6), (5.7) above, and in principle, the same reservations as in point 1 above apply, although

we have not seen any cases where problems occurred in practice. In Hervé (1995) the operators $\mathcal{L}_p$ are studied on the much larger spaces $C^\gamma(-\pi,\pi)$ of Hölder continuous functions with exponent γ. On these spaces the $\mathcal{L}_p$ are not compact; they are quasi-compact, meaning that the radius of their essential spectrum is strictly smaller than the spectral radius itself; this corresponds to a spectrum where only discrete eigenvalues are possible in an outer annulus of the disk with spectral radius. Even though the operators are thus more complicated, Hervé's method has the advantage that he can treat also the case where $p \notin 2\mathbb{N}$ and q has zeros in $[-\pi,\pi]$. It is interesting to note that the eigenvalues α_n in the annulus $\{\lambda;\ 2^{-\gamma}r < |\lambda| \le r\}$ (where γ is the Hölder regularity of $|q(\omega)|^p$) can still be tracked with the Fredholm determinant method (see Theorem 7.1 below), so that even in this case we can use our numerical approach and get absolute error estimates.

3) In practice, one can of course also use the reverse procedure: instead of setting first ϵ and then searching for the appropriate N, as sketched above, one can fix a (relatively large) value for N, find the corresponding smallest zero $z_{N,0}$ of P_N, and then identify ϵ_N so that $|R_N(z)| < \epsilon_N$ on the curve $\Gamma_{N,0}$ defined by (5.2).

6. Examples.

All our examples are motivated by wavelet constructions; we take the refinable function φ to be either the orthonormal scaling function in a multiresolution analysis, or the autocorrelation function of a scaling function. We start by recalling some pertinent definitions.

The refinable function $\varphi(x)$ is said to be *cardinal interpolant* if it satisfies the condition

$$\varphi(k) = \delta_{0,k}\,, \qquad k \in \mathbb{Z}\,; \tag{6.1}$$

it is called *orthonormal* if

$$\int \varphi(x-k)\,\overline{\varphi(x-\ell)}\,dx = \delta_{k,\ell}\,, \qquad k,\ell \in \mathbb{Z}\,. \tag{6.2}$$

These properties correspond to special constraints on $m(\omega)$: (6.1) implies

$$m(\omega) + m(\omega+\pi) = 1\,, \tag{6.3}$$

whereas (6.2) can hold only if

$$|m(\omega)|^2 + |m(\omega + \pi)|^2 = 1\,. \tag{6.4}$$

The conditions (6.3) or (6.4) are necessary for (6.1) or (6.2) to hold, but not sufficient. Under additional technical conditions that ensure uniform convergence of the subdivision algorithm in the first case, or L^2-convergence in the second case, (6.3) implies (6.1) and (6.4) implies (6.2). (For a detailed discussion, see Chapter 6 in Daubechies (1992).)

It is clear that cardinal interpolation and orthonormality are linked: if ϕ is an orthonormal refinable function, then its autocorrelation function $\Phi(x) = \int \phi(y)\,\overline{\phi(x-y)}\,dy$ is interpolating; the corresponding functions m_ϕ and m_Φ are related by $m_\Phi = |m_\phi|^2$. In fact, compactly supported wavelets are usually constructed by first identifying a suitable positive m_Φ and then constructing m_ϕ so that $|m_\phi|^2 = m_\Phi$. It is then obvious that the L^p-Sobolev exponents of ϕ and Φ are related by

$$s_p(\Phi) = 2\,s_{2p}(\phi)\,. \tag{6.5}$$

In particular, using the definitions (1.2) and (1.3), we find

$$\mu(\Phi) = s_1(\Phi) = 2\,s_2(\phi) = 2\,s(\phi)\,, \tag{6.6}$$

where the first equality is a consequence of $\hat{\Phi}(\omega) \geq 0$ for all $\omega \in \mathbb{R}$.

In the first subsection below we concentrate on families of examples where m_Φ is a positive trigonometric polynomial of the form $P_\Phi(\cos\omega)$, so that Φ is real, symmetric and compactly supported. By Riesz' spectral factorization lemma, we can then find a trigonometric polynomial m_ϕ, with real coefficients, so that $|m_\phi|^2 = m_\Phi$. The corresponding refinable functions ϕ are then compactly supported scaling functions from which compactly supported wavelets can be constructed; see Daubechies (1992). In the second subsection we consider examples where $P_\Phi(\cos\omega)$ is no longer a trigonometric polynomial, but the quotient of two such polynomials. The third subsection of examples looks at some two-dimensional examples $m(\omega_1, \omega_2)$ which cannot be written as products $m_1(\omega_1)\,m_2(\omega_2)$ of one-dimensional functions, with matrix dilations. Finally, in the fourth subsection we use the results of the previous subsections to deal with a problem on spline wavelet bases.

6.1. Interpolating and orthonormal scaling functions with compact support.

The minimal degree solution to (6.3) and the factorization requirement (4.3) is given by

$$m_N(\omega) = \left(\cos\frac{\omega}{2}\right)^{2N} \sum_{j=0}^{N-1} \binom{N-1+j}{j} \left(\sin\frac{\omega}{2}\right)^{2j}.$$

(We are interested in factoring out only even powers of $\cos(\omega/2)$ because we want $m_N(\omega) = P_N(\cos\omega)$.) In this case $q(\omega)$ is clearly strictly positive for all $\omega \in [0, 2\pi]$, so that we can apply our theorems for all values $p \geq 1$. The corresponding functions Φ_N, ϕ_N have been studied extensively (see *e.g.* Daubechies (1992) for many references). We have computed the L^p-Sobolev exponents of these functions for different values of N. Table 1 and Figure 4 show the $s_p(\phi_N)$ for $p = 1, 2, 4, 8$, $N = 1, \ldots, 19$.

$N \setminus p$	1	2	4	8
1	−0.322289	0.338856	0.669428	0.834714
2	0.521293	0.999820	1.220150	1.310014
3	0.979675	1.414947	1.587361	1.631686
4	1.391644	1.775305	1.896446	1.912144
5	1.767934	2.096541	2.171522	2.174682
6	2.116733	2.388060	2.430780	2.431755
7	2.441544	2.658569	2.680780	2.680307
8	2.746639	2.914556	2.926425	2.925926
9	3.035292	3.161380	3.166924	3.165533
10	3,309107	3.402546	3.405193	3.405141
11	3.572141	3.639569	3.641221	3.638529
12	3.825525	3.873991	3.874236	3.871917
13	4.071021	4.105802	4.105736	4.105305
14	4.311641	4.336042	4.336476	4.335502
15	4.547368	4.564708	4.564925	4.562449
16	4.780028	4.792323	4.792608	4.792645
17	5.010231	5.018884	5.018754	5.016283
18	5.238588	5.244390	5.244127	5.243230
19	5.464480	5.468841	5.468728	5.466868

Table 1. L^p-Sobolev exponents of φ_N, $p = 1, 2, 4, 8$, $N = 1, \ldots, 19$ (Polynomial).

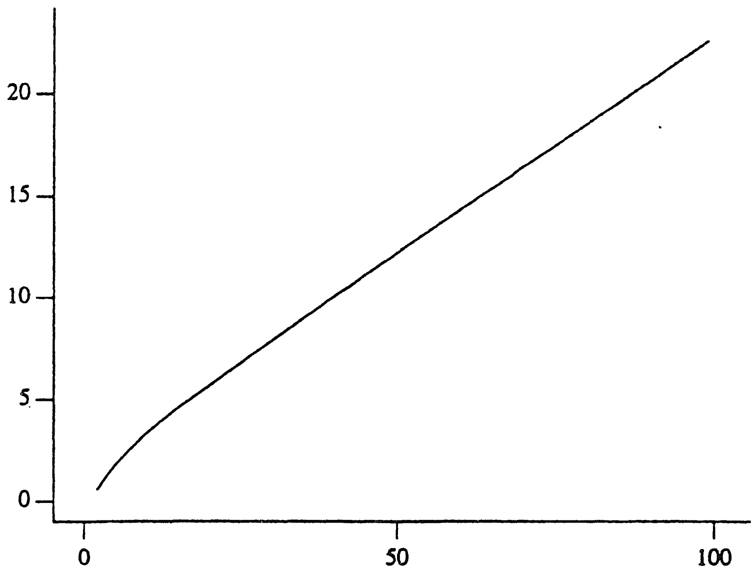


Figure 4. $s(\varphi_N)$ for $0 < N < 100$ (Polynomial).

An interesting observation is that $s_p(\varphi_N)$ becomes independent of p as N goes to $+\infty$. This reflects the fact that φ_N has a lacunary structure in the Fourier domain and that this phenomenon grows with n. More precisely, if the Fourier transform of a function φ decays uniformly at infinity in the sense that $C_1\,(1+|\omega|)^{-\alpha} \leq |\hat{\varphi}(\omega)| \leq C_2\,(1+|\omega|)^{-\alpha}$, then the exponents $s_p(\varphi) = \alpha - 1/p$ are related by $s_p - s_q = 1/q - 1/p$. This is true here only for $N = 1$ (which corresponds to the box function $\varphi_1(x) = \chi_{[0,1]}(x)$); $\hat{\varphi}_1(\omega) = (1 - e^{-i\omega})/(i\omega)$ decays uniformly in $|\omega|^{-1}$, up to the oscillation of the numerator. For larger N, the lacunary structure takes over. As shown in Volkmer (1992) and Cohen and Conze (1992), the worst decay occurs at the points $\omega_j = 2\pi\, 2^j/3$, $j > 0$. A possible explanation for our observation could be that the L^p norm of $\hat{\varphi}_N$ concentrates at these points, as N grows.

We thus conjecture that for all $p, q > 0$, $\lim_{N\to+\infty} |s_p(\varphi_N) - s_q(\varphi_N)| = 0$. If this is true, then we have in particular $\lim_{N\to+\infty} |s(\varphi_N) - \mu(\varphi_N)| = 0$ since $s_1(\varphi_N) \leq \mu(\varphi_N) \leq s(\varphi_N) = s_2(\varphi_N)$.

For large values of N, our method allows us to observe the asymptotic behavior of $\mu(\varphi_N)$.

It was proved by Volkmer (1992) that

$$\lim_{N\to+\infty} \frac{\mu(\varphi_N)}{N} = \lim_{N\to+\infty} \frac{s(\varphi_N)}{N} = 1 - \frac{\log_2 3}{2} \simeq 0.2075\,.$$

The graph of $s_2(\varphi_N) = s_2(N)$ presented in Table 1 shows in addition that $s_2(N) - (1 - (\log_2 3)/2)N$ stays bounded by 3 for $N \leq 100$.

6.2. Interpolating and orthonormal scaling functions with infinite support.

We now turn to the solutions of (6.3) that have the factorized form (4.3) but are not necessarily trigonometric polynomials.

We shall look for solutions of the type

$$m(\omega) = \cos^{2N}(\omega/2)\, R(\cos\omega)\,, \tag{6.7}$$

where $R(z) = P(z)/Q(z)$ is a rational function that is strictly positive on $[-1,1]$. Under this hypothesis we know that we can apply our method to estimate the L^p-Sobolev exponent s_p of the associated scaling function since the Fourier coefficients of $|R(z)|^p$ have exponential decay. Note that the scaling function φ is not compactly supported but typically still has exponential decay at infinity (some restrictions on R, always satisfied in practical examples, are needed to ensure this).

The choice of a rational function is still useful in the applications where one has to perform discrete convolutions with the Fourier coefficients of $m(\omega)$: although they are not finite in number, these convolutions can be implemented in a fast recursive way, the complexity being roughly $2S \times (\deg(P) + \deg(Q) + N)$ where S is the size of the input data.

The simplest rational solution of (6.3) of the form (6.7) is given by the family

$$m_N(\omega) = \cos^{2N}(\omega/2)\, R_N(\cos\omega) \tag{6.8}$$

with $R_N(\cos\omega) = (\cos^{2N}(\omega/2) + \sin^{2N}(\omega/2))^{-1}$. These solutions are well known in signal processing as the transfer functions of the so-called "Butterworth filters" (see Oppenheim and Schafer (1975) for a detailed review).

As in the previous section, we give the estimate of s_p for the orthonormal scaling functions φ_N, $1 \leq N < 20$ and $p = 1, 2, 4, 8$. It is interesting to see that these exponents remain substantially different as N grows: the lacunary behavior does not prevail as much as in the compactly supported case.

For large values of N, we have examined the evolution of $s(\varphi_N) = s(N)$ (see Table 2). It reveals a linear asymptotic behavior, similar to the compactly supported case.

$N \setminus p$	1	2	4	8
1	−0.322289	0.338856	0.669428	0.834714
2	0.677350	1.256211	1.495117	1.604344
3	1.561362	2.044109	2.269688	2.365870
4	2.370365	2.843768	3.059757	3.148599
5	3.183890	3.648646	3.857332	3.940563
6	3.999055	4.456118	4.658210	4.735925
7	4.815040	5.264533	5.460184	5.532265
8	5.630616	6.072947	6.262157	6.328326
9	6.446191	6.881125	7.063818	7.123827
10	7.260947	7.688598	7.864696	7.918627
11	8.075292	8.495600	8.664948	8.712863
12	8.888817	9.301894	9.464414	9.506534
13	9.701520	10.107480	10.263251	10.299921
14	10.513813	10.912358	11.061142	11.093166
15	11.325284	11.716526	11.858558	11.885986
16	12.135933	12.519984	12.655184	12.678805
17	12.946170	13.322968	13.451333	13.471625
18	13.755996	14.125241	14.247006	14.264159
19	14.564999	14.927039	15.042202	15.056836

Table 2. L^p-Sobolev exponents of φ_N, $p = 1, 2, 4, 8$, $N = 1, \ldots, 19$ (Butterworth).

Note that the limit ratio $s(N)/N \simeq .8$ seems to indicate that the worst decay of $\hat{\varphi}_N(\omega)$ occurs at the points $\omega_j = 2\pi\, 2^j/3$. Indeed, we have

$$|\hat{\varphi}_N(\omega_j)| = \left|\hat{\varphi}_N\left(\frac{\pi}{3}\right)\right| \left|m_N\left(\frac{2\pi}{3}\right)\right|^{j/2} = C\,|\omega_j|^{r_N} \tag{6.9}$$

with $r_N = \log_2(R_N(1/2))/2 - 1$. From the definition of R_N, we obtain

$$\lim_{N \to +\infty} \frac{r_N}{N} = -\frac{1}{2} \log_2 3 \simeq -0.7925, \tag{6.10}$$

which seems to coincide with the experimental asymptotic ratio.

The Butterworth functions $R_N(\cos\omega)$ correspond to a choice with $P(z) = 1$, $R(z) = 1/Q(z)$, which makes them in some sense opposites to the polynomial solutions of the previous subsection, for which $Q(z) = 1$, $R(z) = P(z)$. Recently, intermediate solutions that are equally balanced between the numerator and the denominator were proposed by Herley and Vetterli (1993). Such solutions can be built by the following procedure:

- fix $N > 0$ and $0 \leq k \leq N$ such that $N + k$ is odd.
- find a polynomial $P_k(z)$ such that

$$\left(\frac{1+z}{2}\right)^N P_k(z) + \left(\frac{1-z}{2}\right)^N P_k(-z)$$

has degree $N-k+1$. This can be done by solving k linear equations.

- define

(6.11)

$$m_N^k(\omega) = \frac{\cos^{2N}(\omega/2)\, P_k(\cos\omega)}{\cos^{2N}(\omega/2)\, P_k(\cos\omega) + \sin^{2N}(\omega/2)\, P_k(-\cos\omega)}\ .$$

Note that the global complexity of the convolution by the discrete filter associated to $m_N^k(\omega)$ is given by

$$C \simeq 2S(k + (N-k+1) + N) = 2S\,(2N+1) \tag{6.12}$$

and is thus independent of k.

Here we have considered a family of intermediate solutions by taking k close to $N/2$ so that the rational function $R_N^k(\cos\omega)$ has approximately the same number of poles as zeros. Unfortunately, and unlike the polynomial case (which can be viewed as a special case where $k = N - 1$), we do not have an explicit formula for $P_k(\cos\omega)$.

For the values $N = 4, 8, 12, 16, \ldots$, we have used $k(N) = N/2 + 1$ so that $N + k(N)$ is odd.

Figure 5 illustrates the evolution of $s(\varphi_N)$ for these particular intermediate solutions and compares it with the graphs obtained for polynomials and Butterworth solutions.

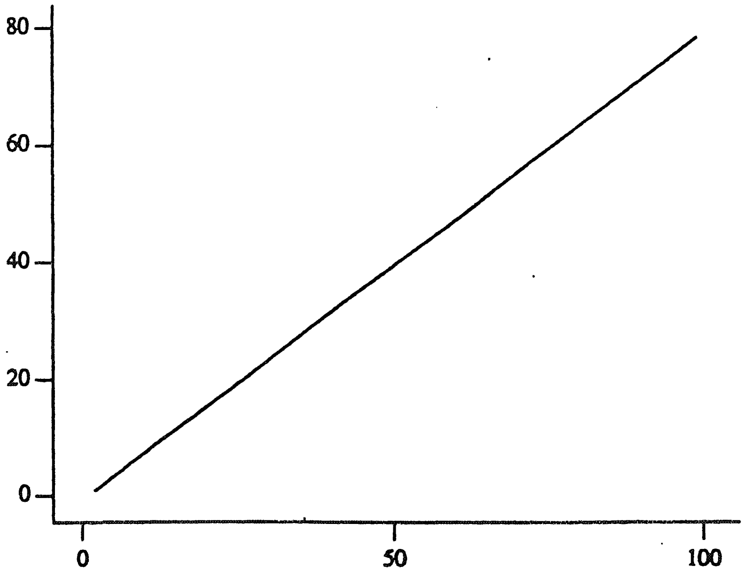


Figure 5. $s(\varphi_N)$ for $0 < N \leq 100$ (Butterworth).

A straightforward observation is that, although these intermediate solutions contain the same number of poles and zeros, the graph of $s(\varphi_N)$ is very close to the graph obtained for Butterworth scaling functions, making these particular discrete rational filters interesting for applications where regularity is desirable.

6.3. Nonseparable bidimensional scaling functions.

The simplest way to generate multivariate scaling functions is to use the tensor product, *i.e.* to define

$$\Phi(x_1, \ldots, x_n) = \varphi_1(x_1) \cdots \varphi_n(x_n), \tag{6.113}$$

where $\varphi_1, \ldots, \varphi_n$ are univariate refinable functions. Note that if the univariate functions are cardinal interpolant or orthonormal, then the same property holds for Φ. The analysis of the regularity of Φ then follows directly from the univariate analysis on the φ_j's.

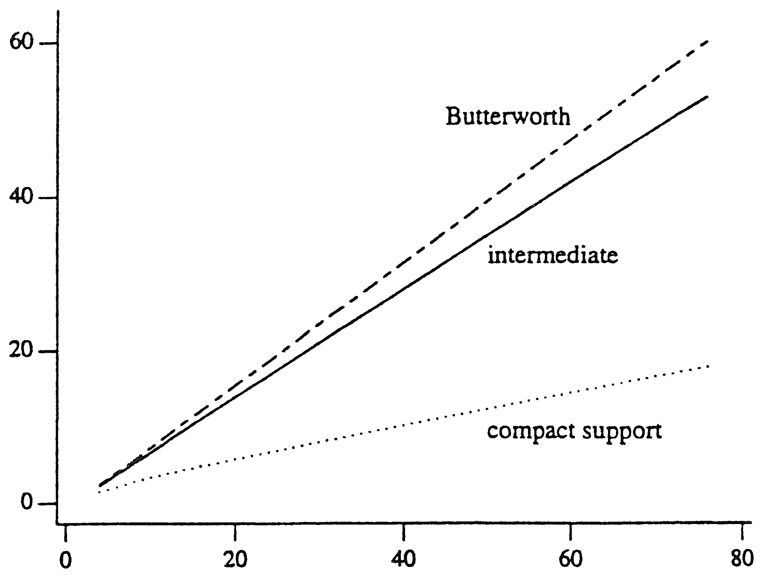


Figure 6. $s(\varphi_N)$, $N \leq 75$ for polynomial, Butterworth and intermediate solution.

One of the simplest -yet instructive- situations where non-separable scaling functions are unavoidable corresponds to the choice

$$D = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \tag{6.14}$$

for the dilation matrix, already introduced in Section 3 (Example 3.8). In that case the interpolatory condition has the following formulation:

$$M(\omega_1, \omega_2) + M(\omega_1 + \pi, \omega_2 + \pi) = 1\,. \tag{6.15}$$

One can use the univariate functions $m_N(\omega)$, defined in the first subsection, to derive a solution of (6.15) as follows:

$$M_N(\omega_1, \omega_2) = \big(c(\omega_1, \omega_2)\big)^N \sum_{j=0}^{N-1} \binom{N-1+j}{j} \big(s(\omega_1, \omega_2)\big)^j\,, \tag{6.16}$$

where

$$c(\omega_1, \omega_2) = \frac{1}{2}\left(\cos^2\left(\frac{\omega_1}{2}\right) + \cos^2\left(\frac{\omega_2}{2}\right)\right)$$

and

$$s(\omega_1, \omega_2) = \frac{1}{2}\left(\sin^2\left(\frac{\omega_1}{2}\right) + \sin^2\left(\frac{\omega_2}{2}\right)\right).$$

We denote by Φ_N the nonseparable cardinal interpolant functions associated to M_N. Note however that the Riesz factorization lemma does not generalize in nD, $n > 1$, so that it is not possible to derive compactly supported orthonormal scaling functions from these Φ_N. Using the preliminary results of Example 3.8, we can compute the Hölder exponents $\mu(\Phi_N) = s_1(\Phi_N)$ as well as $s_p(\Phi_N)$ for $p = 2, 4, 8$. We display their values, for $N = 1, \ldots, 19$ in Table 3. As in the univariate case, this table reveals the increasingly lacunary structure of the functions Φ_N in the Fourier domain.

$N \setminus p$	1	2	4	8
1	0.611268	1.575915	1.939386	1.981617
2	2.285413	3.249338	3.684182	3.862247
3	3.881443	4.778977	5.146044	5.274342
4	5.395644	6.199651	6.487579	6.582461
5	6.841235	7.549708	7.780608	7.868502
6	8.233367	8.854675	9.054724	9.141341
7	9.584611	10.132740	10.318169	10.404515
8	10.904434	11.395230	11.573954	11.660232
9	12.200067	12.648141	12.823734	12.909992
10	13.477248	13.894538	14.068620	14.154872
11	14.740529	15.136095	15.309421	15.395672
12	15.993417	16.373816	16.546755	16.633005
13	17.238515	17.608370	17.781105	17.867355
14	18.477713	18.840232	19.012859	19.099109
15	19.712362	20.069763	20.242333	20.328582
16	20.943434	21.297249	21.469787	21.556036
17	22.171630	22.522919	22.695440	22.781690
18	23.397462	23.746966	23.919477	24.005727
19	24.621314	24.969550	25.142055	25.228305

Table 3. L^p-Sobolev exponent of Φ_n, $p = 1, 2, 4, 8$, $N = 1, \ldots, 19$.

6.4. A problem on spline wavelet bases.

Spline wavelets are generated by a function ψ that is piecewise polynomial (of a fixed degree d) on each integral $[k/2,(k+1)/2[$, $k \in \mathbb{Z}$. There exis' ` several types of spline wavelets:

a) Fully orthonormal wavelets: the family $\{\psi_{j,k}\}_{j,k\in\mathbb{Z}}$ constitutes an orthonormal basis of $L^2(\mathbb{R})$ (Battle (1987), Lemarié (1988)). In that case, ψ cannot be compactly supported except when $d=0$, *i.e.* in the case of piecewise constant functions corresponding to the Haar system. (Note that in a generalized framework, where several scaling functions and wavelets are considered, even for the one dimensional case and dilation factor 2, compactly supported orthonormal wavelets are possible; see Donovan, Geronimo and Hardin (1994).)

b) Semi-orthonormal wavelets: the functions $\psi_{j,k}$ are orthogonal between levels $j \neq j'$ but not within one level j for $k \neq k'$ (Chui-Wang 1990). In that case ψ can be compactly supported but the dual function $\tilde{\psi}$ that generates the dual wavelet basis ($\langle\tilde{\psi}_{j,k},\psi_{j',k'}\rangle = \delta_{j,j'}\,\delta_{k,k'}$) is a noncompactly supported spline function.

c) Biorthogonal wavelet basis: $\{\psi_{j,k}\}_{j,k\in\mathbb{Z}}$ is a Riesz basis for $L^2(\mathbb{R})$ and there exists a dual system $\{\tilde{\psi}_{j,k}\}_{j,k\in\mathbb{Z}}$ as well as a dual scaling function $\tilde{\varphi}$ and multiresolution analysis $\tilde{V}_j$ (Cohen, Daubechies, and Feauveau (1992)). The functions $\varphi,\tilde{\varphi},\psi,\tilde{\psi}$ may be simultaneously compactly supported, but $\tilde{\varphi}$ and $\tilde{\psi}$ are not spline functions in general.

Note that each construction is a particular case of the next one, and that in all cases the wavelet ψ has the expression

$$\psi(x) = \sum_{n\in\mathbb{Z}} g_n\,\varphi(2n-n)\,, \tag{6.17}$$

where $\varphi = \chi_{[0,1]} * \chi_{[0,1]} \cdots * \chi_{[0,1]}$ ($d+1$ times) is the box spline of degree d, and g_n is an oscillating ℓ^2 sequence, *i.e.* $\sum_n g_n = 0$.

One can then address the following general problem: given an arbitrary oscillating sequence g_n, when does the corresponding combination (6.17) of box-splines generate a (Riesz) wavelet basis $\{\psi_{j,k}\}_{j,k\in\mathbb{Z}}$? In particular, we have in mind very simple sequences such as $g_0=1$, $g_1=-1$ or $g_0=-1$, $g_1=2$, $g_2=-1$, etc.

First, note that ψ can be written in the Fourier domain as

$$\hat{\psi}(\omega) = m_1\left(\frac{\omega}{2}\right)\hat{\varphi}\left(\frac{\omega}{2}\right), \tag{6.18}$$

where $m_1(\omega) = \sum_n g_n\, e^{-in\omega}$. Moreover, we have

$$\hat{\varphi}(\omega) = \prod_{k=1}^{+\infty} m_0(2^{-k}\omega), \tag{6.19}$$

where $m_0(\omega) = \left((1+e^{-i\omega})/2\right)^d$. In the case of biorthogonal wavelets (*i.e.* type c) above), $m_1(\omega)$ is equal to $e^{-i\omega}\,\overline{\tilde{m}_0(\omega+\pi)}$, where $\tilde{m}_0$ generates the dual scaling function $\tilde{\varphi}$ in the sense that

$$\hat{\tilde{\varphi}}(\omega) = \prod_{k=1}^{+\infty} \tilde{m}_0(2^{-k}\omega). \tag{6.20}$$

The biorthogonality constraint is expressed by the equation

$$\overline{m_0(\omega)}\,\tilde{m}_0(\omega) + \overline{m_0(\omega+\pi)}\,\tilde{m}_0(\omega+\pi) = 1, \tag{6.21}$$

or, equivalently,

$$e^{i\omega}\left(m_0(\omega)\,m_1(\omega+\pi) - m_0(\omega+\pi)\,m_1(\omega)\right) = 1. \tag{6.22}$$

It is clear that equation (6.22) is a strong restriction on m_1. Given a solution of (6.22), one can however construct other m_1 that still give rise to Riesz bases $\psi_{j,k}$. It suffices to take $m_1(\omega) = m(2\omega)M_1(\omega)$ where $M_1(\omega)$ satisfies equation (6.22) and $m(\omega)$ is a 2π-periodic function such that

$$0 < c \leq |m(\omega)| \leq C < \infty, \tag{6.23}$$

almost everywhere with respect to $\omega \in \mathbb{R}$. This corresponds to the choice

$$\hat{\psi}(\omega) = m(\omega)\,M_1\left(\frac{\omega}{2}\right)\hat{\varphi}\left(\frac{\omega}{2}\right). \tag{6.24}$$

We can then define Ψ by $\hat{\Psi}(\omega) = M_1(\omega/2)\,\hat{\varphi}(\omega/2)$, and use the biorthogonal theory to study if $\{\Psi_{j,k}\}_{j,k\in\mathbb{Z}}$ is a Riesz basis of $L^2(\mathbb{R})$. If this is the case, then the same clearly holds for $\{\psi_{j,k}\}_{j,k\in\mathbb{Z}}$. Note that

$m(\omega)$ and $M_1(\omega)$ are completely determined from $m_1(\omega)$ since we have $M_1(\omega) = m_1(\omega)/m(2\omega)$ and thus, by (6.22),

$$m(2\omega) = e^{i\omega}\left(m_0(\omega)\,m_1(\omega+\pi) - m_0(\omega+\pi)\,m_1(\omega)\right). \tag{6.25}$$

The system $\{\psi_{j,k}\}_{j,k\in\mathbb{Z}}$ will thus constitute a Riesz basis if the two following conditions are satisfied:

i) the function $m(\omega)$ defined by (6.25) is bounded below and above by strictly positive constants.

ii) $\{\Psi_{j,k}\}_{j,k\in\mathbb{Z}}$ is a Riesz basis. A necessary and sufficient condition for this to hold was given in Cohen and Daubechies (1992). In our context this results in the following

Theorem 6.1. *Let*

$$\tilde{M}_0(\omega) = -e^{-i\omega}\,\overline{M_1(\omega+\pi)} \qquad \text{and} \qquad M(\omega) = \frac{|\tilde{M}_0(\omega)|^2}{\cos^2(\omega/2)}\,.$$

Assume that the Fourier coefficients of M satisfy the decay condition (3.3). *Define d as the transition operator associated to $M(\omega)$ and denote by ρ its spectral radius on a space E_α for any $\alpha\in]\gamma,2\gamma[$. Then $\{\Psi_{j,k}\}_{j,k\in\mathbb{Z}}$ constitutes a Riesz basis of $L^2(\mathbb{R})$ if and only if $\rho<4$.*

Note that, according to our results, this condition means that the L^2-Sobolev exponent of the scaling function $\tilde{\Phi}$ associated to $\tilde{M}_0$ is strictly positive.

Since we have

$$|\tilde{M}_0(\omega)|^2 = \frac{|m_1(\omega+\pi)|^2}{|m(2\omega)|^2}\,, \tag{6.26}$$

the function $M(\omega)$ is not a trigonometric polynomial in general, even when $\{g_n\}$ is a finite sequence. This made this application inaccessible to earlier methods that could only deal with finite masks.

An immediate application concerns the case of linear splines, *i.e.* $\varphi(x) = \sup\{0, 1-|x|\}$. In that case, we can propose three simple wavelets corresponding to different choices for the g_n coefficients:

- $\psi_a(x) = \varphi(2x) - \varphi(2x-1)$,
- $\psi_b(x) = 2\,\varphi(2x) - \varphi(2x-1) - \varphi(2x+1)$,

- $\psi_c(x) = 2\,\varphi(2x-1) - \varphi(2x) - \varphi(2x-2)$.

One can easily check that for ψ_b, the associated function $m(\omega)$ vanishes at some point. It is also easy to check that this implies that Span $\{\psi_b(x-k)\}_{k\in\mathbb{Z}}$ cannot complement in a stable manner the space V_0 into V_1.

In the cases of ψ_a and ψ_c, the associated function $m(\omega)$ does not vanish, so that we can further investigate the associated functions $M(\omega)$.

For ψ_a, we find

$$M(\omega) = \frac{16}{10 + 6\cos\omega}\,, \tag{6.27}$$

and our method shows that $\rho > 4$. For ψ_c we find

$$M(\omega) = \cos^2\left(\frac{\omega}{2}\right)\left(\frac{4}{3+\cos\omega}\right)^2, \tag{6.28}$$

and in that case $\rho < 4$.

CONCLUSION: From the three functions above, only ψ_c generates a Riesz wavelet basis.

7. Extension to the computation of the Hölder exponent.

The spaces E_α introduced in Section 3 have shown to be an excellent tool for the computation of s_2 in general, and of s_p for $p \neq 2$ if the h_n satisfy some additional conditions. Could a similar argument also be used for computing the Hölder exponent?

Let us restrict ourselves, for this discussion, to the one-dimensional case of equation (4.1). To start with, assume that only finitely many c_n differ from zero, $c_n = 0$ for $n < 0$ or $n > K$. In this case Daubechies and Lagarias (1992) gave the following technique for computing the Hölder exponent μ of φ. First, we factor out all the zeros at $\omega = \pi$ of $m(\omega)$,

$$\sum_{n=0}^{K} c_n\, e^{-in\omega} = \left(\frac{1+e^{-i\omega}}{2}\right)^N \sum_{n=0}^{K-N} q_n\, e^{-in\omega}\,;$$

then we construct two $(K-N)\times(K-N)$ matrices T_0 and T_1 by taking

$$(T_j)_{k,\ell} = 2\,q_{2k-\ell+j}\,, \qquad 0 \le k,\ell \le K-N-1\,. \tag{7.1}$$

If there exist ν, $C > 0$ so that, for all $k \in \mathbb{N}\backslash\{0\}$ and all $d_1, \ldots, d_k$ chosen in $\{0, 1\}$,

$$\|T_{d_1} \cdots T_{d_k}\| \le C\, 2^{\nu k}\,, \tag{7.2}$$

then $\mu \ge N - \nu$. (Note that we have changed notations slightly with respect to Daubechies and Lagarias (1992); we use the reduced representation that was there introduced in Section 5.) How can we distill from this a strategy that can be generalized to the case where infinitely many c_n are nonzero? First of all, note that T_0, T_1 can be related simply to the operator $\mathcal{L}_w$ corresponding to the choice $w(\omega) = q(\omega)$. Comparing (7.1) with (3.4) we find indeed, for any $f(\omega) = \sum_{n=0}^{K-N} f_n\, e^{-in\omega}$ in the space $\mathcal{P}_{K-N}$ of one-sided trigonometric polynomials of degree $K - N$, that

$$\begin{aligned}(\mathcal{L}f)(\omega) &= \sum_{n=0}^{K-N} (T_0 f)_n\, e^{-in\omega}\,,\\ (\mathcal{L}Sf)(\omega) &= \sum_{n=0}^{K-N} (T_1 f)_n\, e^{-in\omega}\,,\end{aligned} \tag{7.3}$$

where we have introduced the shift operator S,

$$S\Big(\sum_\ell g_\ell\, e^{-i\ell\omega}\Big) = \sum_\ell g_{\ell+1}\, e^{-i\ell\omega} = e^{i\omega}\Big(\sum_\ell g_\ell\, e^{-i\ell\omega}\Big)\,. \tag{7.4}$$

From (3.1) one easily checks that $S\mathcal{L} = \mathcal{L}S^2$. The condition (7.2) can therefore be rewritten as: for all $k \in \mathbb{N}\backslash\{0\}$ and all n, $0 \le n \le 2^k - 1$,

$$\|\mathcal{L}^k S^n\, |_{\mathcal{P}_{K-N}}\,\| \le C\, 2^{\nu k}\,. \tag{7.5}$$

In this form it is easy to generalize (7.2): we could just drop the restriction to $\mathcal{P}_{K-N}$ in (7.5). There is one problem: in (7.5) it doesn't matter which norm we take, because for all $0 \le n \le 2^k - 1$ the operators $\mathcal{L}^k S^n$ map $\mathcal{P}_{K-N}$ to itself, so that we are dealing with a norm on matrices, and all matrix norms are equivalent. Once we look at the case of infinitely many nonvanishing c_n, and we drop the no longer relevant restriction to a finite-dimensional polynomial space, we need to specify which operator norm to use in (7.5). There is in fact a lot of freedom in the choice of this norm; in particular, if E is a Banach space of 2π-periodic functions such that

$$\sup_n |f_n| = \sup_n \left|\frac{1}{2\pi}\int_0^{2\pi} e^{in\omega}\, f(\omega)\, d\omega\right| \le C\, \|f\|_E\,, \tag{7.6}$$

then the operator norm

$$(7.7) \qquad |||A|||_E = \sup_{f \in E,\, \|f\|_E \neq 0} \frac{\|Af\|_E}{\|f\|_E}$$

will do. That is, if for all $k \in \mathbb{N}\backslash\{0\}$, all $0 \le n \le 2^k - 1$,

$$(7.8) \qquad |||\mathcal{L}^k S^n|||_E \le C\, 2^{\nu k},$$

then it will follow that φ has Hölder exponent at least $N - \nu$. The connection between (7.6), (7.8) and this Hölder continuity is explained in the Appendix. Note that (7.8) implicitly assumes that both S and $\mathcal{L}$ map E to itself.

Candidates for spaces E that satisfy (7.6) abound. Examples are all the E_α of Section 3, as well as the $L^p(0, 2\pi)$-spaces, or the C^ν-spaces of 2π-periodic functions that have Hölder exponent ν, and the C^n-spaces of n times continuously differentiable periodic functions (including C^0). The ℓ^p-spaces,

$$\ell^p = \left\{ f \ 2\pi\text{-periodic} : \ \|f\|_{\ell^p} = \Big(\sum_n |f_n|^p \Big)^{1/p} < \infty \right\},$$

also satisfy (7.6). For which of these spaces can we hope to verify (7.8) for some ν?

The spaces E_α, so convenient for the computation of the s_p, are completely useless here. Because $|||S^n|||_{E_\alpha} = |||S^{-n}|||_{E_\alpha} = e^{\alpha|n|}$, we have

$$|||\mathcal{L}^k S^n|||_{E_\alpha} = |||\mathcal{L}^k|||_{E_\alpha}\, e^{\alpha|n|};$$

since $|n|$ can be as large as $2^k - 1$, the only $\mathcal{L}$ for which (7.8) can hold in E_α is the zero operator. In some sense, the E_α-spaces are "too small" for our present purpose: their norm gets affected too much by S.

No such problem exists in the L^p, ℓ^p and C^0-spaces: they all share the property that $|||S|||_E = 1$. This reduces the estimate (7.8) to a spectral radius problem again: it suffices to prove that $\rho_E(\mathcal{L}) < 2^\nu$ in order to conclude that φ has Hölder exponent at least $N - \nu$. If the space E is chosen "too large", then we get a bad estimate for ν, however. Take for example

$$m(\omega) = \Big(\frac{1 + e^{-i\omega}}{2}\Big)^2 \frac{1 + \gamma e^{i\omega}}{1 + \gamma} \qquad \text{with } \gamma > 1;$$

using the techniques of Daubechies and Lagarias (1992) one readily shows that the corresponding φ has Hölder exponent exactly equal to $2-\log_2(2\gamma/(1+\gamma))$. On the other hand, it is easy to find eigenvectors for the corresponding $\mathcal{L}$ in the ℓ^p-spaces. We have $q(\omega) = (1+\gamma e^{i\omega})/(1+\gamma)$, or

$$\mathcal{L}f = \mu f \iff f_{2n} + \gamma f_{2n+1} = \frac{1}{2}\mu(1+\gamma) f_n . \tag{7.9}$$

Let now $\rho \in \mathbb{C}$ be arbitrary (to be fixed below), and define f_n by

$$\begin{aligned} f_n &= 0, \qquad n \le 0, \\ f_1 &= 1, \\ f_{2n} &= \nu\Big(\frac{1+\gamma}{2} - \gamma\rho\Big) f_n, \qquad n \ge 1, \\ f_{2n+1} &= \rho\nu f_n, \qquad n \ge 1; \end{aligned} \tag{7.10}$$

then the f_n obviously satisfy (7.9) with $\mu = \nu$. If, for some $\nu \in \mathbb{C}$, we can find $\rho \in \mathbb{C}$ so that the f_n defined by (7.10) satisfy $\sum_n |f_n|^p < \infty$, then ν is an eigenvalue for $\mathcal{L}$ in ℓ^p, and $\rho_{\ell^p}(\mathcal{L}) \ge |\nu|$. Let us check when this is true. First of all, note that

$$\begin{aligned} \sigma_N &= \sum_{n=2^N}^{2^{N+1}-1} |f_n|^p \\ &= \sum_{n=2^{N-1}}^{2^N-1} \left(|f_{2n}|^p + |f_{2n+1}|^p\right) \\ &= |\nu|^p \left(\left|\frac{1+\gamma}{2} - \gamma\rho\right|^p + |\rho|^p\right) \sigma_{N-1}, \end{aligned}$$

so that

$$\sum_{n=-\infty}^{\infty} |f_n|^p = \sum_{k=0}^{\infty} |\nu|^{pk} \left(\left|\frac{1+\gamma}{2} - \gamma\rho\right|^p + |\rho|^p\right)^k;$$

this is finite if and only if

$$|\nu|^p \left(\left|\frac{1+\gamma}{2} - \gamma\rho\right|^p + |\rho|^p\right) < 1.$$

Consequently

$$\rho_{\ell^p}(\mathcal{L}) \ge \max_{\rho\in\mathbb{C}} \left(\left|\frac{1+\gamma}{2} - \gamma\rho\right|^p + |\rho|^p\right)^{-1/p}.$$

In particular,

$$\rho_{\ell^2}(\mathcal{L}) \geq \frac{2\sqrt{1+\gamma^2}}{1+\gamma}\,, \qquad \rho_{\ell^1}(\mathcal{L}) \geq \frac{2\gamma}{1+\gamma}\,. \tag{7.11}$$

In fact, the lower bound for ρ_{ℓ^1} is exact, *i.e.* $\rho_{\ell^1}(\mathcal{L}) = 2\gamma/(1+\gamma)$. It is then clear from (7.11) that the spectral radius of $\mathcal{L}$ in the larger space ℓ^2 is strictly larger, leading to a nonoptimal estimate of the Hölder exponent μ of φ. The same happens in the other ℓ^p-spaces with $p > 1$.

This example teaches us that it is important to choose the space E carefully. Note that the techniques in the literature for estimating μ can all be viewed in this way. The approaches of Rioul (1992) or Dyn and Levin (1991), Dyn (1991) correspond to estimates of the type

$$|||(\mathcal{L}^*)^k|||_{\ell^\infty} \leq C\,2^{\nu k}\,,$$

which is equivalent with

$$|||\mathcal{L}^k|||_{\ell^1} \leq C\,2^{\nu k}\,,$$

i.e. this corresponds to the choice $E = \ell^1$. In part of Hervé (1995), the choice $E = C^0$ is treated, a slightly larger space than ℓ^1. One can show that the choices $E = \ell^1$ or $E = C^0$ lead to optimal values for μ (see Hervé (1995), Rioul (1992)). The example above also shows, however, that the operators $\mathcal{L}$ on ℓ^1 or C^0 are far from their compact restrictions on the E_α; in our example the entire disk $B(0, \rho_{\ell^1}) = \{z : |z| < \rho_{\ell^1}\}$ consists of eigenvalues of $\mathcal{L}$. This means that many iterative techniques, which usually rely on the fact that the largest eigenvalue is isolated, cannot be applied then.

So far we have seen that the E_α are "too small", the ℓ^p with $p > 1$ "too large" for our purposes; ℓ^1 and C^0 are fine, but the spectrum of $\mathcal{L}$ on these spaces can consist of a disk of unisolated eigenvalues. It turns out that one can use slightly smaller spaces on which the largest eigenvalue of $\mathcal{L}$ becomes again isolated, and can be computed via the zeros of $\det(1 - z\mathcal{L})$. (These are, in fact, the spaces used for the computation of s_p in Hervé (1995), who exploits the fact that the largest eigenvalue is isolated.) This is a consequence of a theorem in Ruelle (1990), of which the following statement is a special case, restricted to the situation under consideration here.

Theorem 7.1. *Let q be a 2π-periodic function satisfying* (4.4), *and let $\mathcal{L}$ be the associated transfer operator (obtained by replacing w by q in* *the definition* (3.1)). *Denote by* $\mathcal{K}$ *the operator obtained by replacing* w *by* $|q|$ *in* (3.1). *Let* ρ *be the spectral radius of* $\mathcal{K}$ *on* $C^0(0,2\pi)$. *Then, for any* $\alpha > 0$, *the spectral radius* ρ_α *of* $\mathcal{K}$ *on* $C^\alpha(0,2\pi)$ *equals* ρ, $\rho_\alpha = \rho$; *the spectral radius* σ_α *of* $\mathcal{L}$ *on* $C^\alpha(0,2\pi)$ *satisfies* $\sigma_\alpha \leq \rho$. *Moreover, the part of the spectrum of* $\mathcal{L}$ *on* $C^\alpha(0,2\pi)$ *that is contained in* $\{\lambda : \ |\lambda| > 2^{-\alpha}\rho\}$ *consists of only eigenvalues with finite multiplicities; these eigenvalues are exactly the inverses of the zeros of the Fredholm determinant* $D(z)$,

$$(7.12) \quad D(z) = \exp\Big(-\sum_{m=1}^{\infty} \frac{1}{m} z^m \frac{1}{1-2^{-m}} \sum_{k=0}^{2^m-1} \prod_{\ell=0}^{m-1} q\Big(2^\ell \frac{2\pi k}{2^m-1}\Big)\Big)$$

in the region $\{z : \ |z| < 2^\alpha \rho^{-1}\}$, *with the same multiplicities.*

If $q(\omega) \geq 0$ for $\omega \in [0,2\pi]$, then this theorem already implies that we can simply look for the smallest zero z_0 of (7.12), exactly like we did before. For any $\epsilon > 0$ it then follows that

$$|||\mathcal{L}^k|||_{C^0} \leq C\,(\rho+\epsilon)^k\,,$$

with $\rho = |z_0|^{-1}$, leading to the estimate $\mu \geq N - \log_2 \rho$ for the Hölder exponent of φ. This is no surprise however: if $q(\omega) \geq 0$, then $|q(\omega)| = q(\omega)$, hence $\rho = r_1$, and $s_1 = N - \log_2 r_1 = N - \log_2 \rho$. Since on the other hand

$$\hat{\varphi}(\omega) = \Big(\frac{1-e^{-i\omega}}{i\omega}\Big)^N \prod_{j=1}^{\infty} q(2^{-j}\omega)\,,$$

we have either

$$e^{i\omega L}\hat{\varphi}(\omega) \geq 0\,, \qquad \text{if } N = 2L \text{ is even,}$$

or

$$e^{iL\omega}\frac{1-e^{-i\omega}}{2}\,\hat{\varphi}(\omega) \geq 0\,, \qquad \text{if } N = 2L-1 \text{ is odd.}$$

It is well known from Littlewood-Paley theory that if $\hat{f}(\omega) \geq 0$ for all ω, then the Hölder exponent of f is exactly equal to the Sobolev exponent $s_1(f)$. It then follows easily, whether N is even or odd, that the Hölder exponent μ of φ is given by $\mu = N - \log_2 \rho$.

If q can also take negative values, then it follows from $\sigma_\alpha \leq \rho$ that for all $\epsilon > 0$,

$$|||\mathcal{L}^k|||_{C^\alpha} \leq C\,(\rho+\epsilon)^k\,.$$

Since $|||S^n|||_{C^\alpha} = |||S^{-n}|||_{C^\alpha} = |n|^\alpha$, we have therefore, for $0 \le n \le 2^k - 1$,

$$|||\mathcal{L}^k S^n|||_{C^\alpha} = |||\mathcal{L}^k|||_{C^\alpha} \, |n|^\alpha \le C\,(2^\alpha(\rho + \epsilon))^k .$$

Here $\alpha, \epsilon > 0$ can be chosen arbitrarily small, so that we have $\mu \ge N - \log_2 \rho$. This bound need not be sharp however: σ_α may well be smaller than ρ for arbitrarily small α. On the other hand, Theorem 7.1 also tells us that as we increase α, the bothersome essential part of the spectrum of $\mathcal{L}$ on C^α shrinks, and at some point eigenvalues are uncovered which correspond to zeros of (7.12), which we can compute accurately. Let us imagine increasing α until Theorem 7.1 guarantees us that the spectral radius σ_α is exactly given by $|z_0|^{-1}$, with z_0 the smallest zero of (7.12). This happens when $2^{-\alpha}\rho \le |z_0|^{-1} < 2^{-\alpha}\rho + \delta$ for some small $\delta > 0$; we have then

$$\begin{aligned} |||\mathcal{L}^k S^n|||_{C^\alpha} &\le |||\mathcal{L}^k|||_{C^\alpha} \, 2^{k\alpha} \\ &\le C\, 2^{k\alpha}\,(|z_0|^{-1} + \epsilon)^k \\ &\le C\,(|2^{-\alpha} z_0|^{-1} + \epsilon')^k \\ &\le C\,(\rho + \epsilon'')^k , \end{aligned}$$

leading again to the same estimate $\mu \ge N - \log_2 \rho$.

This discussion has given a unified picture of the techniques used to find Sobolev and Hölder regularity indices for refinable functions. It also shows that if w is given, and we can prove, for some $\alpha > 0$, that $\sigma_\alpha < 2^{-\alpha}\rho$, then this would lead to a sharper estimate for the Hölder exponent than $s_1(|w|)$; since σ_α corresponds to the spectral radius on a smaller space than C^0 or ℓ^1, it might be easier to tackle σ_α than σ_0. How to do this is an open question however.

Acknowledgment. This work was the direct result of conversations of one of us (I. Daubechies) with David Ruelle, during a visit to IHES in June 1992. We gratefully acknowledge his input. Combining the writing up of this paper with our busy schedules was more difficult than we expected, and it would have taken even longer without the hospitality of various institutes that made it possible for us to get together at various times. We would like to thank all of these institutes; IHES (ID, June '92), AT&T Bell Laboratories (AC, March '93 and July '94), Univ. Bielefeld (ID, August '93).

Appendix. The link between estimates on $\mathcal{L}^k S^n$ and the Hölder exponent of φ.

Let E be a Banach space of 2π-periodic functions such that, for all f in E,

$$\sup_n |f_n| = \sup_n \left| \frac{1}{2\pi} \int_0^{2\pi} e^{in\omega} f(\omega)\, d\omega \right| \leq C \, \|f\|_E \,, \tag{A.1}$$

where C is independent of f. We assume that E contains the function $f(\omega) \equiv 1$, and that E is also invariant for the shift operator S, defined by

$$(Sf)(\omega) = e^{i\omega} f(\omega)\,,$$

or by

$$(Sf)_n = f_{n+1} \,.$$

For an operator A from E to itself, we denote by $\|A\|_E$ its standard operator norm, $\|A\|_E = \sup_{f\in E, f\neq 0} \|Af\|_E / \|f\|_E$. Let φ be a refinable function,

$$\varphi(x) = 2 \sum_n c_n \, \varphi(2x - n) \tag{A.2}$$

and assume that

$$\sum_n c_n \, e^{-in\omega} = \Big(\frac{1 + e^{-i\omega}}{2}\Big)^N \sum_n w_n \, e^{-in\omega} \,, \tag{A.3}$$

where $\sum_n w_n = 1$, $\sum_n (-1)^n w_n \neq 0$. Assume that the associated transfer operator $\mathcal{L}_w$ (defined by (3.1)) maps E to itself. We have one last technical requirement. Let us consider $N = 1$, for simplicity. If we compute $\mathcal{L}_c 1$ formally, where $\mathcal{L}_c$ is the transfer operator associated with $c(\omega) = \sum_n c_n \, e^{-in\omega}$, then

$$\begin{aligned}(\mathcal{L}_c 1)(\omega) &= \Big(\frac{1 + e^{-i\omega/2}}{2}\Big)\, w\big(\frac{\omega}{2}\big) + \Big(\frac{1 - e^{-i\omega/2}}{2}\Big)\, w\big(\frac{\omega}{2} + \pi\big) \\ &= \Big[\mathcal{L}_w \Big(\frac{1+S}{2}\Big) 1\Big] (\omega)\,,\end{aligned}$$

and $(\mathcal{L}_c 1)(0) = w(0) = 1$. It follows that $(\mathcal{L}_c 1) - 1$ has a zero in $\omega = 0$; we shall require that

$$(\mathcal{L}_c 1 - 1)(\omega) = (1 - e^{-i\omega})\, g(\omega)\,, \tag{A.4}$$

where $g \in E$, which can also be written as

$$\mathcal{L}_w\Big(\frac{1+S}{2}\Big)1 - 1 \in (1-S)E\,.$$

For general N, this requirement takes the form that for $1 \le k \le N$, we should have

$$\text{(A.5)}\quad \mathcal{L}_w(1+S)^k 1 \in \operatorname{Span}\{(1-S)^k E,\ (1-S)^m 1, \text{ with } 0 \le m \le k-1\}\,.$$

Then we have the following

Proposition. *Let E, w satisfy all the conditions above. If there exist $c > 0$ and $0 \le \nu < N$ so that, for all $k > 0$ and all n between 0 and 2^k,*

$$\text{(A.6)}\qquad \|\mathcal{L}_w^k\, S^n\|_E \le C\,2^{\nu k}\,,$$

then φ has Hölder exponent $\mu \ge N - \nu$.

PROOF. The proof is essentially a generalization of the arguments in Daubechies and Lagarias (1992), adapted to the case with infinitely many coefficients. This means that alternatives have to be found for some matrix arguments in Daubechies and Lagarias (1992). We shall restrict ourselves here to the case $N = 1$, and discuss the proof in detail for this case; for general N, similar but slightly longer generalizations of Section 3 in Daubechies and Lagarias (1992) do the trick.

1). We start by defining a space $\tilde{E}$ by

$$\tilde{E} = \{f :\ f(\omega) = (1 - e^{-i\omega})\,g(\omega) + C \text{ with } c \in \mathbb{C} \text{ and } g \in E\}\,;$$

the norm on $\tilde{E}$ is simply

$$\|c + (1 - e^{-i\omega})\,g\|_{\tilde{E}} = |c| + \|g\|_E\,.$$

$\tilde{E}$ is clearly a Banach space which contains all trigonometric polynomials (since $1 \in E$ and E is invariant under S). In $\tilde{E}$ we consider the subspace $\tilde{E}_1$ defined by

$$\tilde{E}_1 = \{f \in \tilde{E} :\ \sum_n f_n = f(0) = 0\}\,;$$

$\tilde{E}_1$ can be identified with the original space E, since

$$\tilde{E}_1 = (1 - e^{-i\omega})E\,.$$

2). Let $\mathcal{L}_c$ be the transfer operator associated with the weight function $c(\omega) = \sum_n c_n\, e^{-in\omega}$. Because of the factorization (A.3) we have, for $f \in \tilde{E}_1$,

$$\begin{aligned}(\mathcal{L}_c f)(\omega) &= \Big(\frac{1+e^{-i\omega/2}}{2}\Big)\, w\big(\frac{\omega}{2}\big)\,(1-e^{-i\omega/2})\,g\big(\frac{\omega}{2}\big)\\ &\quad+ \Big(\frac{1-e^{-i\omega/2}}{2}\Big)\, w\big(\frac{\omega}{2}+\pi\big)\,(1+e^{-i\omega/2})\,g\big(\frac{\omega}{2}+\pi\big)\\ &= \frac{1}{2}(1-e^{-i\omega})\,(\mathcal{L}_w g)(\omega)\,.\end{aligned}$$

$\tilde{E}_1$ is thus invariant for $\mathcal{L}_c$, and the action of $\mathcal{L}_c$ on $\tilde{E}_1$ is equivalent to the action of $\mathcal{L}_w$ on E. We have moreover for $f \in \tilde{E}_1$ and $k \geq 1$,

$$\|\mathcal{L}_c^k f\|_{\tilde{E}} = 2^{-k}\,\|\mathcal{L}_w^k g\|_E \leq 2^{-(1-\nu)k}\,\|f\|_{\tilde{E}}\,. \tag{A.7}$$

3). On the other component of $\tilde{E}$, the action of $\mathcal{L}_c$ is completely determined by $\mathcal{L}_c 1$. Because of (A.4), we have $\mathcal{L}_c 1 = 1 + r$, with $r \in \tilde{E}_1$. It follows then from (A.7) that

$$\mathcal{L}_c^k 1 = 1 + \sum_{m=1}^{k-1} \mathcal{L}_c^m r$$

converges, for $k \to \infty$, to $1 + R = a$, with $R \in \tilde{E}_1$; a is an eigenvector of $\mathcal{L}_c$ with eigenvalue 1.

4). Next we rewrite the refinement equation (A.2). For $x \in [0,1]$, we define the sequence-valued function $v(x)$ by

$$[v(x)]_n = \varphi(x+n)\,, \qquad n \in \mathbb{Z}\,.$$

Then (A.2) implies that

$$v(x) = \begin{cases} \mathcal{L}_c v(2x)\,, & \text{if } 0 \leq x \leq 1/2\,,\\ \mathcal{L}_c(Sv)(2x-1)\,, & \text{if } 1/2 \leq x \leq 1\,. \end{cases} \tag{A.8}$$

If, as in Daubechies and Lagarias (1992), we write $d_k(x)$ for the k-th digit in the binary expansion of x, then this becomes

$$v(x) = \mathcal{L}_c(S^{d_1(x)}v)(\sigma x)\,, \tag{A.9}$$

where $\sigma x = 2x$ if $x < 1/2$, $\sigma x = 2x - 1$ if $x \geq 1/2$. Smoothness for φ on $\mathbb{R}$ implies smoothness for v; conversely, smoothness for v on $[0,1]$ together with consistency conditions at the edges (of the style $[v(0)]_{n+1} = [v(1)]_n$) implies smoothness for φ. Solving (A.2) and proving smoothness for φ therefore amounts to finding a fixed point $v(x)$ for the equation (A.9) and proving smoothness for $v(x)$.

5). Define now $v_0(x)$ by

$$[v_0(x)]_n = a_n\,(1-x) + a_{n+1}\,x\,,$$

where a_n is the n-th component of the eigenvector a of $\mathcal{L}_c$ obtained in point 3). This clearly satisfies the consistency condition at $x = 0$ and $x = 1$; moreover $v_0(0)$ is an eigenvector of $\mathcal{L}_c$ with eigenvalue 1, as it should be, according to (A.8). We also define, for $j \geq 1$, and $x \in [0,1]$

$$\begin{aligned} v_j(x) &= \mathcal{L}_c S^{d_1(x)} v_{j-1}(\sigma x) \\ &= \mathcal{L}_c S^{d_1(x)} \mathcal{L}_c S^{d_2(x)} \cdots \mathcal{L}_c S^{d_j(x)} v_0(\sigma^j x)\,. \end{aligned} \tag{A.10}$$

Every component of $v_j(x)$ is a piecewise linear spline with nodes at the dyadic rationals $2^{-j}k$ in $[0,1]$. Since $S\mathcal{L}_c = \mathcal{L}_c S^2$, we can also rewrite (A.10) as

$$v_j(x) = \mathcal{L}_c^j S^{D_j(x)} v_0(\sigma^j x)\,, \tag{A.11}$$

where $D_j(x) = \sum_{\ell=1}^{j} 2^{j-\ell} d_\ell(x)$.

6). Now note that, for all $x \in [0,1]$,

$$\sum_n [v_0(x)]_n = \sum_n a_n = 1\,,$$

since $a = 1 + R$ with $R \in \tilde{E}_1$. Since this will be preserved by both $\mathcal{L}_c$ and S, this implies, for all $j \geq 0$ and all $x \in [0,1]$,

$$\sum_n [v_j(x)]_n = 1\,.$$

It follows that $v_{j+1}(x) - v_j(x) \in \tilde{E}_1$ for all $j \geq 0$ and all $x \in [0,1]$. Because of (A.6), (A.7) and (A.11) this implies

$$\|v_{j+1}(x) - v_j(x)\|_{\tilde{E}} \leq C\, 2^{-(1-\nu)j} \sup_{y\in[0,1]} \|v_1(y) - v_0(y)\|\,.$$

We can use this first to show that the $v_k(x)$ are uniformly bounded in $\tilde{E}$,

$$\begin{aligned}\|v_k(x)\|_{\tilde{E}} &\leq \|v_0(x)\|_{\tilde{E}} + \sum_{j=1}^{k-1} \|v_{j+1}(x) - v_j(x)\|_{\tilde{E}} \\ &\leq \|v_0(x)\|_{\tilde{E}} + \frac{C}{1-2^{-(1-\nu)}} \|v_1(x) - v_0(x)\|_{\tilde{E}} \ ;\end{aligned}$$

$\|v_0(x)\|_{\tilde{E}}$ is obviously bounded uniformly in x, and one easily checks that $\|v_1(x)\|_{\tilde{E}} = \|\mathcal{L}_c S^{d_1(x)} v_0(\sigma x)\|_{\tilde{E}}$ is as well. Next, we use the estimate again to prove that the $v_k(x)$ constitute a Cauchy sequence in $\tilde{E}$, uniformly in x,

$$\|v_{k+m}(x) - v_k(x)\| \leq C'' \, 2^{-(1-\nu)m} \, .$$

7). It follows that the $v_j(x)$ tend to a limit $v(x)$ in $\tilde{E}$, uniformly in x. If we "unfold" the $v_j(x)$ and $v(x)$ to define functions φ_j, φ by

$$\varphi_j(x) = (v_j(x - \lfloor x \rfloor))_{\lfloor x \rfloor}$$

(similarly for φ), then φ_j is piecewise linear with nodes at the $2^{-j}k$, $k \in \mathbb{Z}$, and, for any $x \in \mathbb{R}$, $x = n + y$ with $y \in [0, 1[$,

$$\begin{aligned}|\varphi_j(x) - \varphi(x)| &= |[v_j(y) - v(y)]_n| \\ &\leq C \, \|v_j(y) - v(y)\| \\ &\leq C' \, 2^{-(1-\nu)j} \, .\end{aligned}$$

where the second inequality follows because of property (A.1), which $\tilde{E}$ inherits from E. It then follows from standard results on approximation by splines that φ is Hölder continuous with exponent $1 - \nu$.

For larger values of N, the proof runs along the same lines; the space $\tilde{E}$ is defined by adding the N elements 1, $(1 - e^{-i\omega}), \ldots, (1 - e^{-i\omega})^{N-1}$ to $(1 - e^{-i\omega})^N E$, which then lead to eigenvectors for $\mathcal{L}_c$ with eigenvalues $1, 1/2, \ldots, 2^{-N+1}$. The corresponding eigenvectors are used to define a spline starting point $v_0(x)$ which is piecewise polynomial of degree N, and one ends up with an estimate of type $\|v_j(x) - v(x)\|_{\tilde{E}} \leq C \, 2^{-(N-\nu)j}$, leading to the desired result.

References.

Battle, G., A block spin construction of ondelettes, Part I: Lemarié functions. *Comm. Math. Phys.* **110** (1987). 601-615.

Cavaretta, A. S., Dahmen, W., Micchelli, C., Stationary subdivision. *Mem. Amer. Math. Soc.* **93** (1991), 1-186.

Chui, C. K., Wang, J. Z., 1991. A cardinal spline approach to wavelets. *Proc. Amer. Math. Soc.* **113** (1991), 785-793.

Cohen, A., Conze, J. P., Régularité des bases d'ondelettes et mesures ergodiques. *Revista Mat. Iberoamericana.* **8** (1992), 351-366.

Cohen, A., Daubechies, I., Non-separable bidimensional wavelet bases. *Revista Mat. Iberoamericana.* **9** (1993), 51-137.

Cohen, A., Daubechies, I., Feauveau, J. C., Biorthogonal bases of compactly supported wavelets. *Comm. Pure Appl. Math.* **45** (1992), 485-560.

Cohen, A., Ryan, R., *Wavelets and Multiscale Signal Processing.* Chapman and Hall, 1995.

Conze, J. P., Sur le calcul de la norme de Sobolev des fonctions d'échelles. Preprint, Dept. of Math., Université de Rennes, France, 1989.

Daubechies, I., Orthonormal bases of compactly supported wavelets. *Comm. Pure Appl. Math.* **41** (1988), 909-996.

Daubechies, I., *Ten Lectures on Wavelets.* CBMS-NSF Series in Applied Mathematics **61**, SIAM, 1992.

Daubechies, I., Lagarias, J. C., Two-scale difference equations, I. Global regularity of solutions. *SIAM J. Math. Anal.* **22** (1991), 1388-1410.

Daubechies, I., Lagarias, J. C., Two-scale difference equations II. Infinite matrix products, local regularity bounds and fractals. *SIAM J. Math. Anal.* **23** (1992), 1031-1079.

de Boor, C., *A practical guide to splines.* Applied Math. Science Series **27**, Springer, 1978.

Donovan, G., Geronimo, J., Hardin, D., Intertwining multiresolution analyses and the construction of piecewise polynomial wavelets. Preprint, Dept. of Math., Georgia Tech, Atlanta, 1994.

Dyn, N., Subdivision schemes in computer-aided design. In *Advances in Numerical Analysis II, Wavelets, Subdivision Algorithms and Radial Functions*, 36-104. Light, W. A. (ed), Oxford Univ. Press, 1992.

Dyn, N., Levin, D., Interpolating subdivision schemes for the generation of curves and surfaces. In *Multivariate Interpolation and Approximation*, 257-268. Haussmann, W., Jeller, K. (eds), Birkhauser, 1990.

Eirola, T., Sobolev characterization of solutions of dilation equations. *SIAM J. Math. Anal.* **23** (1992), 1015-1030.

Feauveau, J. C., Presentation at the 1989 International Conference on Wavelets and Applications. Marseille, 1989.

Feauveau, J. C., Analyse multirésolution par ondelettes non orthogonals et bancs de filtres numériques. Ph.D. thesis, Université de Paris-Sud, France, 1990.

Gohberg, I., Krein, M., *Introduction to the theory of linear nonselfadjoint operators.* Amer. Math. Soc. (Translation of 1965 Russian original), 1969.

Gripenberg, G., Unconditional bases of wavelets for Sobolev spaces. *SIAM J. Math. Anal.* **24** (1993), 1030-1042.

Grothendieck, A., 1956. La théorie de Fredholm. *Bull. Soc. Math. France,* **84** (1956), 319-384.

Herley, C., Vetterli, M., Wavelets and recursive filter banks. *IEEE Trans. Sign. Proc.*, **41** (1992), 2536-2556.

Hervé, L., Régularité et conditions de bases de Riesz pour les fonctions d'échelle. *C. R. Acad. Sci. Paris, Série I* **315** (1992), 1029-1032.

Hervé, L., Construction et régularité des fonctions d'échelle. *SIAM J. Math. Anal.* **26** (1995), 1361-1385.

Keane, M., Strongly mixing g-measures. *Invent. Math.* **16** (1972), 309-324.

Kovačević, J., Vetterli, M., Perfect reconstruction filter banks for HDTV representation and coding. *Image Comm.* **2** (1990), 349-364.

Kovačević, J., Vetterli, M., Nonseparable multidimensional perfect reconstruction filter banks and wavelet bases for $\mathbb{R}^n$. *IEEE Trans. Inf. Theory* **38** (1992), 533-555. (Special issue on wavelets).

Lemarié, P. G., Une nouvelle base d'ondelettes de $L^2(\mathbb{R}^n)$. *J. Math. Pures et Appl.* **67** (1988), 227-236.

Lidskii, V. B., Non-selfadjoint operators with a trace. *Dokl. Akad. Nauk SSSR*, **125** (1959), 485-487. English translation, *Amer. Math. Soc. Transl.* **47** (1965), 43-46.

Macdonald, I., *Symmetric functions and Hall polynomials.* Oxford Math. Monographs, Oxford Univ. Press, 1979.

Mallat, S., Multiresolution approximation and wavelet orthonormal bases of L^2. *Trans. Amer. Math. Soc.* **315** (1989), 69-88.

Meyer, Y., *Ondelettes ét opérateurs, I: Ondelettes, II: Opérateurs de Calderón-Zygmund, III: Opérateurs multilinéaires.* Hermann, 1990. An English translation is being prepared by the Cambridge Univ. Press.

Micchelli, C. A., Prautzsch, H., Uniform refinement of curves. *Linear Algebra Appl.* **114/115** (1989), 841-870.

Oppenheim, A., Schafer, R., *Digital signal processing.* Prentice Hall, 1975.

Rioul, O., Simple regularity criteria for subdivision schemes. *SIAM J. Math. Anal.* **23** (1992), 1544-1576.

Rudin, W., *Real and complex analysis.* McGraw-Hill, 1974.

Ruelle, D., Zeta-functions for expanding maps and Anosov flows. *Invent. Math.* **34** (1976), 231-242.

Ruelle, D., An extension of the theory of Fredholm determinants. *Inst. Hautes Etudes Sci. Publ.* **72** (1990), 175-193.

Schaefer, H., *Topological vector spaces.* MacMillan, 1966.

Schaefer, H., *Banach lattices and positive operators.* Springer, 1974.

Simon, B., *Trace ideal methods.* London Math. Soc. Lecture Notes **35**, Cambridge Univ. Press, 1979.

Vetterli, M., Multidimensional subband coding: some theory and algorithms. *Signal Processing* **6** (1984), 97-112.

Villemoes, L., Energy moments in time and frequency for 2-scale difference equation solutions and wavelets. *SIAM J. Math. Anal.* **23** (1992), 1519-1543.

Villemoes, L., Continuity of nonseparable quincunx wavelets. *ACHA* **1** (1994), 180-187.

Volkmer, H., On the regularity of wavelets. *IEEE Trans. Inf. Theory* **38** (1992), 872-876. (Special issue on wavelets).

Recibido: 25 de septiembre de 1.995

Albert Cohen
Laboratoire d'Analyse Numerique
Université Pierre et Marie Curie
4 Place Jussie, 75005 Paris, FRANCE

Ingrid Daubechies
AT&T Bell Laboratories
Murray Hill, NJ 07974, USA

Revista Matemática Iberoamericana
Vol. 12, N.º 3, 1996

Bounds for capacities in terms of asymmetry

Tilak Bhattacharya and Allen Weitsman

1. Introduction.

In [6], a study was initiated by R. Hall, W. Hayman, and A. Weitsman relating the asymmetry of a set to various set parameters such as the diameter, isoperimetric constant, and capacity. For a compact set Ω in $\mathbb{R}^n$, let $V(\Omega)$ denote the volume of Ω, and $B(x,\rho)$ the ball of radius ρ centered at x and volume $V(\Omega)$. The asymmetry $\alpha = \alpha(\Omega)$ is then defined by

$$\alpha = \inf_x \frac{V(\Omega \backslash B(x,\rho))}{V(\Omega)}\,, \qquad \rho = \sqrt[n]{\frac{V(\Omega)}{V(B(0,1))}}\,. \tag{1.1}$$

In $\mathbb{R}^2$, we shall use $A(\Omega)$ to denote the area of Ω. It is clear that $\alpha = 0$ when Ω is a ball.

Let $\mathrm{Cap}(\Omega)$ denote the logarithmic capacity of a set Ω in $\mathbb{R}^2$. In [6] it was shown that there exists an absolute constant K_0 such that

$$\mathrm{Cap}(\Omega) \ge (1 + K_0\, \alpha(\Omega)^3)\sqrt{\frac{A(\Omega)}{\pi}}\,. \tag{1.2}$$

This was improved by W. Hansen and N. Nadirashvili in [7] where it was shown that there exists an absolute constant K_1 such that

$$\mathrm{Cap}(\Omega) \ge (1 + K_1\, \alpha(\Omega)^2)\sqrt{\frac{A(\Omega)}{\pi}}\,. \tag{1.3}$$

The inequality (1.3) was conjectured by L. E. Fraenkel and, as noted in [6], the exponent 2 in (1.3) is sharp. The proof in [7] relies on an inequality between capacity and moment of inertia which had been proved by Pólya and Szegö [10, p. 126] for connected sets. For general sets, this inequality had remained open until Hansen and Nadirashvili's ingenious proof in [7]. They also showed that, in (1.3), $K_1 \geq 1/4$. The proofs in [6] are based on estimates for condensers.

In this work we shall prove an analogue of (1.3) for p-capacities of condensers in the plane. The p-capacities have been studied extensively in recent years, especially in connection with degenerate nonlinear elliptic partial differential equations [10]. Since such capacities are very hard to compute exactly (*cf.* [10, p. 35]), we shall develop a perturbative method to obtain approximations in terms of asymmetry.

A condenser $\Gamma = \Gamma(\Omega, \Omega')$ in $\mathbb{R}^2$ consists of a compact set Ω and a disjoint closed unbounded set Ω'. The p-capacity $(1 < p < \infty)$ of the condenser is then

$$\text{Cap}_p(\Gamma) = \inf \iint_{\mathbb{R}^2} |Du|^p \, dxdy \,, \tag{1.4}$$

the infimum being taken over all functions u absolutely continuous in $\mathbb{R}^2$, with $u = 0$ on Ω and $u = 1$ on Ω'. When $p = 2$, the minimizer is the harmonic function in $\mathbb{R}^2 \backslash (\Omega \cup \Omega')$ having the prescribed boundary values. For other values of p, the minimizer satisfies the "p-Laplace equation", namely, $\text{div}(|Du|^{p-2} Du) = 0$. Although solutions to this equation have only locally Hölder continuous first derivatives [12], they do retain a maximum principle, and the critical values are discrete in $\mathbb{R}^2 \backslash (\Omega \cup \Omega')$ [13]. Furthermore, u is analytic near points where $Du \neq 0$ (*cf.* [11, p. 208]). We will consider p-capacities of condensers $\Gamma = \Gamma(\Omega, \Omega')$ where $A(\Omega) = 1$ and $A(\mathbb{R}^2 \backslash \Omega') = 4$. The main result of this work is

Theorem 1. *Let $1 < p < \infty$. There exist constants K_p depending only on p, such that*

$$\text{Cap}_p(\Gamma) \geq (1 + K_p \, \alpha(\Omega)^2) \text{Cap}_p(\Gamma^*) \,, \tag{1.5}$$

where Γ is as above, and $\Gamma^* = \Gamma\left(B(0, 1/\sqrt{\pi}),\ \mathbb{R}^2 \backslash B(0, 2/\sqrt{\pi})\right)$.

The p-capacity of Γ^* is given explicitly by

$$\text{Cap}_p(\Gamma^*) = \left(\int_1^4 \phi(t) \, dt \right)^{1-p} \,, \tag{1.6}$$

where $\phi(t) = \phi_p(t) = (4\pi t)^{p/2(1-p)}$.

In Section 9 we show that the exponent 2 in (1.5) is sharp.

The methods of this paper can be extended to cover condensers whose inner and outer boundaries exhibit asymmetries, but at a cost of much routine and tedious work. Also, (1.5) in case $p = 2$ can be used to give (1.3). In Section 10 we outline this proof. Although it is impossible, due to the intricacies of the proof, to give any meaningful numerical bounds on the constants K_p in (1.5), with additional work one could allow Ω and Ω' to vary in size. The influence on the constants K_p will be discussed in Section 11.

In higher dimensions only partial results have been obtained relating capacities to asymmetry. Under the assumption of convexity on Ω, if $\mathrm{Cap}(\Omega)$ denotes the Newtonian capacity of Ω, then in [6] the inequality corresponding to (1.3) with exponent $n+1$ on α was obtained. This was improved by Hansen and Nadirashvili [7], [8], again for convex sets, also replacing the asymmetry by the quantity

$$d_e(\Omega) = \frac{R_0(\Omega)}{R(\Omega)} - 1\,,$$

where R_0 is the outradius of Ω and $R(\Omega)$ is the radius of the ball having volume $V(\Omega)$. They proved that for small $d = d_e(\Omega)$,

$$\frac{\mathrm{Cap}(\Omega)}{\mathrm{Cap}(B(0,\rho))} \geq \begin{cases} 1 + A\,\dfrac{d^3}{\log 1/d}\,, & n = 3\,,\\ 1 + A_n\, d^{(n+3)/2}\,, & n \geq 4\,, \end{cases}$$

where $V(B(0,\rho)) = V(\Omega)$.

The main challenge which lies ahead is to determine the effect of asymmetry on Newtonian capacity without the assumption of convexity. Although $\alpha < d_e$, and (1.7) is close to best possible for convex sets [8, p. 8], the quantity d_e has no relevance in the study of general Ω. This stems from the fact that line segments have capacity 0 in $\mathbb{R}^n$ for $n \geq 3$, and so d_e can be depressed with negligible effect on the capacity. On the other hand, the notion of asymmetry, which seems to have been introduced in this context by Fraenkel, remains a natural measure of distortion. It seems reasonable to us to conjecture that

$$\frac{\mathrm{Cap}(\Omega)}{\mathrm{Cap}(B(0,\rho))} \geq (1 + D_n\,\alpha^2) \tag{1.8}$$

for constants D_n where again $V(B(0;\rho)) = V(\Omega)$.

In an unpublished work, Fraenkel has verified (1.8) for starlike regions close to a ball in $\mathbb{R}^3$. However, contrary to the remark attributed to the second author in [9], no general bounds on Newtonian capacity in terms of asymmetry appear to be known. It would be interesting to obtain an inequality of the type (1.8) with some exponent on α, but with no assumption of convexity on Ω.

There are two natural avenues of approach to this problem. The first would be to prove an inequality for the moment of inertia $I(\Omega)$ of Ω about its centroid in terms of $\mathrm{Cap}(\Omega)$ as was done in $\mathbb{R}^2$ by Hansen and Nadirashvili. If one could prove the hypothetical inequality

$$\mathrm{Cap}(\Omega)^{n+2} \geq \frac{(n+2)}{\sigma_n} I(\Omega)\,, \tag{1.9}$$

where σ_n is the $(n-1)$-Hausdorff measure of the unit sphere, and where we have normalized so that the capacity of a ball is its radius, then (1.8) would follow easily from

$$I(\Omega) \geq I(B)\left[1 + \frac{n+2}{n^2}\,\alpha^2\right],$$

where B is the ball of volume $V(\Omega)$. Inequality (1.9) is a natural analogue of the inequality of Hansen and Nadirashvili in $\mathbb{R}^n$.

Another possible approach is along the lines of the present paper, especially in view of the recent results of Hall [5] which give the influence of the asymmetry on the usual isoperimetric inequality. With this in mind, the results of this paper, in particular the symmetrization method introduced in Section 3 can be adapted to $\mathbb{R}^n$ for $n \geq 3$ as long as $p = 2$. The difficulty arises in Section 6 where one needs to prove that if the asymmetry is very small, most of Ω is a set whose boundary lies between two very close concentric balls. The present argument relies on the Bonnesen type inequalities (2.2)-(2.4), and it seems difficult to extend this type of argument to higher dimensions.

In the case of p-capacities of condensers in $\mathbb{R}^n$, $n > 2$, nothing seems to be known regarding an analogue of (1.5), even under the additional assumption of convexity. The problem is more difficult especially because there are no known bounds on the sets of critical points, and in particular whether or not such sets are of measure zero. Nevertheless, it seems likely that (1.5) will continue to hold. More precisely, let R_n be such that $V(B(0,R_n)) = 1$, $\Gamma = \Gamma(\Omega,\Omega')$ be a condenser with $V(\Omega) = 1$, and $V(\mathbb{R}^n\backslash\Omega') = 2^n$. Let Γ^* denote the condenser

$\Gamma(\overline{B}(0,R_n), \mathbb{R}^n \backslash B(0,2R_n))$. Then we conjecture that there is a $K_p > 0$, depending only on p, such that

$$\text{Cap}_p(\Gamma) \geq (1 + K_p\, \alpha^2)\, \text{Cap}_p(\Gamma^*)\,. \tag{1.10}$$

We have divided our work as follows. In Section 2, we state and prove some preliminary results required in the proof of Theorem 1. We also discuss our strategy for achieving the proof of Theorem 1. In Section 3, we introduce a new symmetrization technique. Based on this, we prove a perturbation lemma for 2-capacity in Section 4. The proof of Theorem 1 involves considering several independent cases and is spread over sections 5-8. In Section 9, we present an example to prove the sharpness of the exponent 2 in (1.5); Section 10 contains a proof of (1.3) based on the techniques developed in this paper. Finally, in Section 11, we indicate how our result in (1.5) is modified when the ratio of the areas of the sets involved is different from 4.

As in [6], our proofs will rely in part on connections with the isoperimetric inequality. These ideas have been useful in a number of studies (*cf.* [3], [4], [14], [17]).

2. Preliminary results.

We may assume that the sets we are working with are bounded by a finite number of rectifiable curves. Let D be such a set and $L(\partial D)$ denote the length of its boundary. Then it is proved in [6, Lemma 2.1] that

$$L(\partial D)^2 \geq 4\pi\Big(1 + \frac{\alpha(D)^2}{6}\Big) A(D)\,. \tag{2.1}$$

In proving (2.1), use was made of relations between the inradius R_i and outradius R_o of D. Results of this type are collected in [15]. In this paper, we shall have occasion to use the fact [15, p. 3-4] that if D is bounded by a rectifiable Jordan curve, then

$$L(\partial D)^2 - 4\pi\, A(D) \geq \pi^2 (R_o - R_i)^2\,, \tag{2.2}$$

$$R_o \leq \frac{1}{2\pi}\left(L(\partial D) + \sqrt{L(\partial D)^2 - 4\pi\, A(D)}\right)\,, \tag{2.3}$$

and

$$R_i \geq \frac{1}{2\pi}\left(L(\partial D) - \sqrt{L(\partial D)^2 - 4\pi\, A(D)}\right), \tag{2.4}$$

Proposition 2.1. *Suppose that D is a bounded open set and $D = \cup_{i=1}^{\infty} D_i$, where the D_i's are pairwise disjoint components of D, labelled such that $A(D_1) \geq A(D_2) \geq \cdots$. If $0 < \delta < 1/4$, and*

$$A(D_1) \leq (1-\delta)\, A(D)\,,$$

then

$$L(\partial D)^2 \geq 4\pi\,(1+\sqrt{\delta})\, A(D)\,.$$

PROOF. We assume that the perimeter of each D_i is finite. Set $x_i = A(D_i)$, $i = 1, 2, \ldots$, so that $\sum_{i=1}^{\infty} x_i = A(D)$, and $x_1 \geq x_2 \geq x_3 \geq \cdots$. Also

$$x_1 \leq (1-\delta)\, A(D)\,. \tag{2.5}$$

We first consider the case when $x_1 \geq \delta\, A(D)$. Employing the isoperimetric inequality, we have

$$\begin{aligned}
L(\partial D)^2 &= \left(L(\partial D_1) + \sum_{i=2}^{\infty} L(\partial D_i)\right)^2 \\
&\geq L(\partial D_1)^2 + \sum_{i=2}^{\infty} L(\partial D_i)^2 + 2\, L(\partial D_1) \sum_{i=2}^{\infty} L(\partial D_i) \\
&\geq 4\pi \left(\sum_{i=1}^{\infty} x_i + 2\sqrt{x_1} \sum_{i=2}^{\infty} \sqrt{x_i}\right) \\
&\geq 4\pi \left(A(D) + 2\sqrt{x_1}\Big(\sum_{i=2}^{\infty} x_i\Big)^{1/2}\right) \\
&= 4\pi \left(A(D) + 2\sqrt{x_1(A(D) - x_1)}\right).
\end{aligned}$$

Recalling that $\delta\, A(D) \leq x_1 \leq (1-\delta)\, A(D)$, and using the fact that $x(1-x)$ for $x \in [\delta, 1-\delta]$ has as its minimum $\delta(1-\delta)$, we have

$$L(\partial D)^2 \geq 4\pi\,(1+\sqrt{\delta})\, A(D)\,.$$

Thus the statement of the proposition holds in this case.

We now consider the case when x_1 is small, *i.e.*, $x_1 < \delta A(D)$. Then

$$\delta\, A(D) > x_1 \geq x_2 \geq x_3 \geq \cdots ,$$

and

$$\sum_{i\neq \ell} x_i \geq (1-\delta)\, A(D), \qquad \text{for all } \ell = 1, 2, \ldots \tag{2.6}$$

Clearly,

$$\begin{aligned} L(\partial D)^2 &= \Big(\sum_{i=1}^{\infty} L(\partial D_i)\Big)^2 \\ &= \Big(\sum_{i=1}^{\infty} L(\partial D_i)^2 + \sum_{j=1}^{\infty} L(\partial D_j) \sum_{i\neq j} L(\partial D_i)\Big) \\ &\geq 4\pi \Big(A(D) + \sum_{j=1}^{\infty} \sqrt{x_j} \sum_{i\neq j} \sqrt{x_i}\Big) . \end{aligned} \tag{2.7}$$

Setting $\varepsilon_i = x_i/x_1 \leq 1$, and employing (2.6), we obtain

$$\begin{aligned} \sum_{j=1}^{\infty} \sqrt{x_j} \sum_{i\neq j} \sqrt{x_i} &= x_1 \Big(\sum_{j=1}^{\infty} \sqrt{\varepsilon_j} \sum_{i\neq j} \sqrt{\varepsilon_i}\Big) \\ &\geq x_1 \Big(\sum_{j=1}^{\infty} \varepsilon_j \sum_{i\neq j} \varepsilon_i\Big) \\ &\geq \frac{(1-\delta)\, A(D)^2}{x_1} \\ &\geq \frac{(1-\delta)}{\delta}\, A(D) . \end{aligned} \tag{2.8}$$

The proposition now follows easily in this second case by combining (2.7) and (2.8).

By taking the contrapositive of Proposition 2.1, we have

Proposition 2.2. *Let D be a bounded open set such that, for some δ $(0 < \delta < 1/4)$, $L(\partial D)$ satisfies*

$$L(\partial D)^2 < 4\pi\, (1 + \sqrt{\delta})\, A(D) .$$

If D_1 is a component of D with the largest area, then

$$A(D_1) > (1-\delta)\, A(D)\,.$$

REMARK 2.1. The exponent 1/2 appearing on δ in the statement of Proposition 2.1 is sharp. To see this take $D = D_1 \cup D_2$, where D_1 and D_2 are two disjoint discs of radius $\sqrt{1-\delta}$ and $\sqrt{\delta}$ respectively. Take $\delta < 1/4$. Then $A(D) = \pi$, and $A(D_1) = (1-\delta)A(D)$. Clearly, $L(\partial D)^2 = 4\pi(1 + O(\sqrt{\delta}))A(D)$, as $\delta \to 0$.

For a condenser Γ with inner set Ω and outer set $\mathbb{R}^2 \setminus \Omega'$, if u is the extremal extended to be zero on Ω, we write $F(t) = \{x : u(x) < t\}$ and $A(t) = A(F(t))$ $(0 < t \le 1)$. We will often write $\alpha = \alpha(\Omega)$ for convenience.

Our proof of Theorem 1 will be broken down into two cases. In Case 1, the asymmetry of Ω is propagated through a t interval for the sets $F(t)$. Here the proof follows the methods of [6]. It is easy to construct examples of sets Ω for which $\alpha(F(t))$ is dramatically less than $\alpha(\Omega)$ for t arbitrarily close to zero. Case 2 is designed to cover this possibility.

The plan in Case 2 is as follows. Since $\alpha(F(T))$ is very small for some T close to 0, we first observe that this implies that most of $F(T)$ is a set, which we later call F_1, whose boundary is contained between very close concentric circles. This is the essence of (6.18) below. By using the symmetrization of Section 3, we construct a new condenser with comparable asymmetry and decreased p-capacity by suitably redistributing the portion of F_1 on each ray from the center x_0 of the concentric circles. Using the new configuration, we then obtain a lower bound on the capacities stated in Lemma 4.1.

In what follows, κ and η will denote small positive constants which do not depend on α, and which will be determined later. We assume

$$0 < \kappa < 0.0001,\ \eta \le 0.01, \quad \text{and} \quad \kappa < \frac{\eta^2}{10}\,. \tag{2.9}$$

Case 1. For all t such that

$$1 + \eta \le A(t) \le 1 + 2\eta \tag{2.10}$$

we have

$$L(\partial F(t))^2 \ge 4\pi\,(1 + \kappa\alpha^2)\, A(t)\,. \tag{2.11}$$

Case 2. There exists a value T such that

$$1+\eta \le A(T) \le 1+2\eta \tag{2.12}$$

and

$$L(\partial F(T))^2 < 4\pi\,(1+\kappa\,\alpha^2)\,A(T)\,. \tag{2.13}$$

By the result in [13], in Case 1, Du can vanish on at most a finite number of levels $u=t$ in the interval specified by (2.10). In Case 2, by making a slight adjustment, we may choose T such that Du is nonvanishing on the boundary of $F(T)$. Thus we may take $\partial F(T)$ to be analytic in the latter case.

3. A symmetrization technique.

We now present a new type of symmetrization which will be useful in relating p-capacity to asymmetry. Let Ω_1 and F_1 be two bounded open subsets of $\mathbb{R}^2$. We assume that i) $\overline{\Omega}_1 \subset F_1$, ii) the origin 0 lies in Ω_1, and iii) $\partial\Omega_1$ and ∂F_1 are the unions of finitely many Lipschitz continuous curves. Let $\rho=\sqrt{A(\Omega_1)/\pi}$ and $R=\sqrt{A(F_1)/\pi}$.

For each $\theta\in(-\pi,\pi]$, let $J(\theta)=\{re^{i\theta}:\ 0\le r\}$ be the ray from the origin making an angle θ with the positive x-axis. For a given value of θ, let

$$J(\theta)\cap\Omega_1=[r_0,r_1(\theta))\bigcup_{j\ge1}(r_{2j}(\theta),r_{2j+1}(\theta))\,,\qquad r_0=0\,,$$

the intervals being disjoint. We now introduce the parameters necessary to give a redistribution of the area of Ω_1 relative to $B(0,\rho)$. Set

$$\begin{aligned}
s(\theta)&=\sup\{r:\ re^{i\theta}\in J(\theta)\cap\Omega_1\},\\
t(\theta)&=\inf\{r:\ re^{i\theta}\in J(\theta)\cap\partial F_1\}\\
&=\sup\{r:\ [0,r)\subset J(\theta)\cap F_1\},\\
\hat s(\theta)&=\sup\{r:\ re^{i\theta}\in J(\theta)\cap\Omega_1, r<t(\theta)\},\\
\hat t(\theta)&=\inf\{r:\ re^{i\theta}\in J(\theta)\cap\partial F_1, r>s(\theta)\}\\
&=\sup\{r:\ [s(\theta),r)\subset J(\theta)\cap F_1\},\\
N&=\{re^{i\theta}\in\Omega_1:\ s(\theta)>t(\theta),\ r>\hat s(\theta)\},\\
E&=\{\theta:\ J(\theta)\cap N\neq\varnothing\}\,.
\end{aligned}\tag{3.1}$$

Figure 1. $\hat{s}(\theta) = s(\theta) < t(\theta) = \hat{t}(\theta)$.

Figure 2. $\hat{s}(\theta) < t(\theta) < s(\theta) < \hat{t}(\theta)$ (Shaded Region in N).

Note that $\hat{s}(\theta) \le s(\theta)$ and $\hat{t}(\theta) \ge t(\theta)$ with equality if and only if $s(\theta) < t(\theta)$.

We distinguish two possibilities in our redistribution of Ω_1.

Case A. Suppose first that $\hat{s}(\theta) \le \rho$. Then we define $\xi(\theta) > 0$ by

$$\xi(\theta)^2 = \sum_{j \in K} r_{2j+1}^2 - r_{2j}^2 , \tag{3.2}$$

where $K = \{j:\ r_{2j+1} \le \hat{s}(\theta)\}$.

Case B. If $\hat{s}(\theta) > \rho$ we distinguish two subcases to define $\xi(\theta) > 0$ and $\lambda(\theta) > 0$.

i) If $\rho \in J(\theta) \cap \Omega_1$, *i.e.*, $r_{2m} < \rho < r_{2m+1}$ for some m, then

$$\xi(\theta)^2 = r_{2m+1}^2 + \sum_{j \in L} r_{2j+1}^2 - r_{2j}^2\ , \tag{3.3}$$

where $L = \{j:\ r_{2m+1} \le r_{2j} < r_{2j+1} \le \hat{s}(\theta)\}$; also let

$$\lambda(\theta)^2 = \rho^2 - r_{2m}^2 + \sum_{j \in M} r_{2j+1}^2 - r_{2j}^2\ , \tag{3.4}$$

where $M = \{j:\ r_{2j+1} \le r_{2m}\}$.

ii) If $\rho \notin J(\theta) \cap \Omega_1$, we set

$$\xi(\theta)^2 = \rho^2 + \sum_{j \in L'} r_{2j+1}^2 - r_{2j}^2\ , \tag{3.5}$$

where $L' = \{j:\ \rho \le r_{2j} < r_{2j+1} \le \hat{s}(\theta)\}$, and

$$\lambda(\theta)^2 = \sum_{j \in M'} r_{2j+1}^2 - r_{2j}^2\ , \tag{3.6}$$

where $M' = \{j:\ r_{2j+1} \le \rho\}$. It is useful to observe that $\xi(\theta) > \rho$ and $\lambda(\theta) \le \rho$, whenever $\hat{s}(\theta) > \rho$.

For each $\theta \in (-\pi, \pi]$, let $\Omega_1^*(\theta) \subset J(\theta)$ be defined by

$$\Omega_1^*(\theta) = \begin{cases} [0, \xi(\theta)]\,, & \text{if } \hat{s}(\theta) \le \rho\,, \\ [0, \lambda(\theta)] \cup (\rho, \xi(\theta)]\,, & \text{if } \hat{s}(\theta) > \rho \text{ and } \lambda(\theta) < \rho\,, \\ [0, \xi(\theta)]\,, & \text{if } \hat{s}(\theta) > \rho \text{ and } \lambda(\theta) = \rho\,. \end{cases}$$

Define $\Omega_1^* = \cup_\theta \Omega_1^*(\theta)$; by the definitions in (3.1)-(3.6), it is clear that i) $A(\Omega_1) = A(\Omega_1^*) + A(N)$ (see (3.10)), ii) if $B(0,r) \subset \Omega_1$, then $B(0,r) \subset \Omega_1^*$, and iii) $\Omega_1^* \cap B(0,\rho)$ is starlike with respect to the origin 0.

Now suppose that $0 < R_i' \le \rho$ and $R_i \le R_o$ are such that $\bar{B}(0, R_i') \subset \Omega_1$, and $\bar{B}(0, R_i) \subset F_1 \subset \bar{F_1} \subset B(0, R_o)$. Then we conclude from

(3.1)-(3.6) that

$$
\begin{aligned}
&\text{i) } R_i' \le \xi(\theta) \le \hat{s}(\theta) \le s(\theta) \le \hat{t}(\theta) \le R_o\,,\\
&\text{ii) } R_i \le t(\theta) \le \hat{t}(\theta) \le R_o\,,\\
&\text{iii) } R_i \le t(\theta) < s(\theta) < \hat{t}(\theta) \le R_o,\ \theta \in E\,,\\
&\text{iv) } R_i' \le \xi(\theta) < t(\theta) \le R_o\,,\\
&\text{v) } R_i' \le \min\{\rho, R_i\} \le \max\{\rho, R_i\} \le R \le R_o\,,\\
&\text{vi) If } \hat{s}(\theta) > \rho,\ \text{then } \xi(\theta) > \rho,\ \text{and } \lambda(\theta) \le \rho\,.
\end{aligned}
\tag{3.7}
$$

Based on (3.7) we now make some easy observations. These will be useful in Section 4 and Section 8. Suppose that $\beta = A(\Omega_1 \backslash B(0,\rho))/A(\Omega_1) > 0$. By consideration of $\Omega_1 \backslash B(0,\rho)$ we have

$$
\begin{aligned}
0 &< 2\pi\rho^2 \left(\beta - \frac{A(N\backslash B(0,\rho))}{\pi\rho^2}\right)\\
&= \int_{\{\xi(\theta)\ge\rho\}} \left(\xi(\theta)^2 - \rho^2\right) d\theta \le 2\pi\rho^2\beta\,.
\end{aligned}
\tag{3.8}
$$

By consideration of $\Omega_1 \cap B(0,\rho)$,

$$
\begin{aligned}
0 &< \int_{\{\xi(\theta)\le\rho\}} (\rho^2 - \xi(\theta)^2)\, d\theta + \int_{\{\hat{s}(\theta)>\rho\}} (\rho^2 - \lambda(\theta)^2)\, d\theta\\
&= 2\pi\rho^2 \left(\beta + \frac{A(N\cap B(0,\rho))}{\pi\rho^2}\right).
\end{aligned}
\tag{3.9}
$$

Subtracting (3.8) from (3.9), we then have

$$
0 < \int_{-\pi}^{\pi} (\rho^2 - \xi(\theta)^2)\, d\theta + \int_{\{\hat{s}(\theta)>\rho\}} (\rho^2 - \lambda(\theta)^2)\, d\theta = 2\,A(N)\,,
\tag{3.10}
$$

and adding we obtain

$$
\begin{aligned}
\int_{-\pi}^{\pi} |\rho^2 - \xi(\theta)^2|\, d\theta &= \int_{\{\xi(\theta)\ge\rho\}} (\xi(\theta)^2 - \rho^2)\, d\theta\\
&\quad + \int_{\{\xi(\theta)\le\rho\}} (\rho^2 - \xi(\theta)^2)\, d\theta\\
&\le 4\pi\rho^2 \left(\beta + \frac{A(N)}{\pi\rho^2}\right).
\end{aligned}
\tag{3.11}
$$

Also, let

$$(3.12) \qquad \begin{cases} \mu = \dfrac{1}{\rho^2} \displaystyle\int_{\{\hat{s}(\theta) > \rho\}} (\rho^2 - \lambda(\theta)^2)\, d\theta \geq 0\,, \\ \bar{\mu} = \dfrac{1}{R^2} \displaystyle\int_{-\pi}^{\pi} (R^2 - t(\theta)^2)\, d\theta \geq 0\,. \end{cases}$$

In the next section we will use this symmetrization technique to deduce a perturbation result for 2-capacity.

4. A perturbation lemma for 2-capacity.

We will now prove a perturbation lemma based on the symmetrization introduced in Section 3. As before, Ω_1 and F_1, subsets of $\mathbb{R}^2$, are bounded open sets such that i) $\overline{\Omega}_1 \subset F_1$, ii) the origin 0 lies in Ω_1, and iii) $\partial\Omega_1$ and ∂F_1 are the unions of finitely many Lipschitz continuous curves. Set $\rho = \sqrt{A(\Omega_1)/\pi}$ and $R = \sqrt{A(F_1)/\pi}$. Let $0 < R_i' \leq \rho$ and $R_i \leq R_o$ be such that $\bar{B}(0, R_i') \subset \Omega_1$, $\bar{B}(0, R_i) \subset F_1 \subset \bar{F}_1 \subset B(0, R_o)$. Suppose furthermore that

$$(4.1) \qquad \begin{aligned} &\text{i) For a fixed } \varepsilon,\ 0 < \varepsilon \leq 1/2,\ R_o(1-\varepsilon) \leq R_i \leq R \leq R_o\,, \\ &\text{ii) } 1/2 \leq R_i'/R_o \leq R_i/R_o \leq 1\,, \\ &\text{iii) For } 0 < \delta \leq 1/2,\ 1/4 \leq (\rho/R)^2 \leq 1/(1+\delta) < 1\,. \end{aligned}$$

By the definition in (1.4), if $\Gamma = \Gamma(\overline{\Omega}_1, \mathbb{R}^2 \backslash F_1)$, then

$$I = \operatorname{Cap}_2(\Gamma) = \inf_w \int_{F_1 \backslash \Omega_1} |Du|^2\, dx\, dy\,,$$

where, w is absolutely continuous and takes the value 1 on $\mathbb{R}^2 \backslash F_1$ and 0 on $\overline{\Omega}_1$. Let v denote the minimizer. Then it is harmonic in $F_1 \backslash \overline{\Omega}_1$ and assumes the appropriate boundary values. Set

$$\beta = \frac{A(\Omega_1 \backslash B(0,\rho))}{A(\Omega_1)} > 0\,,$$

We prove

Lemma 4.1. *Let $\Omega_1, F_1, \rho, R, R_i, R_i', R_o, \beta, \varepsilon, \delta$, and v be as described above. Assume that* (4.1) *holds. Then for all sufficiently small ε, we have*

$$I = \int_{F_1\setminus\Omega_1} |Dv|^2 dx\, dy \geq \frac{2\pi}{\log\ R/\rho} + B_0\,\beta^2 - B_1\,\varepsilon^2 - B_2\,\varepsilon\,\beta\,,$$

where B_0, B_1 and B_2 are positive constants depending only on δ.

PROOF. Throughout the proof we shall let C, with or without subscripts, denote positive constants depending only on δ, and which need not be the same at each occurrence. We employ the symmetrization introduced in Section 3, and use the same notations as in (3.1)-(3.6). Then from (3.7) and (4.1), we may conclude that

$$\begin{aligned} &\text{i) } 0 < \hat{t}(\theta) - s(\theta) \leq \varepsilon R_o\,,\ \ \theta \in E\,,\\ &\text{ii) } (1/e)^2 < (1/2)^2 \leq \min\{\xi(\theta)^2/R^2, \xi(\theta)^2/t(\theta)^2\}\,,\\ &\text{iii) } |R^2 - t(\theta)^2| \leq 2\,\varepsilon R_o^2\,.\\ &\text{iv) } 1-\varepsilon \leq t(\theta)/R_o \leq 1\,. \end{aligned} \tag{4.2}$$

Now

$$\begin{aligned} I &= \int_{F_1\setminus\Omega_1} \Big(v_r^2 + \frac{1}{r^2}v_\theta^2\Big)\, r\, dr\, d\theta\\ &\geq \int_{F_1\setminus\Omega_1} v_r^2\ r\, dr\, d\theta\\ &\geq \int_{-\pi}^{\pi} \Big(\inf \int_{J(\theta)\cap\{F_1\setminus\Omega_1\}} z_r^2 r\, dr\Big)\, d\theta\,, \end{aligned} \tag{4.3}$$

where the infimum is taken over all $z = z(r,\theta)$ such that $z = 1$ on $J(\theta)\cap\partial F_1$ and $z = 0$ on $J(\theta)\cap\partial\Omega_1$. The minimizer $\bar{z}$ satisfies the one variable Euler equation $(r\bar{z}')' = 0$ in $J(\theta)\cap\{F_1\setminus\bar{\Omega}_1\}$. We will now estimate I by employing the symmetrization in Section 3 and obtaining a lower bound for the inner integral on the right side of (4.3). We do this by first solving for $\bar{z}$ from the aforementioned o.d.e over the disjoint intervals $(\hat{s}(\theta), t(\theta))$ and $(s(\theta), \hat{t}(\theta))$, the latter occurring whenever $s(\theta) > t(\theta)$. Note that $\bar{z}$ vanishes on the left end points of these intervals and takes the value 1 on the right end points. Also see (3.7). Thus a lower bound for I is obtained by calculating the inner integral for this function $\bar{z}$

over the above mentioned intervals. Recalling the definition of E from (3.1), it follows from (4.3), (3.7.i)), and (3.1) that

$$
\begin{aligned}
I &\geq \int_{-\pi}^{\pi} \frac{1}{\log(t(\theta)/\hat{s}(\theta))}\, d\theta + \int_E \frac{1}{\log(\hat{t}(\theta)/s(\theta))}\, d\theta \\
&\geq \int_{-\pi}^{\pi} \frac{1}{\log(t(\theta)/\xi(\theta))}\, d\theta + \int_E \frac{1}{\log(\hat{t}(\theta)/s(\theta))}\, d\theta\,.
\end{aligned} \tag{4.4}
$$

If the second integral, on the right hand side of (4.4), is larger than $4\pi/\log(R/\rho)$ then Lemma 4.1 follows trivially from (4.1.iii)). Otherwise,

$$\int_E \frac{1}{\log(\hat{t}(\theta)/s(\theta))}\, d\theta \leq \frac{4\pi}{\log(R/\rho)}\,.$$

But, $\log(\hat{t}(\theta)/s(\theta)) \leq (\hat{t}(\theta)/s(\theta) - 1)$, so it then follows from (4.2.i)), (4.1.ii)-iii)) and (3.7.iii)) that

$$\operatorname{meas}_\theta E \leq C_1\, \varepsilon\,.$$

Note that C_1 depends only on δ. Since

$$N = \{re^{i\theta} \in \Omega_1 :\ s(\theta) > t(\theta),\ r > \hat{s}(\theta)\}\,,$$

(4.1.i)) then yields

$$A(N) \leq C_2\, \varepsilon^2 R_o^2\,. \tag{4.5}$$

Now, from (4.4),

$$
\begin{aligned}
I &\geq \int_{-\pi}^{\pi} \frac{1}{\log(t(\theta)/\xi(\theta))}\, d\theta \\
&= 2\int_{-\pi}^{\pi} \frac{-1}{\log(\xi(\theta)^2/t(\theta)^2)}\, d\theta\,.
\end{aligned} \tag{4.6}
$$

To estimate (4.6) we observe that the function $f(x) = -1/\log x$ satisfies

$$
\begin{cases}
\text{i)} & f(x) > 0 \quad (0 < x < 1)\,, \\
\text{ii)} & f'(x) > 0 \quad (0 < x < 1)\,, \\
\text{iii)} & f''(x) > 0 \quad (1/e^2 < x < 1)\,.
\end{cases} \tag{4.7}
$$

We shall use (4.7) in the form

$$f(x)-f(\bar{x})=f(\bar{x})\,(x-\bar{x})+\frac{f''(\zeta)}{2}\,(x-\bar{x})^2 \tag{4.8}$$

for some $\zeta\in(x,\bar{x})$ or $(\bar{x},x)$. Then with $\bar{x}=\rho^2/R^2$, it follows from (4.1), (4.2), (4.6), (4.7) and (4.8) that

$$\begin{aligned} I-\frac{2\pi}{\log(R/\rho)} &\geq 2\int_{-\pi}^{\pi}\Big(\frac{-1}{\log(\xi(\theta)^2/t(\theta)^2)}+\frac{1}{\log(\rho^2/R^2)}\Big)\,d\theta \\ &\geq 2\,f'(\rho^2/R^2)\int_{-\pi}^{\pi}\Big(\frac{\xi(\theta)^2}{t(\theta)^2}-\frac{\rho^2}{R^2}\Big)\,d\theta \\ &\quad + C_3\int_{-\pi}^{\pi}\left(\frac{\xi(\theta)^2}{t(\theta)^2}-\frac{\rho^2}{R^2}\right)^2 d\theta\,. \end{aligned} \tag{4.9}$$

The positive absolute constant C_3 in (4.9) results from the fact that (4.2.ii)) implies that $\xi(\theta)^2/t(\theta)^2>1/e^2$.

Next we estimate the quantities

$$S=\int_{-\pi}^{\pi}\Big(\frac{\xi(\theta)^2}{t(\theta)^2}-\frac{\rho^2}{R^2}\Big)\,d\theta\,,\qquad \bar{S}=\int_{-\pi}^{\pi}\left(\frac{\xi(\theta)^2}{t(\theta)^2}-\frac{\rho^2}{R^2}\right)^2 d\theta\,.$$

We may rewrite S as

$$S=\int_{-\pi}^{\pi}\Big((\xi(\theta)^2-\rho^2)\Big(\frac{1}{t(\theta)^2}-\frac{1}{R^2}\Big)+\rho^2\Big(\frac{1}{t(\theta)^2}-\frac{1}{R^2}\Big)+\frac{\xi(\theta)^2-\rho^2}{R^2}\Big)\,d\theta\,.$$

By (3.10) and (3.12)

$$\int_{-\pi}^{\pi}\frac{\xi(\theta)^2-\rho^2}{R^2}\,d\theta=\frac{\mu\rho^2}{R^2}-\frac{2\,A(N)}{R^2}\geq-\frac{2\,A(N)}{R^2}\,. \tag{4.10}$$

Also, by (3.11), (4.1.ii)-iii)), (4.2.iii-iv)),

$$\begin{aligned} &\left|\int_{-\pi}^{\pi}(\xi(\theta)^2-\rho^2)\left(\frac{1}{t(\theta)^2}-\frac{1}{R^2}\right)\,d\theta\right| \\ &\qquad\leq\int_{-\pi}^{\pi}|\xi(\theta)^2-\rho^2|\left|\frac{1}{t(\theta)^2}-\frac{1}{R^2}\right|\,d\theta \\ &\qquad\leq\frac{C_4\,\varepsilon}{R^2}\int_{-\pi}^{\pi}|\xi(\theta)^2-\rho^2|\,d\theta \\ &\qquad\leq C_5\,\varepsilon\left(\beta+\frac{A(N)}{\pi\rho^2}\right)\,. \end{aligned} \tag{4.11}$$

By (3.12)

$$\int_{-\pi}^{\pi}\Big(\frac{1}{t(\theta)^2}-\frac{1}{R^2}\Big)\,d\theta \tag{4.12}$$
$$=\int_{-\pi}^{\pi}\Big(\frac{R^2-t(\theta)^2}{R^2t(\theta)^2}-\frac{R^2-t(\theta)^2}{R^4}+\frac{\bar\mu}{2\pi R^2}\Big)\,d\theta$$
$$=\int_{-\pi}^{\pi}\Big(\frac{(R^2-t(\theta)^2)^2}{R^4t(\theta)^2}+\frac{\bar\mu}{2\pi R^2}\Big)\,d\theta$$
$$\geq 0\,.$$

Putting together (4.10), (4.11), and (4.12) we have

$$S=\int_{-\pi}^{\pi}\Big(\frac{\xi(\theta)^2}{t(\theta)^2}-\frac{\rho^2}{R^2}\Big)\,d\theta \geq -\frac{2\,A(N)}{R^2}-C_5\,\varepsilon\left(\beta+\frac{A(N)}{\pi\rho^2}\right)\,. \tag{4.13}$$

We now estimate $\bar{S}$. Observe that

$$\frac{1}{2}\left(\frac{\xi(\theta)^2}{R^2}-\frac{\rho^2}{R^2}\right)^2\leq\left(\frac{\xi(\theta)^2}{R^2}-\frac{\xi(\theta)^2}{t(\theta)^2}\right)^2+\left(\frac{\xi(\theta)^2}{t(\theta)^2}-\frac{\rho^2}{R^2}\right)^2\,.$$

Integrating with respect to θ and recalling (3.7.i)), (4.1.i)-ii)) and (4.2.ii)), we have

$$\int_{-\pi}^{\pi}\left(\frac{\xi(\theta)^2}{R^2}-\frac{\xi(\theta)^2}{t(\theta)^2}\right)^2 d\theta\leq C_6\,\varepsilon^2\,. \tag{4.14}$$

Using Hölder's inequality,

$$\left(\int_{\xi(\theta)\geq\rho}\left(\xi(\theta)^2-\rho^2\right)d\theta\right)^2\leq\left(\int_{-\pi}^{\pi}|\xi(\theta)^2-\rho^2|\,d\theta\right)^2$$
$$\leq 2\pi\int_{-\pi}^{\pi}(\xi(\theta)^2-\rho^2)^2d\theta\,,$$

so by (3.8) and (4.1.iii)),

$$\frac{1}{2}\int_{-\pi}^{\pi}\left(\frac{\xi(\theta)^2}{R^2}-\frac{\rho^2}{R^2}\right)^2 d\theta\geq C_7\left(\beta-\frac{A(N)}{\pi\rho^2}\right)^2\,. \tag{4.15}$$

Putting together (4.14) and (4.15) we obtain

$$\bar{S}=\int_{-\pi}^{\pi}\left(\frac{\xi(\theta)^2}{t(\theta)^2}-\frac{\rho^2}{R^2}\right)^2 d\theta\geq C_8\,\beta^2-C_9\,\varepsilon^2-C_{10}\,\frac{A(N)}{\pi\rho^2}\,. \tag{4.16}$$

By virtue of (4.1) and (4.2), the positive constants C_1-C_{10} depend only on δ. The estimates in (4.13), (4.16) and (4.5) in (4.9) then give

$$I \geq \frac{2\pi}{\log(R/\rho)} + B_0\,\beta^2 - B_1\,\varepsilon^2 - B_2\,\varepsilon\,\beta\,.$$

where B_0, B_1, and B_2 are positive constants depending only on δ. This concludes the proof of Lemma 4.1.

A p-analogue of Lemma 4.1 appears in Section 8.

Remark 4.1. The constants B_0, B_1 and B_2 appearing in the statement of the Lemma 4.1, become absolute once a numerical value for δ is chosen. In our application of Lemma 4.1, a positive value for δ will be fixed once a positive value for η, appearing in (2.9)-(2.13), is chosen. In particular, we may take $\delta = 0.9\,\eta$. See (6.29.x)).

In the next four sections, we will present the proof of Theorem 1, based on the strategy outlined in Section 2. The proof in Case 1 appears in Section 5, while the proof in Case 2 will be presented in sections 6, 7 and 8.

5. Proof of (1.5) in Case 1.

We will first prove Theorem 1 in the situation when asymmetry propagates, that is, when (2.10) implies (2.11). It is easy to see that $A(t)$ is continuous and increasing. If we set

$$s_0 = \inf\{t \in [0,1] :\ A(t) \geq 1+\eta\} \tag{5.1}$$

and

$$T_0 = \sup\{t \in [0,1] :\ A(t) \leq 1+2\eta\}\,, \tag{5.2}$$

then

$$A(s_0) \leq A(t) \leq A(T_0)\,, \qquad t \in [s_0, T_0]\,. \tag{5.3}$$

Recall from Section 1 that u is locally $C^{1,\gamma}$. Hence an application of the coarea formula [2, p. 248] yields, for almost everywhere t,

$$A'(t) = \int_{\partial F(t)} \frac{1}{|Du|}\,d\sigma\,. \tag{5.4}$$

The formula in (5.4) holds except for possibly a discrete set of t's since the set of critical points of u is discrete. We now prove

Lemma 5.1. *Let $1 < p < \infty$. If u is the extremal for the condenser with inner set Ω and outer set $\mathbb{R}^2 \backslash \Omega'$ and T_0 is as in (5.2), then*

$$T_0 \leq \left(\int_{F(T_0)} |Du|^p dx\, dy \right)^{1/p} \cdot \left(\frac{1}{1 + C\alpha^2} \int_1^{1+2\eta} \phi(t)\, dt \right)^{(p-1)/p}, \tag{5.5}$$

where $\phi(t) = \phi_p(t) = (4\pi t)^{p/2(1-p)}$, $\alpha = \alpha(\Omega)$, and C is a constant which depends only on κ, η and p.

PROOF. By the coarea formula and (5.4) we have outside a discrete set of t's,

$$\begin{aligned} \int_{\partial F(t)} 1\, d\sigma &\leq \left(\int_{\partial F(t)} |Du|^{p-1} d\sigma \right)^{1/p} \left(\int_{\partial F(t)} \frac{1}{|Du|}\, d\sigma \right)^{(p-1)/p} \\ &= \left(\int_{\partial F(t)} |Du|^{p-1} d\sigma \right)^{1/p} (A'(t))^{(p-1)/p}. \end{aligned}$$

Using (2.10) and (2.11) it follows, for almost everywhere t with $s_0 < t \leq T_0$ (see (5.1)-(5.3)),

$$1 \leq \left(\int_{\partial F(t)} |Du|^{p-1} d\sigma \right)^{1/p} \left(\frac{(A'(t))^{(p-1)/p}}{\sqrt{4\pi\,(1 + \kappa\,\alpha^2)\, A(t)}} \right). \tag{5.6}$$

We now integrate (5.6) from s_0 to T_0. An application of Hölder's inequality then yields

$$\begin{aligned} T_0 - s_0 &\leq \int_{s_0}^{T_0} \left(\int_{\partial F(t)} |Du|^{p-1} d\sigma \right)^{1/p} \left(\frac{(A'(t))^{(p-1)/p}}{\sqrt{4\pi\,(1 + \kappa\,\alpha^2)\, A(t)}} \right) dt \\ &\leq \left(\int_{s_0}^{T_0} \left(\int_{\partial F(t)} |Du|^{p-1} d\sigma \right) dt \right)^{1/p} \\ &\quad \cdot \left(\int_{s_0}^{T_0} \frac{A'(t)}{(4\pi\,(1 + \kappa\,\alpha^2)\, A(t))^{p/2(p-1)}}\, dt \right)^{(p-1)/p}. \end{aligned} \tag{5.7}$$

Thus, by the coarea formula and the formula for ϕ as described in (1.6), we have

$$\tag{5.8} \begin{aligned} T_0 - s_0 \leq & \left(\int_{F(T_0)\setminus F(s_0)} |Du|^p dx\, dy \right)^{1/p} \\ & \cdot \left(\frac{1}{\sqrt{1+\kappa\,\alpha^2}} \left(\int_{1+\eta}^{1+2\eta} \phi(t)\, dt \right)^{(p-1)/p} \right). \end{aligned}$$

Using the same procedure on $(0, s_0)$ and the usual isoperimetric inequality in place of (2.11), we see that

$$\tag{5.9} s_0 \leq \left(\int_{F(s_0)} |Du|^p dx\, dy \right)^{1/p} \left(\int_1^{1+\eta} \phi(t)\, dt \right)^{(p-1)/p}.$$

Adding (5.8) and (5.9) and applying the Hölder inequality, we may show that

$$\begin{aligned} T_0 \leq & \left(\int_{F(T_0)} |Du|^p dx\, dy \right)^{1/p} \\ & \cdot \left(\int_1^{1+\eta} \phi(t)\, dt + \left(\frac{1}{1+\kappa\,\alpha^2} \right)^{p/2(p-1)} \int_{1+\eta}^{1+2\eta} \phi(t)\, dt \right)^{(p-1)/p} \\ = & \left(\int_{F(T_0)} |Du|^p dx\, dy \right)^{1/p} \\ & \cdot \left(1 - \left(1 - \left(\frac{1}{1+\kappa\,\alpha^2} \right)^{p/2(p-1)} \right) \frac{\displaystyle\int_{1+\eta}^{1+2\eta} \phi(t)\, dt}{\displaystyle\int_1^{1+2\eta} \phi(t)\, dt} \right)^{(p-1)/p} \\ & \cdot \left(\int_1^{1+2\eta} \phi(t)\, dt \right)^{(p-1)/p}. \end{aligned}$$

The inequality in the lemma now follows with an appropriate constant $C = C(\kappa, \eta, p)$.

PROOF OF (1.5) IN CASE 1. Using the usual isoperimetric inequality and the above procedure, we may show that

$$
\begin{aligned}
1 - T_0 \le & \left(\int_{F(1)\setminus F(T_0)} |Du|^p dx\, dy \right)^{1/p} \\
& \cdot \left(\int_{1+2\eta}^{4} \phi(t)\, dt \right)^{(p-1)/p} .
\end{aligned} \tag{5.10}
$$

We now add (5.5) and (5.10), and then use the Hölder inequality to deduce that

$$
\begin{aligned}
1 \le & \left(\int_{F(1)} |Du|^p dx\, dy \right)^{1/p} \\
& \cdot \left(\frac{1}{1 + C\,\alpha^2} \int_1^{1+2\eta} \phi(t)\, dt + \int_{1+2\eta}^{4} \phi(t)\, dt \right)^{(p-1)/p} \\
= & \left(\int_{F(1)} |Du|^p dx\, dy \right)^{1/p} \left(1 - \frac{C\alpha^2}{1 + C\alpha^2} \frac{\int_1^{1+2\eta} \phi(t)\, dt}{\int_1^4 \phi(t)\, dt} \right)^{(p-1)/p} \\
& \cdot \left(\int_1^4 \phi(t)\, dt \right)^{(p-1)/p} .
\end{aligned}
$$

Noting (1.6) we easily obtain the statement of Theorem 1.

6. Geometry of the Sets in Case 2.

Assume Case 2 holds. In this section we shall use (2.12) and (2.13) to construct a subcondenser whose inner set is close to a disc. Lemma 4.1 will then provide the necessary estimates for obtaining the 2-capacity of the original condenser.

We may assume, as in [8, p. 5], that the components of Ω are simply connected, so that by the maximum principle, the components of the set $F(t)$ for each t in $(0, 1]$, are simply connected. Let $F_1(t)$ be one having largest area, and $F_2(t) = F(t)\setminus F_1(t)$. We first show that it suffices to assume that for some t such that

$$
A(t) < 1 + \kappa\,\alpha^2 , \tag{6.1}
$$

we have

$$A(F_1(t)) > (1-\eta/10)\,A(t) \tag{6.2}$$

and

$$L(\partial F_1(t))^2 < 4\pi\,(1+\eta)\,A(F_1(t))\,. \tag{6.3}$$

Let $\tau = \sup\{t : A(t) < 1 + k\alpha^2\}$. Suppose that (6.2) were false for all t such that $0 < t \le \tau$. It follows from Proposition 2.1 and (2.9) that

$$L(\partial F(t))^2 \ge 4\pi\,(1+\sqrt{\eta/10})\,A(t)\,, \qquad 0 < t \le \tau\,. \tag{6.4}$$

If, on the other hand, (6.2) holds but (6.3) does not, then instead of (6.4) we get

$$\begin{aligned} L(\partial F(t))^2 &\ge L(\partial F_1(t))^2 \\ &\ge 4\pi\,(1+\eta)\,A(F_1(t)) \\ &\ge 4\pi\,(1+\eta)\,(1-\eta/10)\,A(t) \\ &\ge 4\pi\,(1+4\eta/5)\,A(t)\,. \end{aligned} \tag{6.5}$$

Since the right hand side of (6.4) is greater than that of (6.5) for $\eta < 0.01$, we find that if (6.2) or (6.3) were to fail, then at least (6.5) would hold.

If we were to repeat the steps in Lemma 5.1 leading to (5.8) we would get

$$\begin{aligned} \tau \le &\left(\int_{F(\tau)} |Du|^p dx\,dy\right)^{1/p} \\ &\cdot \left(\frac{1}{\sqrt{1+4\eta/5}}\left(\int_1^{1+\kappa\alpha^2} \phi(t)\,dt\right)^{(p-1)/p}\right)\,. \end{aligned} \tag{6.6}$$

Also, corresponding to (5.10) we would have

$$1-\tau \le \left(\int_{F(1)\setminus F(\tau)} |Du|^p dx\,dy\right)^{1/p} \left(\int_{1+\kappa\alpha^2}^{4} \phi(t)\,dt\right)^{(p-1)/p}\,. \tag{6.7}$$

Adding (6.6) and (6.7) we would obtain

$$
\begin{aligned}
1 \leq & \left(\int_{F(1)} |Du|^p dx\, dy \right)^{1/p} \\
& \cdot \left(1 - \left(1 - \left(\frac{1}{1+4\eta/5} \right)^{p/2(p-1)} \right) \frac{\int_1^{1+\kappa\alpha^2} \phi(t)\, dt}{\int_1^4 \phi(t)\, dt} \right)^{(p-1)/p} \\
& \cdot \left(\int_1^4 \phi(t)\, dt \right)^{(p-1)/p} .
\end{aligned}
$$

It is easy to see that (1.5) follows for an appropriate constant $K = K(\kappa, \eta, p)$.

Thus we may assume the existence of $t = t_0$ such that (6.1)-(6.3) hold. Then $F(t_0)$ has a simply connected component $F_1(t_0)$ such that (6.1)-(6.3) become

$$1 < A(t_0) < 1 + \kappa\, \alpha^2 , \tag{6.8}$$

$$A(F_1(t_0)) > (1 - \eta/10)\, A(t_0) , \tag{6.9}$$

and

$$L(\partial F_1(t_0))^2 < 4\pi\, (1+\eta)\, A(F_1(t_0)) . \tag{6.10}$$

Now, with T as in (2.12) and (2.13), $F_1(T)$ is a component of $F(T)$ having largest area and $F_2(T) = F(T) \backslash F_1(T)$. From (2.9) and (6.8), $T > t_0$ and $F(T)$ contains $F(t_0)$. From (2.13) and Proposition 2.2, it follows easily that

$$A(F_1(T)) \geq (1 - \kappa^2 \alpha^4)\, A(T) . \tag{6.11}$$

It is clear from (6.11) that $A(F_2(T)) \leq \kappa^2 \alpha^4 A(T)$. From (6.8), (6.9), (2.9) and (2.12) it follows that $F_1(t_0)$ cannot be completely contained in $F_2(T)$. Now, since $F_1(t_0)$ and $F_1(T)$ are both connected and $F_1(t_0) \subseteq F(T)$, it follows that

$$F_1(t_0) \subseteq F_1(T) \quad \text{and} \quad A(F_2(T)) < \kappa^2 \alpha^4 A(T) . \tag{6.12}$$

Let $\Omega_1 = F_1(T) \cap F(t_0)$. Then the set $F(t_0)\backslash\Omega_1$ is contained in $F_2(T)$. From (2.12) and (6.12) we have

$$A(F(t_0)\backslash\Omega_1) \leq A(F_2(T)) \leq \kappa^2\alpha^4 A(T) \leq 4\,\kappa^2\alpha^4 \leq \kappa\,\alpha^2\,.$$

Hence,

$$A(\Omega_1) \geq A(t_0) - \kappa\,\alpha^2 \geq 1 - \kappa\,\alpha^2\,. \tag{6.13}$$

Based on (6.8)-(6.11) we now form an auxiliary condenser with some observations on the geometry of the sets.

Now, by (2.2), $\partial F_1(T)$ lies between two circles $C_o = \{x : |x - x_o| = R_o\}$ and $C_i = \{x : |x - x_i| = R_i\}$, $R_o > R_i$, where by (2.12), (2.13) and (6.11),

$$\begin{aligned} R_o - R_i &\leq \frac{1}{\pi}\sqrt{L(\partial F_1(T))^2 - 4\pi A(F_1(T))} \\ &\leq \frac{1}{\pi}\sqrt{L(\partial F(T))^2 - 4\pi\,(1 - \kappa^2\alpha^4)\,A(T)} \\ &\leq \frac{1}{\pi}\sqrt{4\pi\,[(1 + \kappa\,\alpha^2) - (1 - \kappa^2\alpha^4)]\,A(T)} \\ &\leq 2\sqrt{\kappa}\,\alpha\,. \end{aligned} \tag{6.14}$$

In particular, the centers of C_o and C_i satisfy

$$|x_o - x_i| \leq 2\sqrt{\kappa}\;\alpha\,. \tag{6.15}$$

Also, by (2.3), (2.9), (2.12), (2.13) and (6.11),

$$\begin{aligned} R_o &\leq \frac{1}{2\pi}\big(L(\partial F_1(T)) + \sqrt{L(\partial F_1(T))^2 - 4\pi A(F_1(T))}\;\big) \\ &\leq \frac{1}{2\pi}\big(L(\partial F(T)) + \sqrt{L(\partial F(T))^2 - 4\pi\,(1 - \kappa^2\alpha^4)\,A(T)}\;\big) \\ &\leq \sqrt{\frac{A(T)}{\pi}}\;\big(\sqrt{1 + \kappa\,\alpha^2} + \sqrt{\kappa\,\alpha^2 + \kappa^2\alpha^4}\big) \\ &\leq \sqrt{\frac{1 + 2\eta}{\pi}}\;\big(1 + 3\sqrt{\kappa}\,\alpha\big)\,. \end{aligned} \tag{6.16}$$

Regarding the position of $F_1(t_0)$ in $F_1(T)$, we note that (6.8), (6.9), (6.10) and (2.4) imply that $F_1(t_0)$ contains a disc $B(\bar{x}, \bar{R}_i)$ where

$$\begin{aligned} \bar{R}_i &\geq (1 - \sqrt{\eta})\sqrt{1 - \eta/10}\;\sqrt{\frac{A(t_0)}{\pi}} \\ &\geq (1 - 1.1\sqrt{\eta})\;\sqrt{\frac{A(t_0)}{\pi}} \\ &\geq \frac{1 - 1.1\sqrt{\eta}}{\sqrt{\pi}}\,. \end{aligned} \tag{6.17}$$

Recalling that $\Omega_1 = F_1(T) \cap F(t_0)$ and comparing (6.12)-(6.17) we conclude that

$$\text{(6.18)} \qquad \begin{cases} \text{i)} & B(x_o, R_o) \supseteq F_1(T)\,, \\ \text{ii)} & B(x_o, R_o(1-\varepsilon)) \subseteq F_1(T)\,, \quad \varepsilon = 7.5\sqrt{\kappa}\alpha\,, \\ \text{iii)} & B(x_o, R_i') \subseteq \Omega_1\,, \end{cases}$$

where

$$\text{(6.19)} \qquad \sqrt{\frac{A(F_1(T))}{\pi}} - 2\sqrt{\kappa}\,\alpha \le R_i \le R_o \le \sqrt{1+2\eta}\,\frac{1+3\sqrt{\kappa}\,\alpha}{\sqrt{\pi}}\,,$$

and

$$\text{(6.20)} \qquad R_i' = 2\bar{R}_i - R_o \ge \frac{1 - 0.2\,\eta - 3\,(1+2\eta)\sqrt{\kappa}\,\alpha}{\sqrt{\pi}}\,.$$

By (6.8), (6.11), (6.13), and (2.12)

$$\text{(6.21)} \qquad \begin{cases} \text{i)} & 1 - \kappa\,\alpha^2 \le A(\Omega_1) \le 1 + \kappa\,\alpha^2, \\ \text{ii)} & (1 - \kappa^2\alpha^4)A(T) \le A(F_1(T)) \le A(T)\,, \\ \text{iii)} & 1 + \eta \le A(T) \le 1 + 2\eta\,. \end{cases}$$

It follows from (2.9) and (6.21) that

$$\text{(6.22)} \qquad 1 + 0.9\,\eta \le \frac{A(F_1(T))}{A(\Omega_1)} \le 1 + 2.1\,\eta\,.$$

If $B(x_o, \rho)$ has the same area as Ω_1 and $B(\tilde{x}, \sqrt{1/\pi})$ is such that $\alpha = A(\Omega \backslash B(\tilde{x}, \sqrt{1/\pi}))$, then by (1.1), (6.18) and (6.21)

$$\text{(6.23)} \qquad \begin{aligned} A(\Omega_1 \backslash B(x_o, \rho)) &\ge A(\Omega \backslash B(x_o, \rho)) - A(\Omega \backslash \Omega_1) \\ &\ge A(\Omega \backslash B(x_o, r)) \\ &\quad - A(B(x_o, \rho) \backslash B(x_o, r)) - A(\Omega \backslash \Omega_1) \\ &\ge A(\Omega \backslash B(\tilde{x}, r)) \\ &\quad - A(B(x_o, \rho) \backslash B(x_o, r)) - A(\Omega \backslash \Omega_1) \\ &\ge \alpha - \kappa\,\alpha^2 - \kappa\,\alpha^2 \\ &> \frac{\alpha}{2}\,, \end{aligned}$$

where $r = \sqrt{1/\pi}$. The third inequality follows from the definition of $\alpha(\Omega)$. Thus, if

$$\beta = \frac{A(\Omega_1 \backslash B(x_o, \rho))}{A(\Omega_1)}, \qquad \rho = \sqrt{\frac{A(\Omega_1)}{\pi}}, \tag{6.24}$$

we have, from (2.9), (6.23) and (6.21) i) that

$$\beta > \frac{\alpha}{2\,(1 + \kappa\,\alpha^2)} > \frac{\alpha}{3}. \tag{6.25}$$

We set $F_1 = F_1(T)$ for convenience, and let $u = u_p$ be the minimizer for (1.4). Clearly,

$$\int_{F(T)} |Du|^p dx\, dy \geq \int_{F_1 \backslash \Omega_1} |Du|^p dx\, dy. \tag{6.26}$$

Also, since ∂F_1 and $\partial \Omega_1$ are level sets for u, we may use u, renormalized, as the extremal for the condenser having inner set $\overline{\Omega_1}$ (closure of Ω) and outer set $\mathbb{R}^2 \backslash F_1$, and in this way estimate the right hand side of (6.26). For $p = 2$, this will be done by using Lemma 4.1, while for $p \neq 2$, the p-analogue (see Section 8) will be used.

In fact, with $u = t_0$ on $\partial\Omega_1$ and $u = T$ on ∂F_1, then

$$v = \frac{u - t_0}{T - t_0} \tag{6.27}$$

is the minimizer for

$$\int_{F_1 \backslash \Omega_1} |Dw|^p dx\, dy, \qquad w = \begin{cases} 1, & \text{on } \partial F_1, \\ 0, & \text{on } \partial \Omega_1. \end{cases}$$

Thus,

$$\begin{aligned} \inf_w \int_{F_1 \backslash \Omega_1} |Dw|^p dx\, dy &= \int_{F_1 \backslash \Omega_1} |Dv|^p dx\, dy \\ &= \frac{1}{(T - t_0)^p} \int_{F_1 \backslash \Omega_1} |Du|^p dx\, dy. \end{aligned} \tag{6.28}$$

Thus, with $\Gamma = \Gamma(\overline{\Omega}_1, \mathbb{R}^2 \backslash F_1)$ as the subcondenser, the next step in the proof of Theorem 1 is to obtain estimates for $\mathrm{Cap}_p(\Gamma)$. To this end, we first employ the symmetrization introduced in Section 3. Setting

$\rho = \sqrt{A(\Omega_1)/\pi}$ and $R = \sqrt{A(F_1)/\pi}$, and using the notations (3.1)-(3.6), we conclude from (3.7), (6.14), (6.16), (6.18)-(6.22) that

$$(6.29)\quad \begin{cases} \text{i)} & \text{if } \hat{s}(\theta) > \rho\,, \text{ then } \xi(\theta) > \rho \text{ and } \lambda(\theta) \le \rho\,, \\ \text{ii)} & R_i' \le \xi(\theta) \le R_o\,, \\ \text{iii)} & R_o(1-\varepsilon) \le R_i \le R \le R_o\,, \\ \text{iv)} & R_o(1-\varepsilon) \le t(\theta) \le R_o\,, \\ \text{v)} & R_i' \le \rho < R \le R_o\,, \\ \text{vi)} & |R^2 - t(\theta)^2| \le 2\,\varepsilon R_o^2\,, \\ \text{vii)} & 0 < \hat{t}(\theta) - s(\theta) \le \varepsilon R_o\,, \qquad \theta \in E\,, \\ \text{viii)} & \xi(\theta) \le s(\theta) \le \hat{t}(\theta) \le R_o\,, \\ \text{ix)} & R_i' \le \xi(\theta) < t(\theta) \le R_o\,, \\ \text{x)} & \sqrt{1+0.9\,\eta} \le R/\rho \le \sqrt{1+2.1\,\eta}\,, \\ \text{xi)} & R_o(1-\varepsilon) \le t(\theta) < s(\theta) < \hat{t}(\theta) \le R_o\,, \qquad \theta \in E\,. \end{cases}$$

In Section 7, we will prove Theorem 1 when $p = 2$. The details of the proof, when $p \neq 2$, together with the p-analogue of Lemma 4.1 will be presented in Section 8.

7. Proof of (1.5) for $p = 2$ in Case 2.

We now prove Theorem 1, in Case 2, when $p = 2$. We specify $\eta = .01$ when $p = 2$.

We now take a) $\Omega_1 = \Omega_1(t_0)$, $F_1 = F_1(T)$, $\rho = \sqrt{A(\Omega_1)/\pi}$, and $R = \sqrt{A(F_1)/\pi}$, and b) $R_i', R_i, R_o, \varepsilon$ and v as in (6.20), (6.19), (6.16), (6.18) and (6.27), and c) $x_o = 0$ in (6.18). As in Remark 4.1, we take $\delta = 0.9$, $\eta = 0.009$ (see (6.29.x))). These observations together with (6.29) imply that the hypotheses of Lemma 4.1 are satisfied. It is easily seen from (6.18) and (6.21) that

$$(7.1)\qquad \frac{1}{2}\log\frac{A(T)}{1-\kappa\,\alpha^2} \ge \log\frac{R}{\rho}\,.$$

We apply the conclusion of Lemma 4.1, together with (6.25)-(6.28), (7.1) and the definition of ε in (6.18), to conclude that there are absolute constants C and κ_1 such that $\kappa \le \kappa_1$,

$$\int_{F(T)} |Du|^2 dx\,dy \ge \int_{F_1\setminus\Omega_1} |Du|^2 dx\,dy$$

$$
\begin{aligned}
&\geq (T-t_0)^2\left(\frac{4\pi}{\log(A(T)/(1-\kappa\alpha^2))}\right.\\
&\qquad\qquad\left. + B_0\,\beta^2 - B_1\,\varepsilon^2 - B_2\,\varepsilon\,\beta\right) \qquad (7.2)\\
&\geq (T-t_0)^2(1+C\alpha^2)\,\frac{4\pi}{\log A(T)}\,.
\end{aligned}
$$

Henceforth, we take $\kappa \leq \kappa_1$.

To estimate t_0 in (7.2) we recall that $u = t_0$ on $\partial F(t_0)$, with t_0 as in (6.8) so that (*cf.* [3, p. 3])

$$\frac{1}{t_0^2}\int_{F(t_0)} |Du|^2 dx\,dy \geq \frac{4\pi}{\log A(t_0)}\,,$$

that is,

$$t_0^2 \leq \frac{1}{4\pi}\log(1+\kappa\,\alpha^2)\int_{F(t_0)} |Du|^2 dx\,dy\,. \tag{7.3}$$

By Green's theorem and the fact that u is harmonic,

$$
\begin{aligned}
\int_{F(t_0)} |Du|^2 dx\,dy &= t_0 \int_{\partial F(t_0)} \frac{\partial u}{\partial n}\,ds\\
&= t_0 \int_{\partial F(1)} \frac{\partial u}{\partial n}\,ds \qquad (7.4)\\
&= t_0\,\mathrm{Cap}_2(\Gamma)\,.
\end{aligned}
$$

Thus, from (7.3) and (7.4) we have,

$$t_0 \leq \frac{\kappa\,\alpha^2}{4\pi}\mathrm{Cap}_2(\Gamma) := M\,. \tag{7.5}$$

We now have two cases to examine, namely, i) $T > M$, and ii) $T \leq M$.

First we work out case i). From (7.2),

$$\int_{F(T)} |Du|^2 dx\,dy \geq \frac{4\pi\,(T-M)^2}{\log A(T)}(1+C\alpha^2)\,, \tag{7.6}$$

We now use the usual isoperimetric inequality for $T < t < 1$ as was done in (5.10) to obtain

$$1 - T \leq \left(\int_{F(1)\setminus F(T)} |Du|^2 dx\,dy\right)^{1/2}\left(\frac{1}{4\pi}\log\frac{4}{A(T)}\right)^{1/2}.$$

This together with (7.6) and Hölder's inequality gives

$$
\begin{aligned}
4\pi\,(1-M)^2 &\le \left(\int_{F(1)} |Du|^2 dx\,dy\right) \\
&\quad\cdot\left(\log\frac{4}{A(T)} + \frac{1}{1+C\alpha^2}\,\log A(T)\right) \\
&= \left(\int_{F(1)} |Du|^2 dx\,dy\right) \\
&\quad\cdot\left(\log 4 - \frac{C\alpha^2}{1+C\alpha^2}\,\log A(T)\right) \\
&\le \left(\int_{F(1)} |Du|^2 dx\,dy\right) \\
&\quad\cdot\left(1 - \frac{C\alpha^2}{1+C\alpha^2}\,\frac{\log A(T)}{\log 4}\right)\log 4\,.
\end{aligned}
\tag{7.7}
$$

Now set $G = \mathrm{Cap}_2(\Gamma)/\mathrm{Cap}_2(\Gamma^*)$. Then $G \ge 1$. Recalling that $\mathrm{Cap}_2(\Gamma^*) = 4\pi/\log 4$, (7.7), (2.12), and $\eta = 0.01$ yield

$$(1-M)^2 \le G\left(1 - \frac{C\alpha^2}{2}\,\frac{\log 1.01}{\log 4}\right).$$

This together with (7.5) gives

$$1 - \frac{\kappa\,\alpha^2 G}{\log 4} \le \sqrt{G\,(1-C_1\alpha^2)} \le G\,(1-C_2\alpha^2)\,.$$

Thus,

$$G \ge \frac{1}{1 - C_2\alpha^2 + \kappa\,\alpha^2/\log 4}\,.$$

For sufficiently small κ we then have

$$\mathrm{Cap}_2(\Gamma) \ge (1 + C_3\alpha^2)\,\mathrm{Cap}_2(\Gamma^*)\,. \tag{7.8}$$

We now examine case ii), *i.e.* $T \le M$. Observe that

$$\int_{F(1)} |Du|^2 dx\,dy = \frac{1}{T}\int_{F(T)} |Du|^2 dx\,dy\,.$$

Now, from (7.5) we deduce that

$$T \le \frac{\kappa\,\alpha^2}{4\pi}\,\frac{1}{T}\,\int_{F(T)} |Du|^2 dx\,dy\,,$$

which in turn implies,

$$T \le \alpha\sqrt{\frac{\kappa}{4\pi}}\left(\int_{F(T)} |Du|^2 dx\,dy\right)^{1/2}. \tag{7.9}$$

By employing a procedure, similar to the one used in deriving (5.10), we again write

$$1 - T \le \left(\int_{F(1)\setminus F(T)} |Du|^2 dx\,dy\right)^{1/2}\left(\frac{1}{4\pi}\log\left(\frac{4}{A(T)}\right)\right)^{1/2}. \tag{7.10}$$

Adding (7.9) and (7.10), using (2.10) and $\eta = 0.01$, and applying Hölder's inequality we have

$$\begin{aligned}
1 &\le \left(\int_{F(1)} |Du|^2 dx\,dy\right)^{1/2} \\
&\quad\cdot\left(\frac{1}{4\pi}\log\left(\frac{4}{A(T)}\right) + \frac{\kappa\,\alpha^2}{4\pi}\right)^{1/2} \\
&= \left(\int_{F(1)} |Du|^2 dx\,dy\right)^{1/2} \\
&\quad\cdot\left(\frac{\log 4}{4\pi} + \frac{\kappa\,\alpha^2}{4\pi} - \frac{\log A(T)}{4\pi}\right)^{1/2} \\
&\le \left(\int_{F(1)} |Du|^2 dx\,dy\right)^{1/2} \\
&\quad\cdot\left(1 + \frac{\kappa\,\alpha^2}{\log 4} - \frac{\log 1.01}{\log 4}\right)^{1/2}\left(\frac{\log 4}{4\pi}\right)^{1/2}.
\end{aligned} \tag{7.11}$$

For sufficiently small κ, (7.11) then yields

$$\frac{4\pi}{\log 4} \le \left(1 - \frac{\log 1.01}{2\,\log 4}\right)\left(\int_{F(1)} |Du|^2 dx\,dy\right),$$

which implies (1.5) trivially, that is, with no dependence on α. Thus we have shown that (1.5) holds when $T > M$ and $T \le M$, so the proof of (1.5) is complete for $p = 2$.

8. Remarks on Case 2 for $p \neq 2$.

The procedure for obtaining the analogue of Lemma 4.1 will now follow for general p, with different constants, much as was done in Section 4. Inequality (4.3) becomes

$$(8.1)\qquad \begin{aligned} I &= \int_{F_1\setminus\Omega_1} (v_r^2 + \frac{1}{r^2}\, v_\theta^2)^{p/2}\, r\, dr\, d\theta \\ &\ge \int_{-\pi}^{\pi} \left(\inf \int_{J(\theta)\cap\{F_1\setminus\Omega_1\}} |f_r|^p\, r\, dr \right) d\theta\,, \end{aligned}$$

where $f = f(r,\theta)$ is absolutely continuous and $f = 1$ on $J(\theta) \cap \partial F_1$ and $f = 0$ on $J(\theta) \cap \partial\Omega_1$. We then use the solution to the one variable Euler equation $(r|z'|^{p-2}z')' = 0$ and (4.4) becomes

$$(8.2)\quad I \ge |d|^{p-1} \left(\int_{-\pi}^{\pi} \frac{d\theta}{|t(\theta)^d - \xi(\theta)^d|^{p-1}} + \int_E \frac{d\theta}{|\hat{t}(\theta)^d - s(\theta)^d|^{p-1}} \right),$$

where $d = (p-2)/(p-1)$. This follows from the observation that for $d \neq 0$ and $\xi(\theta) \le \hat{s}(\theta) \le t(\theta)$,

$$|t(\theta)^d - \xi(\theta)^d|^{p-1} \ge |t(\theta)^d - \hat{s}(\theta)^d|^{p-1}\,.$$

Our objective is to prove the analogue

$$(8.3)\qquad I \ge |d|^{p-1} \frac{2\pi}{|R^d - \rho^d|^{p-1}}\, (1 + K_1\, \beta^2 - K_2\, \varepsilon^2 - K_3\, \varepsilon\, \beta)$$

of Lemma 4.1, where the constants K_1, K_2, and K_3 now depend only on p for small ε. We first consider the case $p > 2$. We write

$$(8.4)\qquad \begin{aligned} \left(\frac{1}{t^d - \xi^d}\right)^{p-1} &= \left(\frac{1}{R^d - \rho^d}\right)^{p-1} \\ &\quad \cdot \left(1 - \frac{(R^d - t^d) - (\rho^d - \xi^d)}{(R^d - \rho^d)}\right)^{1-p}. \end{aligned}$$

Now the condition (2.9) and (6.29) already imply that t/R and ξ/ρ are close to 1; certainly

$$\frac{1}{2} < \frac{\xi}{\rho}, \qquad \frac{t}{R} < 2\,.$$

In addition, by (2.9), (6.19), (6.20), (6.29.v))-(6.29.ix-x)), we also have

$$0 < \frac{t^d - \xi^d}{R^d - \rho^d} \leq \sigma \tag{8.6}$$

for some constant $\sigma = \sigma_p > 0$, which depends only on p.

Let $h(x) = (1-x)^{1-p}$. Then, $h(0) = 1$, $h'(x) = (p-1)(1-x)^{-p}$, and $h''(x) = p\,(p-1)(1-x)^{-p-1}$ which is positive and increasing for $-\infty < x < 1$. Using these on the interval $[1-\sigma, 1)$, we find that

$$h(x) \geq 1 + (p-1)\,x + h''(1-\sigma)\,\frac{x^2}{2}, \qquad 1-\sigma \leq x < 1\,. \tag{8.7}$$

Combining (8.4), (8.6) and (8.7), we may then write

$$\begin{aligned}\left(\frac{1}{t^d - \xi^d}\right)^{p-1} \geq \left(\frac{1}{R^d - \rho^d}\right)^{p-1} \\ \cdot \left(1 + (p-1)\left(\left(\frac{R^d - t^d + \xi^d - \rho^d}{R^d - \rho^d}\right)\right.\right. \\ \left.\left. + \frac{p}{2}\,\sigma^{-p-1}\left(\frac{R^d - t^d + \xi^d - \rho^d}{R^d - \rho^d}\right)^2\right)\right).\end{aligned} \tag{8.8}$$

In (8.8), we shall use the following four expansions with (8.5). First we have

$$\begin{aligned}\int_{-\pi}^{\pi} \frac{R^d - t^d}{R^d - \rho^d}\,d\theta &= \frac{R^d}{R^d - \rho^d}\int_{-\pi}^{\pi}\left(1 - \left(\left(\frac{t}{R}\right)^2\right)^{d/2}\right)d\theta \\ &\geq \frac{d\,R^d}{2\,(R^d - \rho^d)}\int_{-\pi}^{\pi}\frac{R^2 - t^2}{R^2}\,d\theta \geq 0\,.\end{aligned} \tag{8.9}$$

The fact that the right hand side is nonnegative follows from (3.12).

Also,

$$\begin{aligned}\int_{-\pi}^{\pi} \frac{\xi^d-\rho^d}{R^d-\rho^d}\,d\theta &= \frac{\rho^d}{R^d-\rho^d} \\ &\quad\cdot \int_{-\pi}^{\pi}\left(\left(\left(\frac{\xi}{\rho}\right)^2\right)^{d/2}-1\right)d\theta \\ &\geq \frac{d\,\rho^d}{2\,(R^d-\rho^d)}\int_{-\pi}^{\pi}\frac{\xi^2-\rho^2}{\rho^2}\,d\theta \\ &\quad-\ 2^{2-d}d\left(1-\frac{d}{2}\right)\left(\frac{\rho^d}{R^d-\rho^d}\right) \\ &\quad\cdot\int_{-\pi}^{\pi}\left(\frac{\xi^2-\rho^2}{\rho^2}\right)^2 d\theta\,, \end{aligned} \tag{8.10}$$

and

$$\begin{aligned}\int_{-\pi}^{\pi}\left(\frac{\xi^d-\rho^d}{R^d-\rho^d}\right)^2 d\theta &= \left(\frac{\rho^d}{R^d-\rho^d}\right)^2 \\ &\quad\cdot\int_{-\pi}^{\pi}\left(\left(\left(\frac{\xi}{\rho}\right)^2\right)^{d/2}-1\right)^2 d\theta \\ &\geq d^2\,4^{d-3}\left(\frac{\rho^d}{R^d-\rho^d}\right)^2\int_{-\pi}^{\pi}\left(\frac{\xi^2-\rho^2}{\rho^2}\right)^2 d\theta\,. \end{aligned} \tag{8.11}$$

Similarly,

$$\begin{aligned}\int_{-\pi}^{\pi}\left(\frac{R^d-t^d}{R^d-\rho^d}\right)^2 d\theta &\geq d^2\,4^{d-3}\frac{\rho^d R^d}{(R^d-\rho^d)^2} \\ &\quad\cdot\int_{-\pi}^{\pi}\left(\frac{R^2-t^2}{R^2}\right)^2 d\theta\,. \end{aligned} \tag{8.12}$$

Using (8.8) in (8.2) we obtain

$$\begin{aligned}I &\geq \frac{d^{p-1}}{(R^d-\rho^d)^{p-1}} \\ &\quad\cdot\Bigg(2\pi+(p-1)\Bigg(\int_{-\pi}^{\pi}\Bigg(\frac{\xi^d-\rho^d}{R^d-\rho^d}+\frac{R^d-t^d}{R^d-\rho^d} \\ &\qquad\qquad+\frac{p}{2}\,\sigma^{-p-1}\left(\frac{\xi^d-\rho^d}{R^d-\rho^d}+\frac{R^d-t^d}{R^d-\rho^d}\right)^2\Bigg)\,d\theta\Bigg)\Bigg)\end{aligned}$$

$$\geq \frac{d^{p-1}}{(R^d-\rho^d)^{p-1}}$$
$$\cdot \Bigg(2\pi + (p-1)\Bigg(\int_{-\pi}^{\pi}\Bigg(\frac{\xi^d-\rho^d}{R^d-\rho^d} + \frac{R^d-t^d}{R^d-\rho^d}$$
$$+ \frac{p}{2}\,\sigma^{-p-1}\Bigg(\Bigg(\frac{\xi^d-\rho^d}{R^d-\rho^d}\Bigg)^2 + \Bigg(\frac{R^d-t^d}{R^d-\rho^d}\Bigg)^2$$
$$-2\left|\frac{\xi^d-\rho^d}{R^d-\rho^d}\right|\left|\frac{R^d-t^d}{R^d-\rho^d}\right|\Bigg)\Bigg)\,d\theta\Bigg)\Bigg).$$

We now use the inequalities in (8.9)-(8.12) to estimate I. It follows that

$$I \geq \frac{d^{p-1}}{(R^d-\rho^d)^{p-1}}\,(2\pi + T_1 + T_2 + T_3 + T_4)\,, \tag{8.13}$$

where,

$$T_1 = \frac{(p-1)\,d}{2}\,\frac{\rho^d}{R^d-\rho^d}\int_{-\pi}^{\pi}\Big(\frac{R^2-t^2}{R^2} + \frac{\xi^2-\rho^2}{\rho^2}\Big)\,d\theta\,,$$

$$T_2 = (p-1)\Bigg(\frac{p}{2}\,\sigma^{-p-1}d^2 4^{d-3}\Bigg(\frac{\rho^d}{R^d-\rho^d}\Bigg) - 2^{2-d}d\,\big(1-\frac{d}{2}\big)\Bigg)$$
$$\cdot\,\frac{\rho^d}{R^d-\rho^d}\int_{-\pi}^{\pi}\Bigg(\frac{\xi^2-\rho^2}{\rho^2}\Bigg)^2 d\theta\,,$$
$$T_3 = (p-1)\,\frac{p}{2}\,\sigma^{-p-1}d^2 4^{d-3}\frac{\rho^d R^d}{(R^d-\rho^d)^2}\int_{-\pi}^{\pi}\Bigg(\frac{R^2-t^2}{R^2}\Bigg)^2 d\theta\,,$$

and

$$T_4 = -p\,(p-1)\,\sigma^{-p-1}\frac{\rho^d R^d}{(R^d-\rho^d)^2}\int_{-\pi}^{\pi}\left|\Bigg(\Bigg(\frac{\xi}{\rho}\Bigg)^2\Bigg)^{d/2} - 1\right|$$
$$\cdot\left|\Bigg(\Bigg(\frac{t}{R}\Bigg)^2\Bigg)^{d/2} - 1\right| d\theta\,.$$

Now, for some $C_1 > 0$, $T_1 \geq -C_1 A(N)/\rho^2$ by (3.12) and (4.10), nd $T_3 \geq 0$. We may estimate T_4 by using, (6.29.vi)), (3.11), and (8.5) to obtain

$$|T_4| \leq 200\,p\,(p-1)\,\sigma^{-p-1}\frac{R^d\rho^d}{(R^d-\rho^d)^2}\,\varepsilon\left(\beta + \frac{A(N)}{\pi\rho^2}\right).$$

It is at this stage that we constrain our parameter η for each $p \neq 2$. We now assume that η is sufficiently small so that

$$T_2 \geq \frac{p\,(p-1)}{4}\, d^2 4^{d-3} \sigma^{-p-1} \left(\frac{\rho^d}{R^d - \rho^d} \right)^2 \int_{-\pi}^{\pi} \left(\frac{\xi^2 - \rho^2}{\rho^2} \right)^2 d\theta\,. \tag{8.14}$$

This is possible due to (6.29.x)). Using these estimates in (8.13) along with (4.15) we then obtain

$$\begin{aligned} I \geq \frac{d^{p-1}}{(R^d - \rho^d)^{p-1}} \Bigg(2\pi + C_3 \left(\beta - \frac{A(N)}{\pi \rho^2} \right)^2 \\ - C_2 \beta \varepsilon - C_1 \frac{A(N)}{\rho^2} \Bigg)\,, \end{aligned} \tag{8.15}$$

Finally, we need an estimate for $A(N)$. We first make a preliminary estimate using (8.4), (8.8), (8.9), and ignoring the second order term in (8.8). Observe that from (8.5), $|(\xi^2 - \rho^2)/\rho^2| \leq 4$. Using this and (3.11) in (8.10), (8.8) yields

$$\int_{-\pi}^{\pi} \frac{d\theta}{|t(\theta)^d - \xi(\theta)^d|^{p-1}} \geq \frac{2\pi}{(R^d - \rho^d)^{p-1}} \,(1 - C_4 \beta)\,. \tag{8.16}$$

If

$$\begin{aligned} \int_{-\pi}^{\pi} \frac{d\theta}{|t(\theta)^d - \xi(\theta)^d|^{p-1}} + \int_E \frac{d\theta}{|\hat{t}(\theta)^d - s(\theta)^d|^{p-1}} \\ \geq \frac{2\pi}{(R^d - \rho^d)^{p-1}} \,(1 + C_4\, \beta)\,, \end{aligned}$$

then (8.3) follows trivially. Otherwise, from (8.16) we have

$$\int_E \frac{d\theta}{|\hat{t}(\theta)^d - s(\theta)^d|^{p-1}} \leq \frac{4\pi}{(R^d - \rho^d)^{p-1}} \, C_4\, \beta\,.$$

Using (6.29.vii)) to estimate $A(N)$ as in Section 4, we then obtain

$$A(N) \leq C_5\, \varepsilon^p\, \beta\, R_o^2\,. \tag{8.17}$$

Using (8.17) in (8.15) and fixing η so that (8.14) holds, we then obtain (8.3) with constants depending only on p.

A similar analysis can be carried out for $1 < p < 2$.

Finally, we give the analogue of Section 4 for $p \neq 2$. Now,

$$
\begin{aligned}
\frac{2\pi |d|^{p-1}}{|R^d - \rho^d|^{p-1}} &= \left(\int_{A(\Omega_1)}^{A(F_1)} \phi(t)\, dt \right)^{1-p} \\
&\geq \left(\int_{(1-\kappa\alpha^2)}^{A(T)} \phi(t)\, dt \right)^{1-p} \\
&\geq \left(\int_{1}^{A(T)} \phi(t)\, dt \right)^{1-p} (1 - C_6\, \kappa\, \alpha^2)\, .
\end{aligned}
\tag{8.18}
$$

By (8.3) and (8.18), there exist constants C_7 and κ_1 such that for $0 < \kappa \leq \kappa_1$, we have

$$
\begin{aligned}
\int_{F(T)} |Du|^p dx\, dy &\geq \int_{F_1 \setminus \Omega_1} |Du|^p dx\, dy \\
&\geq \frac{2\pi |d|^{p-1}}{|R^d - \rho^d|^{p-1}} \\
&\quad \cdot (T - t_0)^p (1 + K_1\, \beta^2 - K_2\, \varepsilon^2 - K_3\, \varepsilon\, \beta) \\
&\geq (T - t_0)^p\ (1 + C_7 \alpha^2) \\
&\quad \cdot \left(\int_{1}^{A(T)} \phi(t)\, dt \right)^{1-p} .
\end{aligned}
\tag{8.19}
$$

To estimate t_0 in (8.19), we recall that $u = t_0$ on $\partial F(t_0)$ with t_0 as in (6.8), so that

$$
\frac{1}{t_0^p} \int_{F(t_0)} |Du|^p dx\, dy \geq \left(\int_{1}^{A(t_0)} \phi(t)\, dt \right)^{1-p} .
$$

Hence,

$$
\begin{aligned}
t_0^p &\leq \left(\int_{F(t_0)} |Du|^p dx\, dy \right) \left(\int_{1}^{1+\kappa\alpha^2} \phi(t)\, dt \right)^{p-1} \\
&\leq C_8\, (\kappa\, \alpha^2)^{p-1} \int_{F(t_0)} |Du|^p dx\, dy\, .
\end{aligned}
\tag{8.20}
$$

By Green's theorem,

$$
\int_{F(t_0)} |Du|^p dx\, dy = t_0 \int_{\partial F(t_0)} |Du|^{p-2} \frac{\partial u}{\partial n}\, ds = t_0\, \mathrm{Cap}_p(\Gamma)\, . \tag{8.21}
$$

By (8.20) and (8.21),

$$t_0 \le C_9\,\kappa\,\alpha^2\,\mathrm{Cap}_p(\Gamma)^{1/(p-1)} := M\,, \qquad C_9 = C_8^{1/(p-1)}\,. \tag{8.22}$$

As in Section 7, we distinguish two possibilities, namely, i) $T > M$, and ii) $T \le M$. Let us first assume that i) holds. Thus for $0 < \kappa \le \kappa_1$, (8.19) yields

$$\int_{F(T)} |Du|^p \ge (T-M)^p\,(1 + C_7\,\alpha^2)\,\left(\int_1^{A(T)} \phi(t)\,dt\right)^{1-p}\,. \tag{8.23}$$

We may now use the usual isoperimetric inequality over the interval $(T, 1)$ to obtain

$$1 - T \le \left(\int_{F(1)\setminus F(T)} |Du|^p\right)^{1/p} \left(\int_{A(T)}^4 \phi(t)\,dt\right)^{(p-1)/p}\,.$$

This together with (8.23) and Hölder's inequality gives us

$$\begin{aligned}(1-M)^p &\le \left(\int_{F(1)} |Du|^p dx\,dy\right)\\ &\quad\cdot\left(\left(\frac{1}{1+C_7\alpha^2}\right)^{1/(p-1)}\int_1^{A(T)} \phi(t)\,dt + \int_{A(T)}^4 \phi(t)\,dt\right)^{p-1}\\ &= \left[1 + \left(\left(\frac{1}{1+C_7\alpha^2}\right)^{1/(p-1)} - 1\right)\frac{\displaystyle\int_1^{A(T)} \phi(t)\,dt}{\displaystyle\int_1^4 \phi(t)\,dt}\right]^{p-1}\\ &\quad\cdot\left(\int_{F(1)} |Du|^p dx\,dy\right)\left(\int_1^4 \phi(t)\,dt\right)^{p-1}\,.\end{aligned} \tag{8.24}$$

Set Z to be the square bracket term on the right hand side of (8.24), and take $S = \mathrm{Cap}_p(\Gamma)/\mathrm{Cap}_p(\Gamma^*)$. Then $S \ge 1$, and (8.24) says that $(1-M) \le S^{1/p} Z^{1/p}$, or by (8.22),

$$1 - C_9\,\kappa\,\alpha^2\,S^{1/(p-1)}\,\mathrm{Cap}_p(\Gamma^*)^{1/(p-1)} \le S^{1/p}\,Z^{1/p}\,.$$

Since $S^{1/(p-1)} \geq S^{1/p}$, it follows that

$$S^{1/(p-1)} \geq \frac{1}{Z^{1/p} + C_9\ \kappa\,\alpha^2\ \mathrm{Cap}_p(\Gamma^*)^{1/(p-1)}}\,.$$

This in turn implies,

$$\mathrm{Cap}_p(\Gamma) \geq \left(\frac{1}{Z^{1/p} + C_9\,\kappa\,\alpha^2\ \mathrm{Cap}_p(\Gamma^*)^{1/(p-1)}}\right)^{p-1}\ \mathrm{Cap}_p(\Gamma^*)\,. \tag{8.25}$$

Since it is easy to see that $Z \leq 1 - C_{10}\alpha^2$, the result then follows from (8.25) for sufficiently small κ.

We next consider case ii), *i.e.*, $T \leq M$. Now,

$$\int_{F(1)} |Du|^p dx\,dy = \frac{1}{T}\ \int_{F(T)} |Du|^p dx\,dy,$$

so that by (8.22),

$$T \leq C_9\,\kappa\,\alpha^2 \left(\frac{1}{T}\ \int_{F(T)} |Du|^p dx\,dy\right)^{1/(p-1)}\,.$$

Hence,

$$T \leq (C_9\,\kappa\,\alpha^2)^{(p-1)/p}\ \left(\int_{F(T)} |Du|^p dx\,dy\right)^{1/p}\,. \tag{8.26}$$

We employ the usual isoperimetric inequality and the coarea formula over the interval $(T,1)$ (see Section 5) to obtain

$$1 - T \leq \left(\int_{F(1)\setminus F(T)} |Du|^p dx\,dy\right)^{1/p} \left(\int_{A(T)}^{1} \phi(t)\,dt\right)^{(p-1)/p}\,.$$

This together with (8.26), (2.12), and Hölder's inequality results in

$$1 \leq \left(\int_{F(1)} |Du|^p dx\,dy\right) \left(C_9\,\kappa\,\alpha^2 + \int_{A(T)}^{1} \phi(t)\,dt\right)^{p-1}$$

$$\leq \left(\int_{F(1)} |Du|^p dx\, dy \right) \left(1 + \frac{C_9\, \kappa\, \alpha^2}{\displaystyle\int_1^4 \phi(t)\, dt} - \frac{\displaystyle\int_1^{A(T)} \phi(t)\, dt}{\displaystyle\int_1^4 \phi(t)\, dt} \right)^{p-1}$$

$$\cdot \left(\int_1^4 \phi(t) dt \right)^{p-1}$$

$$\leq \left(\int_{F(1)} |Du|^p dx\, dy \right) \left(1 + \frac{C_9\, \kappa\, \alpha^2 - \displaystyle\int_1^{1+\eta} \phi(t)\, dt}{\displaystyle\int_1^4 \phi(t)\, dt} \right)^{p-1}$$

$$\cdot \left(\int_1^4 \phi(t)\, dt \right)^{p-1},$$

which again gives the result for κ sufficiently small. Thus, the proof of Theorem 1 is complete for $p \neq 2$.

9. Sharpness of the exponent 2.

In this section we show that the condenser with elliptical inner set of small eccentricity gives the proper order of magnitude for capacity to show that the exponent 2 is sharp. Although there is no reason to believe that this case gives the sharp constant K_p in Theorem 1, it is convenient from the standpoint of calculations. On the other hand, there is some delicacy in choosing the inner set. For example, putting a small bump or a circle would result in an exponent of 1 instead of 2 on α.

Let ε be a small positive number. For each ε, let E_ε denote the closed domain bounded by the ellipse $x = r_0(1+\varepsilon)^{1/2}\cos\theta$, $y = r_0 \sin\theta$, where $r_0 = 1/(\sqrt{\pi}\,(1+\varepsilon)^{1/4})$. Then $A(E_\varepsilon) = 1$. Let Γ_ε denote the condenser $\Gamma(E_\varepsilon, \mathbb{R}^2 \backslash B(0, 2/\sqrt{\pi}))$. From [6, p. 88-89] we have that $\alpha = \alpha(E_\varepsilon) = \varepsilon/2\pi + O(\varepsilon^2)$, as $\varepsilon \to 0$. In order to prove our claim, we note from (1.4) and (1.5) that it is sufficient to exhibit a function u, belonging to the class of admissible functions for (1.4), with the property that

$$\iint_{\mathbb{R}^2} |\nabla u|^p dx\, dy = \mathrm{Cap}_p(\Gamma^*) + O(\varepsilon^2) \qquad \text{as } \varepsilon \to 0\,,$$

where Γ^* is as in Theorem 1. This will then imply that

$$\mathrm{Cap}_p(\Gamma_\varepsilon) = \mathrm{Cap}_p(\Gamma^*) + O(\varepsilon^2) \qquad \text{as } \varepsilon \to 0\,. \tag{9.1}$$

Theorem 2. *Let $\varepsilon > 0$, be small, Γ_ε be the condenser whose inner set is E_ε and outer set is $\mathbb{R}^2 \backslash B(0, 2/\sqrt{\pi})$. Then for each fixed $p > 1$, there is a function $u = u_{\varepsilon,p}$ with $u = 0$ on E_ε and $u = 1$ on $\mathbb{R}^2 \backslash B(0, 2/\sqrt{\pi})$, such that*

$$\iint_{\mathbb{R}^2} |\nabla u|^p dx\, dy = \mathrm{Cap}_p(\Gamma^*) + O(\varepsilon^2) \qquad \text{as } \varepsilon \to 0\,. \tag{9.2}$$

Proof. We shall present details for $p \neq 2$; the case $p = 2$ is similar. Set $R = 2/\sqrt{\pi}$ and $\rho = 1/\sqrt{\pi}$. Then $r_0 = \rho/(1+\varepsilon)^{1/4}$. By (1.6),

$$\mathrm{Cap}_p(\Gamma^*) = \frac{2\pi |d|^{p-1}}{|R^d - \rho^d|^{p-1}}\,, \tag{9.3}$$

where $d = (p-2)/(p-1)$.

Let r, θ be the polar coordinates, and define $u(r,\theta) = u_{\varepsilon,p}(r,\theta)$ as

$$u(r,\theta) = 1 - \frac{R^d - r^d}{R^d - r_0^d\,(1+\varepsilon\cos^2\theta)^{d/2}}\,, \tag{9.4}$$

in $B(0, 2/\sqrt{\pi})\backslash E_\varepsilon$, $u = 0$ on E_ε, and $u = 1$ on $\mathbb{R}^2 \backslash B(0, 2/\sqrt{\pi})$. Then u is absolutely continuous, and in $B(0, 2/\sqrt{\pi})\backslash E_\varepsilon$,

$$|\nabla u| = \frac{|d|\, r^{d-1}}{|R^d - r_0^d\,(1+\varepsilon\cos^2\theta)^{d/2}|} + O(\varepsilon^2) \qquad \text{as } \varepsilon \to 0\,. \tag{9.5}$$

Then, by (9.5),

$$\begin{aligned}
&\iint_{\mathbb{R}^2} |\nabla u|^p dx\, dy \\
&\quad = |d|^p \int_0^{2\pi} \int_{r_0\sqrt{1+\varepsilon\cos^2\theta}}^{R} \frac{r^{p/(1-p)}}{|R^d - r_0^d\,(1+\varepsilon\cos^2\theta)^{d/2}|^p}\, r\, dr\, d\theta \\
&\qquad + O(\varepsilon^2) \\
&\quad = |d|^{p-1} \int_0^{2\pi} \frac{|R^d - r_0^d\,(1+\varepsilon\cos^2\theta)^{d/2}|}{|R^d - r_0^d\,(1+\varepsilon\cos^2\theta)^{d/2}|^p}\, d\theta \\
&\qquad + O(\varepsilon^2) \\
&\quad = |d|^{p-1} \int_0^{2\pi} \frac{1}{|R^d - r_0^d\,(1+\varepsilon\cos^2\theta)^{d/2}|^{p-1}}\, d\theta \\
&\qquad + O(\varepsilon^2)\,,
\end{aligned} \tag{9.6}$$

as $\varepsilon \to 0$. By the definition of r_0 and ρ,

$$
\begin{aligned}
|R^d - r_0^d (1+\varepsilon \cos^2\theta)^{d/2}| &= \left| R^d - \rho^d \left(\frac{1+\varepsilon\cos^2\theta}{\sqrt{1+\varepsilon}} \right)^{d/2} \right| \\
&= \left| R^d - \rho^d + \rho^d \left(1 - \left(\frac{1+\varepsilon\cos^2\theta}{\sqrt{1+\varepsilon}} \right)^{d/2} \right) \right| . \qquad (9.7)
\end{aligned}
$$

Set

$$
h(\varepsilon) = 1 - \left(\frac{1+\varepsilon\cos^2\theta}{\sqrt{1+\varepsilon}} \right)^{d/2} .
$$

Now,

$$
h(\varepsilon) = -\frac{d}{2} (\cos^2\theta - \frac{1}{2})\, \varepsilon + O(\varepsilon^2) \qquad \text{as } \varepsilon \to 0 . \qquad (9.8)
$$

Thus, (9.7) and (9.8) imply, as $\varepsilon \to 0$,

$$
\begin{aligned}
|R^d - r_0^d (1+\varepsilon\cos^2\theta)^{d/2}|^{1-p} &= |R^d - \rho^d|^{1-p} \Big(1 + \frac{\rho^d\, h(\varepsilon)}{(R^d-\rho^d)} \Big)^{1-p} \\
&= |R^d - \rho^d|^{1-p} \Big(1 - \frac{(p-1)\,\rho^d\, h(\varepsilon)}{(R^d-\rho^d)} \Big) \\
&\quad + O(\varepsilon^2) \\
&= |R^d - \rho^d|^{1-p} \\
&\quad \cdot \Big(1 + \frac{(p-1)\, d\, \rho^d\, (\cos^2\theta - 1/2)}{2\,(R^d-\rho^d)} \varepsilon \Big) \\
&\quad + O(\varepsilon^2) . \qquad (9.9)
\end{aligned}
$$

Using (9.9) in (9.6), we have, as $\varepsilon \to 0$,

$$
\begin{aligned}
\iint_{\mathbb{R}^2} |\nabla u|^p dx\, dy = \frac{|d|^{p-1}}{|R^d - \rho^d|^{p-1}} \int_0^{2\pi} 1 + \frac{d\,(p-1)\,\rho^d}{2\,(R^d-\rho^d)} \Big(\cos^2\theta - \frac{1}{2} \Big)\, \varepsilon\, d\theta \\
+ O(\varepsilon^2) .
\end{aligned}
$$

Since

$$
\int_0^{2\pi} \Big(\cos^2\theta - \frac{1}{2} \Big)\, d\theta = 0 ,
$$

we obtain (9.2).

10. Logarithmic Capacity.

We now outline the proof of (1.3). Let Ω be a compact subset of the complex plane $\mathbb{C}$ with $\partial\Omega$ a finite union of rectifiable curves. Let $G(z)$ denote Green's function for $\hat{\mathbb{C}}\backslash\Omega$ with pole at ∞, extended to be 0 on Ω. Then

$$-\log \operatorname{Cap}(\Omega) = \lim_{z\to\infty} \left(G(z) - \log|z|\right) . \tag{10.1}$$

For $\lambda > 0$, let $\Omega_\lambda = \{z : G(z) \le \lambda\}$. Then $G(z) - \lambda$ is Green's function for the complement of Ω_λ. Let Γ_λ be the condenser $\Gamma(\Omega, \mathbb{C}\backslash\Omega_\lambda)$. The definition of $\operatorname{Cap}(\Gamma_\lambda)$ is as given in (1.4) with $p = 2$. In this instance, the minimizer is harmonic and is given by $G(z)/\lambda$. For $0 < t \le \lambda$, write $F(t) = \{z : G(z) < t\}$, and $A(t) = A(F(t))$. We will assume throughout that λ is larger than some λ_0 in order to ensure that $A(\Omega_\lambda) \ge 2A(\Omega) = 2$. We continue to assume that $A(\Omega) = 1$. In the event that $A(\Omega) \ne 1$, all areas may be scaled by $1/A(\Omega)$ to recover the result. We will apply the coarea formula directly to $G(z)$. We take $\eta = 0.01$ in (2.10)-(2.13) and begin with Case 1. Set $s_0 = \inf\{t > 0 : A(t) \ge 1.01\}$ and $T_0 = \sup\{t : A(t) \le 1.02\}$. Inserting $p = 2$ and $\eta = 0.01$ in Lemma 5.1, we obtain

Lemma 10.1. *For $\lambda \ge \lambda_0$, if T_0 is such that $A(T_0) = 1.02$, then*

$$\iint_{F(T_0)} |DG|^2 dx\, dy = \frac{4\pi\, T_0^2}{\log 1.02} \left(1 + D_1\alpha^2\right), \tag{10.2}$$

where D_1 depends only on κ.

We now proceed as in Section 5. Applying the usual isoperimetric inequality over the interval $T_0 < t < \lambda$, we obtain

$$(\lambda - T_0)^2 \le \frac{1}{4\pi} \log \frac{A(\lambda)}{A(T_0)} \iint_{\Omega_\lambda \backslash F(T_0)} |DG|^2 dx\, dy\, .$$

Combining this with (10.2) via Hölder's inequality, we see that

$$\iint_{\Omega_\lambda} |DG|^2 dx\, dy \ge \frac{4\pi\lambda^2}{\log\left(A(\lambda)(1.02)^{-D_1\alpha^2/(1+D_1\alpha^2)}\right)} \tag{10.3}$$

Since $G(z) - \log|z|$ is harmonic at ∞, it follows that with $r = |z|$, $\partial G/\partial r = 1/r + o(1/r^2)$ as $r \to \infty$. By Green's Theorem, we have as $r \to \infty$,

$$\begin{aligned}
\iint_{\Omega_\lambda} |DG|^2 dx\, dy &= \lambda \int_{\partial\Omega_\lambda} \frac{\partial G}{\partial n}\, ds \\
&= \lambda \int_{|z|=r} \frac{\partial G}{\partial r}\, ds \\
&= \lambda\, 2\pi\, r \left(\frac{1}{r} + o(\frac{1}{r^2})\right) \to 2\pi\lambda\,.
\end{aligned} \tag{10.4}$$

It follows from (10.1) that for $z \in \partial\Omega_\lambda$, $|z| = \mathrm{Cap}(\Omega)e^\lambda(1 + o(1))$, so that

$$A(\lambda) = \pi\left(\mathrm{Cap}(\Omega)\, e^\lambda\right)^2 (1 + o(1)) \qquad \text{as } \lambda \to \infty\,.$$

This with (10.3) and (10.4), gives

$$\frac{2\pi}{\lambda} \geq \frac{4\pi}{\log\left(\pi\, (\mathrm{Cap}(\Omega)\, e^\lambda)^2\, (1 + o(1))\, (1.02)^{-D_1\alpha^2/(1+D_1\alpha^2)}\right)}\,.$$

Thus,

$$\mathrm{Cap}(\Omega) \geq (1.02)^{D_1\alpha^2/(2(1+D_1\alpha^2))}\sqrt{\frac{1}{\pi}}\,.$$

The inequality in (1.3) now follows in Case 1.

We now discuss Case 2. As in Section 6, we may assume that there is a $t_0 > 0$ such that (6.8)-(6.10) hold. Let $F_1 = F_1(T)$, $\Omega_1 = F_1(T) \cap F(t_0)$ as in Section 6 and let Γ_c be the condenser $\Gamma(\Omega_1, \mathbb{C}\backslash F_1)$. Since F_1 and Ω_1 are both level sets for $G(z)$, it follows that

$$\begin{aligned}
\mathrm{Cap}(\Gamma_c) &= \iint_{F_1\backslash\Omega_1} |Dv|^2 dx\, dy \\
&= \frac{1}{(T - t_0)^2} \iint_{F_1\backslash\Omega_1} |DG|^2 dx\, dy\,,
\end{aligned} \tag{10.5}$$

where

$$v(z) = \frac{G(z) - t_0}{T - t_0}\,.$$

Using Lemma 4.1 in (10.5) and choosing $0 < \kappa < \kappa_0$ for some small κ_0, we may show that

$$\text{(10.6)} \qquad \iint_{F(T)} |DG|^2 dx\, dy \geq \frac{2\pi\,(T - t_0)^2}{\log R/\rho}\,(1 + B\alpha^2)\,,$$

where B is an absolute constant. From (6.29) x), R/ρ depends only on η. As was done in Case 1, we apply the usual isoperimetric inequality on $T < t \leq \lambda$, and combine the result with (10.6) via Hölder's inequality to obtain

$$\text{(10.7)} \quad 4\pi\,(\lambda - t_0)^2 \leq \log(A(\lambda)\,(1.01)^{-B\alpha^2/(1+B\alpha^2)}) \iint_{\Omega_\lambda} |DG|^2 dx\, dy\,.$$

To estimate t_0, observe that $F(t_0)$ is a level set of $G(z)$, and $G(z)/t_0$ is harmonic in $F(t_0)\backslash\Omega$. Thus,

$$\operatorname{Cap}\left(\Gamma(\Omega, \mathbb{C}\backslash F(t_0))\right) = \frac{1}{t_0^2} \iint_{F(t_0)\backslash\Omega} |DG|^2 dx\, dy \geq \frac{4\pi}{\log A(t_0)}\,.$$

Using the inequality (6.8) and an argument similar to that in (10.4) we have

$$\text{(10.8)} \quad t_0^2 \leq \frac{1}{4\pi} \log(1 + \kappa\,\alpha^2) \iint_{F(t_0)} |DG|^2 dx\, dy = \frac{t_0}{2} \log(1 + \kappa\,\alpha^2)\,.$$

Clearly then, $t_0 \leq \kappa\alpha^2$. Using (10.4), the estimate on $A(\lambda)$ (see Case 1) and the bound on t_0, in (10.7), we have

$$4\pi\,(\lambda - \kappa\,\alpha^2)^2 \leq 2\pi\lambda\Big(\log(\pi\big(\operatorname{Cap}(\Omega)e^{\lambda}\big)^2(1 + o(1))\,(1.01)^{-B\alpha^2/(1+\kappa\alpha^2)})\Big)\,.$$

Simplifying the above,

$$\operatorname{Cap}(\Omega) \geq e^{-2\kappa\alpha^2}(1.01)^{B\alpha^2/(2(1+B\alpha^2))}\sqrt{\frac{1}{\pi}}\,.$$

Fixing κ such that $0 < \kappa \leq \kappa_0$, we obtain (1.3).

11. The constants K_p.

Let $\Gamma(\Omega, \Omega')$ be a condenser as in Section 1, and set $\chi = A(\mathbb{R}^2 \backslash \Omega') / A(\Omega)$. Let $B(0,R)$ and $B(0,\bar{R})$ be discs such that $A(B(0,R)) = A(\Omega)$ and $A(B(0,\bar{R})) = A(\mathbb{R}^2 \backslash \Omega')$. Let $\Gamma^* = \Gamma(\bar{B}(0,R), \mathbb{R}^2 \backslash B(0,\bar{R}))$ and set $d = (p-2)/(p-1)$. Then

$$\mathrm{Cap}_2(\Gamma^*) = 4\pi / \log \chi \,,$$

and, for $p \neq 2$,

$$\mathrm{Cap}_p(\Gamma^*) = \frac{2\pi^{p/2}\,|d|^{p-1} A(\Omega)^{(2-p)/2}}{|\xi^{d/2} - 1|^{p-1}} = \frac{2\pi\,|d|^{p-1}}{|\bar{R}^d - R^d|^{p-1}} \,.$$

In this section we will discuss how the constants $K_p = K_p(\chi)$ in (1.5) behave as χ varies. Note that we have taken $\chi = 4$ in Theorem 1. Although determining the dependence on χ involves only routine modifications of the proofs, this was avoided in the text since such consideration involves carrying along additional parameters and the introduction of numerous subcases. In what follows, $\hat{K}_p$ represents positive constants depending only on p. Our methods give the following:

i) $1 < p < 2$,

$$K_p = \begin{cases} \hat{K}_p\,(\chi - 1)^2, & 1 < \chi \leq 2\,, \\ \hat{K}_p \text{ (independent of } \chi), & \chi > 2\,, \end{cases}$$

ii) $p = 2$,

$$K_2 = \begin{cases} \hat{K}_2\,(\chi - 1)^2, & 1 < \chi \leq 2\,, \\ \hat{K}_2 / \log \chi\,, & \chi > 2\,, \end{cases}$$

iii) $p > 2$,

$$K_p = \begin{cases} \hat{K}_p\,(\chi - 1)^2, & 1 < \chi \leq 2\,, \\ \hat{K}_p / |\chi^{d/2} - 1|\,, & \chi > 2\,. \end{cases}$$

Acknowledgements. We wish to thank the referee for many helpful suggestions.

References.

[1] Andrievski, V., Hansen, W., Nadirashvili, N., Isoperimetric inequalities for capacities in the plane. *Math. Ann.* **292** (1992), 191-195.

[2] Federer, H., *Geometric Measure Theory.* Springer-Verlag, 1969.

[3] Federer, H. and Fleming, W., Normal and integral currents. *Ann. of Math.* **72** (1960), 458-520.

[4] Gehring, F., Inequalities for condensers, hyperbolic capacity, and extremal length. *Michigan Math. J.* **18** (1971), 1-20.

[5] Hall, R., A quantitative isoperimetric inequality in n dimensional space. *J. Reine angew. Math.* **428** (1992), 61-76.

[6] Hall, R., Hayman, W., Weitsman, A., On capacity and asymmetry. *J. d'Analyse Math.* **56** (1991), 87-123.

[7] Hansen, W., Nadirashvili, N., *Isoperimetric inequalities for capacities, Harmonic Analysis and Discrete Potential Theory.* Editor M. Picardello, Plenum Press, 1992, 193-206.

[8] Hansen, W., Nadirashvili, N., Isoperimetric inequalities in Potential Theory. *Potential Analysis* **3** (1994), 1-14.

[9] Hayman, W., A problem in Fourier series arising from an isoperimetric inequality. *Current Problems in Analysis and Physics* (Taormina 1992), 119-125, Univ. Roma, 1993.

[10] Heinonen, J., Kilpeläinen, T., Martio, O., *Nonlinear Potential Theory of Degenerate Elliptic Equations.* Oxford Science Publications, 1993.

[11] Lewis, J., Capacitary functions in convex rings. *Arch. Rational Mech. Anal.* **66** (1977), 201-224.

[12] Lewis, J., Regularity of derivatives of solutions to certain degenerate elliptic equations. *Indiana Univ. Math. J.* **32** (1983), 849-858.

[13] Manfredi, J., p-harmonic functions in the plane. *Proc. Amer. Math. Soc.* **103** (1988), 473-479.

[14] Maz'ya, V., Classes of domains and imbedding theorems for function spaces. *Dokl. Akad. Nauk SSR* **133** (1960), 527-530.

[15] Osserman, R., Bonnesen-style isoperimetric inequalities. *Amer. Math. Monthly* **86** (1972), 1-29.

[16] Pólya, G., Szegö, G., *Isoperimetric inequalities in Mathematical Physics.* Annals of Math. Studies **27**, Princeton Univ. Press, 1951.

[17] Talenti, G., Best constant in Sobolev inequality. *Ann. Mat. Pura. Appl.* **110** (1976), 353-372.

Recibido: 6 de octubre de 1.994
Revisado: 15 de agosto de 1.995

Tilak Bhattacharya
Indian Statistical Institute
7, S.J.S. Sansanwal Marg
New Dehli 110016
INDIA
tlk@isid.ernet.in

and

Allen Weitsman
Department of Mathematics
Purdue University
W. Lafayette, IN 47907
U.S.A.
weits@math.purdue.edu

Revista Matemática Iberoamericana
Vol. 12, N.º 3, 1996

An endpoint estimate for some maximal operators

Daniel M. Oberlin

Suppose μ is a finite positive Borel measure on $\mathbb{R}^n$. It is proved in [DR] that if the Fourier transform of μ satisfies a decay estimate

$$|\hat{\mu}(\xi)| \leq C|\xi|^{-\alpha} \tag{1}$$

for some $\alpha > 0$, then the maximal operator

$$Mf(x) = \sup_{k \in \mathbb{Z}} \int_{\mathbb{R}^n} |f(x - 2^k y)| \, d\mu(y) \tag{2}$$

is bounded on $L^p(\mathbb{R}^n)$ for $1 < p < \infty$. On the other hand, Theorem 4 in [C2] states that if μ is the Lebesgue measure σ_{n-1} on the unit sphere Σ_{n-1} in $\mathbb{R}^n$, then (2) maps $H^1(\mathbb{R}^n)$ into $L^{1,\infty}(\mathbb{R}^n)$. The purpose of this paper is to adapt the method of [C2] to prove an H^1-$L^{1,\infty}$ result for (2) requiring, in the spirit of [DR], only a certain decay of $\hat{\mu}$.

Theorem. *Suppose μ is a finite positive Borel measure on $\mathbb{R}^n$ with support in $[-1,1]^n$. If*

$$|\hat{\mu}(\xi)| \leq C|\xi|^{-n/2},$$

then (2) *maps $H^1(\mathbb{R}^n)$ into $L^{1,\infty}(\mathbb{R}^n)$.*

As indicated, our proof follows the method of proof of Theorem 4 of [C2]. Our view is that the interest of this paper lies as much in a demonstration of the flexibility of that method (see [C2, Remark 7.2]) as in our result. Although many of the details differ, the main novelty here lies in the use of the auxiliary functions φ_N to handle the control (see (7)) of

$$\Big\| \Big(\sum_{\kappa(Q)=j-s} \lambda_Q \, a_Q \Big) * \mu_j \Big\|_2 .$$

The proof in [C2] used the curvature of the support of σ_{n-1} in the analogous estimate. Our argument proceeds, albeit in the same spirit, with no knowledge of μ aside from the decay of $\hat{\mu}$. But we pay by requiring a higher rate of decay -$\hat{\sigma}_{n-1}(\xi)$ decays, as is well-known, like $|\xi|^{(1-n)/2}$. Still, there exist singular measures on $\mathbb{R}^n$ satisfying our hypothesis. (This was proved in [I-M] for $n = 1$ - see Lemma 1 [K, p. 165] for the extension from Fourier coefficients to Fourier transform. To get a singular measure μ on $\mathbb{R}^n$ with $|\hat{\mu}(\xi)| = O(|\xi|^{-n/2})$, let ν be the measure from [I-M] translated to have support in [1,2] and define the measure μ on $\mathbb{R}^n$ by

$$\int_{\mathbb{R}^n} f \, d\mu = \int_1^2 \int_{\Sigma_{n-1}} f(ry) \, d\sigma_{n-1}(y) \, r^{(n-1)/2} d\nu(r) .$$

Then asymptotic estimates for Bessel functions such as those in [SW, Lemma 3.11] combine with the decay of $\hat{\nu}$ to give $|\hat{\mu}(\xi)| = O(|\xi|^{-n/2})$.) It may be that our $n/2$ can be replaced by smaller $\alpha > 0$, thus yielding a more satisfying endpoint analog of the result of [DR]. The referee has pointed out that the paper [S] contains a point of similarity to the proof of our theorem (in its use of the Fourier transform for the L^2 estimate) and that ideas equivalent to some of those in [DR] are present in [C1]. We begin with two lemmas.

Lemma 1. *For any $\alpha > 0$ and any finite collection of dyadic cubes $Q \subseteq \mathbb{R}^n$ and associated positive scalars λ_Q, there exists a collection $\mathcal{S}$ of pairwise disjoint dyadic cubes S such that*

a) $\sum_{Q \subseteq S} \lambda_Q \le 2^n \alpha \, |S| \, , \text{ if } S \in \mathcal{S} \, ,$

b) $\sum |S| \le \alpha^{-1} \sum \lambda_Q \, ,$

c) $\Big\| \sum_{\substack{Q \text{ not contained} \\ \text{in any } S}} \lambda_Q \, |Q|^{-1} \chi_Q \Big\|_\infty \le \alpha \, .$

PROOF. In the proof of Lemma 4.1 of [C2], simply replace 8 by 2^n and interpret dyadic in the n-dimensional Euclidean sense (instead of the parabolic sense in $\mathbb{R}^2$).

NOTATION. If Q is a dyadic cube in $\mathbb{R}^n$ with side-length 2^j, write $\sigma(Q)$ to stand for j. If $\sigma \in \mathbb{Z}$, let $\mathcal{R}_\sigma$ be the collection of dyadic cubes $Q \subseteq \mathbb{R}^n$ with $\sigma(Q) = \sigma$. Finally, if $Q \in \mathcal{R}_\sigma$, define $Q^* = Q + [-2^\sigma, 2^\sigma]^n$. Thus Q^* is the union of 3^n cubes in $\mathcal{R}_\sigma$.

Lemma 2. (*cf.* [C2, Lemma 5.1]) *Suppose given the following: some $\alpha > 0$, a collection $\mathcal{S}$ of pairwise disjoint dyadic cubes $S \subseteq \mathbb{R}^n$, a finite collection $\mathcal{C}$ of dyadic cubes $Q \subseteq \mathbb{R}^n$ such that each $Q \in \mathcal{C}$ is contained in some $S = S(Q) \in \mathcal{S}$, and for each $Q \in \mathcal{C}$ a positive number λ_Q. Then there exist a measurable $E \subseteq \mathbb{R}^n$ and a function $\kappa : \mathcal{C} \to \mathbb{Z}$ such that*

a) $|E| \leq 3^n(\alpha^{-1} \sum \lambda_Q + \sum |S|)$,

b) $Q + [-2^j, 2^j]^n \subseteq E$, *if* $j < \kappa(Q)$ *and* $Q \in \mathcal{C}$,

c) $\sigma(S(Q)) < \kappa(Q) \qquad (Q \in \mathcal{C})$,

d) *for* $\sigma \in \mathbb{Z}$ *any* $q \in \mathcal{R}_\sigma$, $\displaystyle\sum_{\substack{Q \subseteq q \\ \kappa(Q) \leq \sigma}} \lambda_Q \leq \alpha\, 2^{n(\sigma+1)}$.

PROOF. The proof is an adaptation of (and simpler than) that of Lemma 5.1 in [C2]. But we give the details for completeness and for the convenience of the reader.

Let $m = \min\{\sigma(Q)\}$. Find $\sigma_0 \in \mathbb{Z}$ such that

$$\sum \lambda_Q < \alpha\, 2^{n\sigma_0}, \qquad \sigma_0 > \max\{\sigma(Q)\}\,.$$

The proof is a stopping time argument on the descending parameter σ and proceeds by dividing $\mathcal{C}$ into disjoint subcollections $\mathcal{C}_1$ and $\mathcal{C}_2$. We begin with $\sigma = \sigma_0 - 1$ and define, for $q \in \mathcal{R}_\sigma$,

$$\Lambda_\sigma(q) = \sum_{Q \subseteq q} \lambda_Q\ .$$

Say that $q \in \mathcal{R}_\sigma$ is "selected at step σ" if

$$\Lambda_\sigma(q) > \alpha\, 2^{n\sigma}\,.$$

Put into $\mathcal{C}_1$ every Q such that $Q \subseteq q$ for some q selected at step σ, and for such Q define

$$\kappa(Q) = \max\{1 + \sigma, 1 + \sigma(S(Q))\}\,. \tag{3}$$

Next, put into $\mathcal{C}_2$ every $Q \in \mathcal{C} \sim \mathcal{C}_1$ such that $\sigma(Q) > \sigma$ - such a Q will actually satisfy $\sigma(Q) = \sigma + 1$ - and for such Q define

$$\kappa(Q) = 1 + \sigma(S(Q)) . \tag{4}$$

Note that (3) and (4) guarantee that (c) holds. Now replace σ by $\sigma - 1$ and repeat the process with

$$\Lambda_\sigma(q) = \sum_{\substack{Q \subseteq q \\ Q \notin \mathcal{C}_1}} \lambda_Q = \sum_{\substack{Q \subseteq q \\ Q \notin \mathcal{C}_1 \cup \mathcal{C}_2}} \lambda_Q , \qquad q \in \mathcal{R}_\sigma .$$

(The last equality holds because $Q \in \mathcal{C}_2$ at the beginning of step σ implies $\sigma(Q) \geq \sigma + 2$.) After the step $\sigma = m$ we will have $\mathcal{C} = \mathcal{C}_1 \cup \mathcal{C}_2$ and κ defined on all of $\mathcal{C}$. Next define

$$E_1 = \bigcup_{q \text{ selected}} q^*, \quad E_2 = \bigcup S^*, \quad E = E_1 \cup E_2 .$$

Then, since distinct selected q are disjoint,

$$|E_1| \leq 3^n \sum_{q \text{ selected}} 2^{n\sigma(q)} < \frac{3^n}{\alpha} \sum_{q \text{ selected}} \Lambda_{\sigma(q)}(q) \leq \frac{3^n}{\alpha} \sum \lambda_Q .$$

Now a) follows since $|S^*| = 3^n |S|$.

If $\kappa(Q) = 1 + \sigma(S(Q))$ and if $j < \kappa(Q)$, then

$$Q + [-2^j, 2^j]^n \subseteq S^* \subseteq E_2 .$$

If $\kappa(Q) \neq 1 + \sigma(S(Q))$, then $Q \subseteq q$ for some q selected at some step σ and $\kappa(Q) = 1 + \sigma(q)$. Thus if $j < \kappa(Q)$,

$$Q + [-2^j, 2^j]^n \subseteq E_1 .$$

So b) is verified.

Finally, if $q \in \mathcal{R}_\sigma$ for $\sigma \geq \sigma_0 - 1$, then d) is clear from the choice of σ_0. So suppose $\sigma < \sigma_0 - 1$ and $q \in \mathcal{R}_\sigma$. Now

$$\Lambda_\sigma(q) \leq \alpha\, 2^{n(\sigma+1)}$$

or else the $q_1 \in \mathcal{R}_{\sigma+1}$ that contains q would have been selected at stage $\sigma + 1$. Since $\kappa(Q) \leq \sigma$ implies that $Q \notin \mathcal{C}_1$ at the beginning of step σ,

$$\sum_{\substack{Q \subseteq q \\ \kappa(Q) \leq \sigma}} \lambda_Q \leq \Lambda_\sigma(q) ,$$

and so d) is proved.

Now suppose μ is a positive Borel probability measure supported on $[-1,1]^n$ and satisfying $|\hat{\mu}(\xi)| \le C|\xi|^{-n/2}$. Let $f \in H^1(\mathbb{R}^n)$ have the form of a finite sum

$$f = \sum \lambda_Q \, a_Q \ ,$$

where $\lambda_Q > 0$ and a_Q, supported in a cube Q, satisfies

$$\|a_Q\|_\infty \le |Q|^{-1}, \qquad \int_Q a_Q = 0 \, .$$

As in [C2], a device of Garnett and Jones involving auxiliary dyadic grids allows us to assume that each Q is dyadic. Fix $\alpha > 0$. It is enough to show that

$$|\{Mf > 2\alpha\}| \le \frac{C}{\alpha} \sum \lambda_Q \ , \tag{5}$$

where C depends only on μ and n.

Following [C2], let $\mathcal{S}$ be as in Lemma 1 and define

$$b = \sum_{S \in \mathcal{S}} \sum_{Q \subseteq S} \lambda_Q \, a_Q \ , \qquad g = f - b \, .$$

Then $\|g\|_\infty \le \alpha$ from Lemma 1.c) and so $|Mg| \le \alpha$ (because μ has mass 1). Thus (5) will follow from

$$|\{Mb > \alpha\}| \le \frac{C}{\alpha} \sum \lambda_Q \ .$$

Now, with $\mathcal{S}$ as above and with $\mathcal{C}$ the collection of Q's appearing in the definition of b, let κ and E be as in Lemma 2. Since $|E| \le C\alpha^{-1} \sum \lambda_Q$, it is enough to prove

$$\|Mb\|^2_{L^2(\mathbb{R}^n \sim E)} \le C\alpha \sum \lambda_Q \ . \tag{6}$$

Let μ_j be the dilate of μ defined by

$$\langle \varphi, \mu_j \rangle = \int_{\mathbb{R}^n} \varphi(2^j x) \, d\mu(x)$$

so that μ_j is supported in $[-2^j, 2^j]^n$ and

$$Mb(x) = \sup_{j\in\mathbb{Z}} |b * \mu_j(x)|\,.$$

If $Q \in \mathcal{C}$, then, by Lemma 2.b), $a_Q * \mu_j$ is supported in E unless $j \ge \kappa(Q)$. Thus if $x \notin E$,

$$\begin{aligned}
|Mb(x)|^2 &\le \sum_j |b * \mu_j(x)|^2 \\
&= \sum_j \Big| \Big(\sum_{\kappa(Q)\le j} \lambda_Q a_Q \Big) * \mu_j(x) \Big|^2 \\
&= \sum_j \Big| \sum_{s=0}^{\infty} \Big(\sum_{\kappa(Q)=j-s} \lambda_Q a_Q \Big) * \mu_j(x) \Big|^2 .
\end{aligned}$$

So, for $x \notin E$

$$|Mb(x)| \le \sum_{s=0}^{\infty} \Big(\sum_j \Big| \Big(\sum_{\kappa(Q)=j-s} \lambda_Q a_Q \Big) * \mu_j(x) \Big|^2 \Big)^{1/2} .$$

Now (6) will follow from

$$\Big\| \Big(\sum_j \Big| \Big(\sum_{\kappa(Q)=j-s} \lambda_Q a_Q \Big) * \mu_j \Big|^2 \Big)^{1/2} \Big\|_2^2 \le C\alpha(s+1)2^{-s} \sum \lambda_Q$$

and so from

$$\Big\| \Big(\sum_{\kappa(Q)=j-s} \lambda_Q a_Q \Big) * \mu_j \Big\|_2^2 \le C\alpha(s+1)2^{-s} \sum_{\kappa(Q)=j-s} \lambda_Q\,. \tag{7}$$

The proof of (7) requires another lemma.

Lemma 3. *For $N = 1, 2, \ldots$, there exist functions $\varphi_N \in L^1(\mathbb{R}^n)$ such that*

a) $|\hat{\varphi}_N(\xi)| \ge (1 + |\xi|)^{-n/2}/C$, *if* $|\xi| \le N - 1$,

b) $|\hat{\varphi}_N(\xi)| \le C\,|\xi|^{-n/2}$,

and if $L_N = \varphi_N * \tilde{\varphi}_N$ $(\tilde{\varphi}_N(x) = \varphi_N(-x))$, *then*

c) $\operatorname{supp}(L_N) \subseteq [-1,1]^n$,

d) $|L_N(x) - L_N(y)| \le C\,|x-y|/\min\{|x|,|y|\}$.

PROOF. We will construct L_N first and then φ_N. Define $h_N \in C(\mathbb{R}^n)$ by

$$\hat{h}_N(\xi) = \begin{cases} 1\,, & \text{if } |\xi| \le 1\,, \\ |\xi|^{-n}\,, & \text{if } 1 < |\xi| \le N\,, \\ 0\,, & \text{if } |\xi| > N\,. \end{cases}$$

Choose a radial function $\rho \in C_C^\infty(\mathbb{R}^n)$ such that

$$\int \rho = 1\,, \qquad \operatorname{supp}(\rho) \subseteq [-1,1]^n, \qquad \hat{\rho} \ge 0\,.$$

Now let $L_N = \rho\, h_N$. Clearly c) holds. It is easy to check that

$$\hat{L}_N(\xi) \ge (1+|\xi|)^{-n}/C \quad \text{if } |\xi| \le N-1\,,$$
$$0 \le \hat{L}_N(\xi) \le C|\xi|^{-n} \quad \text{if } \xi \in \mathbb{R}^n\,.$$

So if φ_N is the inverse Fourier transform of $(\hat{L}_N)^{1/2}$, then a) and b) hold. Since

$$|L_N(x) - L_N(y)| \le |\rho(x) - \rho(y)|\,|h_N(x)| + \rho(y)\,|h_N(x) - h_N(y)|\,,$$

d) will follow from

$$|h_N(x)| \le C\Big(\log^+\Big(\frac{1}{|x|}\Big) + 1\Big)\,, \tag{8}$$

and

$$\Big|\frac{\partial}{\partial|x|} h_N(x)\Big| \le \frac{C}{|x|}\,, \qquad |x| \le 1\,. \tag{9}$$

Now

$$h_N(x) = \int_0^1 \int_{\Sigma_{n-1}} e^{irx\cdot\omega}\, d\sigma_{n-1}(\omega)\, r^{n-1} dr + \int_1^N \int_{\Sigma_{n-1}} e^{irx\cdot\omega}\, d\sigma_{n-1}(\omega)\, \frac{dr}{r}\,,$$

with the important contribution coming from the second integral. For (8) just use the well-known estimate

$$\left| \int_{\Sigma_{n-1}} e^{irx\cdot\omega}\, d\sigma_{n-1}(\omega) \right| \le \frac{C}{(1+r|x|)^{(n-1)/2}}\,.$$

For (9) note that

$$\int_{\Sigma_{n-1}} e^{irx\cdot\omega}\, d\sigma_{n-1}(\omega) = \int_0^1 \cos(|x|rs)\omega(s)\, ds\,,$$

for some $\omega \in L^1([0,1])$. Now

$$\begin{aligned}\left| \frac{d}{dt} \int_1^N \int_0^1 \cos(trs)\,\omega(s)\, ds\, \frac{dr}{r} \right| &= \left| \int_0^1 \int_1^N \sin(trs)\, s\, dr\, \omega(s)\, ds \right| \\ &\le \int_0^1 \left| \int_s^{Ns} \sin(tu)\, du \right| \omega(s)\, ds \\ &\le \frac{C}{|t|}\,.\end{aligned}$$

Returning to (7) we have, because of our estimate on $\hat{\mu}$ combined with Lemma 3.a),

$$\begin{aligned}\left\| \Big(\sum_{\kappa(Q)=j-s} \lambda_Q a_Q \Big) * \mu_j \right\|_2^2 &= \int_{\mathbb{R}^n} \left| \Big(\sum_{\kappa(Q)=j-s} \lambda_Q a_Q \Big)^{\wedge}(\xi) \right|^2 |\hat{\mu}(2^j\xi)|^2 d\xi \\ &\le C \int_{\mathbb{R}^n} \left| \Big(\sum_{\kappa(Q)=j-s} \lambda_Q a_Q \Big)^{\wedge}(\xi) \right|^2 \\ &\qquad \cdot \liminf_N \left| \hat{\varphi}_N(2^j\xi) \right|^2 d\xi\,.\end{aligned}$$

Thus, letting $\varphi_{N,j}(x) = 2^{-nj}\varphi_N(2^{-j}x)$, (7) will follow from the estimates, uniform in N,

$$\text{(10)} \qquad \left\| \Big(\sum_{\kappa(Q)=j-s} \lambda_Q a_Q \Big) * \varphi_{N,j} \right\|_2^2 \le C\alpha(s+1) \sum_{\kappa(Q)=j-s} \lambda_Q\,.$$

So fix N, j, and s and write φ for φ_N, φ_j for $\varphi_{N,j}$. For $q \in \mathcal{R}_{j-s}$, let

$$A_q = \sum_{\substack{\kappa(Q)=j-s \\ Q\subseteq q}} \lambda_Q a_Q, \qquad \lambda_q = \sum_{\substack{\kappa(Q)=j-s \\ Q\subseteq q}} \lambda_Q \,.$$

Then

$$\begin{aligned}
\left\| \Big(\sum_{\kappa(Q)=j-s} \lambda_Q a_Q \Big) * \varphi_j \right\|_2^2 &\leq \sum_{q,q'\in\mathcal{R}_{j-s}} \left| \langle A_q * \varphi_j, A_{q'} * \varphi_j \rangle \right| \\
&\leq \sum_{q'} \sum_{q\subseteq (q')^*} + \sum_{q'} \sum_{q\cap (q')^* = \varnothing} \\
&= \mathrm{I} + \mathrm{II}\,.
\end{aligned}$$

The inequality

$$\|a_Q * \varphi_j\|_2 \leq C\, 2^{-nj/2}$$

follows easily from Lemma 3.b) and the well-known estimates

$$\left| \hat{a}_Q(\xi) \right| \leq C\, |\xi| \operatorname{diam}(Q)\,,$$

$$\|a_Q\|_2^2 \leq \frac{C}{|Q|}\,.$$

This leads, via Lemma 2.d), to

$$\begin{aligned}
\mathrm{I} &\leq C\, 2^{-nj} \sum_{q'} \lambda_{q'} \sum_{q\subseteq (q')^*} \lambda_q \\
&\leq C\, 2^{-nj} \sum_{q'} \lambda_{q'} \sum_{\substack{Q\subseteq (q')^* \\ \kappa(Q)=j-s}} \lambda_Q \\
&\leq C\, 2^{-nj} \sum_{q'} \lambda_{q'}\, \alpha\, 2^{n(j-s+1)} \\
&= C\alpha\, 2^{n(1-s)} \sum_{\kappa(Q)=j-s} \lambda_Q \,.
\end{aligned} \tag{12}$$

To estimate II, begin by fixing q, q' ($\in \mathcal{R}_{j-s}$) with $q \cap (q')^* = \varnothing$. We write

$$\langle A_q * \varphi_j, A_{q'} * \varphi_j \rangle = \int A_q(x) A_{q'} * L_j(x)\, dx\,, \tag{13}$$

where $L_j(x) = \varphi_j * \tilde{\varphi}_j(x) = 2^{-nj} L(2^{-j}x)$ and so, by Lemma 3.d),

$$|L_j(x) - L_j(y)| \leq C\, 2^{-nj} |x-y| / \min\{|x|, |y|\}\,.$$

Now if $\kappa(Q) = j - s$, $Q \subseteq q'$, $x \in q$, and $y_0 \in Q$, then

$$a_Q * L_j(x) = \int a_Q(y) \big(L_j(x-y) - L_j(x-y_0) \big)\, dy\,.$$

Thus

$$|a_Q * L_j(x)| \leq \frac{C\, 2^{-nj} \operatorname{diam}(Q)}{d(x,Q)} \leq \frac{C\, 2^{-nj+\sigma(Q)}}{d(x,Q)} \leq \frac{C\, 2^{-(n-1)j-s}}{d(x,Q)}\,,$$

since $\sigma(Q) \leq \sigma(S(Q)) < \kappa(Q) = j-s$ by Lemma 2. Also, if $a_Q * L_j(x) \neq 0$, then $d(x,Q) \leq C\, 2^j$ (since L_j is supported in $[-2^j, 2^j]^n$). Thus

$$|a_Q * L_j(x)| \leq \frac{C\, 2^{-s}}{d(x,Q)^n}\,.$$

Now suppose $x \in q$. If $Q \subseteq q'$ and $\kappa(Q) = j-s$, then $\sigma(S(Q)) < \kappa(Q) = j - s = \sigma(q')$. Since $S(Q) \cap q' \neq \varnothing$, $S(Q) \subseteq q'$. Because $q \cap (q')^* = \varnothing$, we must have $d(x, S(Q)) \geq 2^{j-s}$. Coupled with $d(x,S(Q)) \leq d(x,Q) \leq C\, 2^j$ if $a_Q * L_j(x) \neq 0$, we estimate, for fixed $q \in \mathcal{R}_{s-j}$ and $x \in q$,

$$\begin{aligned}
\sum_{(q')^* \cap q = \varnothing} |A_{q'} * L_j(x)| &\leq \sum_{(q')^* \cap q = \varnothing} \; \sum_{\substack{Q \subseteq q', \kappa(Q) = j-s \\ 2^{j-s} \leq d(x,S(Q)) \leq C2^j}} \lambda_Q |a_Q * L_j(x)| \\
&\leq C \sum_{(q')^* \cap q = \varnothing} \; \sum_{\substack{Q \subseteq q', \kappa(Q) = j-s \\ 2^{j-s} \leq d(x,S(Q)) \leq C2^j}} \lambda_Q \frac{2^{-s}}{d(x,Q)^n} \\
&\leq C 2^{-s} \sum_{2^{j-s} \leq d(x,S) \leq C2^j} \frac{1}{d(x,S)^n} \sum_{\substack{Q \subseteq S \\ \kappa(Q) = j-s}} \lambda_Q\,.
\end{aligned}$$

By Lemma 1.a) this last term is dominated by

$$C\alpha\, 2^{-s} \sum_{2^{j-s}\le d(x,S)\le C2^j} \frac{|S|}{d(x,S)^n} \le C\alpha\, 2^{-s} \int_{2^{j-s}}^{C2^j} \frac{dr}{r} \le C\alpha\, 2^{-s}(s+1)\,.$$

That is, if $x \in q$, then

$$\sum_{(q')^*\cap q=\varnothing} |A_{q'} * L_j(x)| \le C\alpha\, 2^{-s}(s+1))\,.$$

Thus, from (13),

$$\begin{aligned} \mathrm{II} &\le \sum_q \int |A_q(x)| \sum_{(q')^*\cap q=\varnothing} |A_{q'} * L_j(x)|\, dx \\ &\le C\alpha\, 2^{-s}(s+1) \sum_q \lambda_q = C\alpha\, 2^{-s}(s+1) \sum_{\kappa(Q)=j-s} \lambda_Q\,. \end{aligned}$$

With (11) and (12) this gives (10) and completes the proof of our theorem.

Acknowledgements. The author thanks the referee for his comments on a previous version of this paper.

References.

[C1] Christ, M., Hilbert transforms along curves, I: Nilpotent groups. *Ann. of Math.* **122** (1985), 575-596.

[C2] Christ, M., Weak type $(1,1)$ bounds for rough operators. *Ann. of Math.* **128** (1988), 19-42.

[DR] Duoandikoetxea, J. and Rubio de Francia, J. L., Maximal and singular integral operators via Fourier transform estimates. *Invent. Math.* **84** (1986), 541-561.

[I-M] Ivašev-Musatov, O.S., The coefficients of trigonometric null series. *Izvestia* **21** (1957), 559-578.

[K] Kahane, J. P., *Some Random Series of Functions.* Heath, 1968.

[S] Seeger, Andreas, Singular integral operators with rough convolution kernels. *J. Amer. Math. Soc.* **9** (1996), 95-105.

[SW] Stein, E. M. and Weiss, G., *Introduction to Fourier Analysis on Euclidean Spaces.* Princeton University Press, 1971.

Recibido: 22 de febrero de 1.995
Revisado: 31 de julio de 1.995

Daniel M. Oberlin
Department of Mathematics
Florida State University
Tallahassee, FL 32306-3027, U.S.A.
oberlin@math.fsu.edu

Revista Matemática Iberoamericana
Vol. 12, N.º 3, 1996

Transformée en paquets d'ondelettes des signaux stationnaires: comportement asymptotique des densités spectrales

Loïc Hervé

Résumé. On considère la transformée en paquets d'ondelettes associée à un filtre polynomial QMF. Soit $X = \{X_n\}_{n\in\mathbb{Z}}$ un signal aléatoire stationnaire à densité spectrale f continue. On démontre que les 2^n signaux, générés à partir de X après n itérations de la transformée, "convergent" vers des bruits blancs quand $n \to +\infty$. Si f est hölderienne, la vitesse de convergence est exponentielle.

Abstract. We consider quadrature mirror filters, and the associated wavelet packet transform. Let $X = \{X_n\}_{n\in\mathbb{Z}}$ be a stationary signal which has a continuous spectral density, f. We prove that the 2^n signals, obtained from X by n iterations of the transform, "converge" to white noises when $n \to +\infty$. If f is holderian, the convergence rate is exponential.

1. Enoncé des résultats.

On désigne par $(E, \|\cdot\|_\infty)$ l'espace des fonctions continues 1-périodiques muni de la norme uniforme, et pour $\eta \in]0,1]$, on note $(E^\eta, \|\cdot\|_\eta)$ le sous-espace des fonctions uniformément η-hölderiennes, avec

$$\|f\|_\eta = \|f\|_\infty + m_\eta(f)\,,$$

où

$$m_\eta(f) = \sup\left\{\frac{|f(y)-f(x)|}{|y-x|^\eta},\ x \neq y\right\}.$$

Soit $H_0(\lambda) = \sum_{k=0}^N h_k^0 e^{2i\pi k\lambda}$ un polynôme trigonométrique tel que $H_0(0) = 1$ et

$$\text{(1)} \qquad \left|H_0(\frac{\lambda}{2})\right|^2 + \left|H_0(\frac{\lambda}{2}+\frac{1}{2})\right|^2 = 1\,, \quad \text{pour tout } \lambda \in [0,1]\,.$$

On suppose que les h_k^0 sont des nombres réels, et on définit $h_k^1 = (-1)^{k+1}h_{-1-k}^0$ pour $k = -N-1, \ldots, -1$, et

$$H_1(\lambda) = \sum_{k=-N-1}^{-1} h_k^1 e^{2i\pi k\lambda} = e^{-2i\pi\lambda}\overline{H_0(\lambda+\frac{1}{2})}\,.$$

Notons que $H_0(0) = H_0(1) = H_1(1/2) = 1$, $H_1(0) = H_1(1) = H_0(1/2) = 0$, et que H_1 vérifie également la relation (1). Un tel couple (H_0, H_1) est appelé QMF (quadrature mirror filters) [3], [8]. La méthode d'analyse spectrale des signaux aléatoires, développée dans [2], consiste à itérer, à partir d'un processus aléatoire initial $X = (X_n)_{n\in\mathbb{Z}}$, les deux opérations de filtrage T_0 et T_1 définies par

$$(T_jX)_n = \sqrt{2}\sum_{k\in\mathbb{Z}} h_k^j X_{2n-k}, \qquad j = 0,1, \quad n \in \mathbb{Z}\,,$$

où par convention $h_k^0 = 0$ si $k \notin [0,N]$, et $h_k^1 = 0$ si $k \notin [-N-1,-1]$.

Cet algorithme, qui en d'autres termes effectue la transformée en paquets d'ondelettes de X [1], fournit une famille arborescente de signaux stationnaires: après n itérations, on dispose des 2^n processus $T_{\omega_n}\cdots T_{\omega_1}X$, où les ω_i décrivent $\{0,1\}$.

On note $\Omega = \{0,1\}^{\mathbb{N}^*}$, $u_0 = |H_0|^2$, $u_1 = |H_1|^2$, et P_0, P_1 les opérateurs de transition définis sur E par

$$\text{(2)} \qquad P_jf(\lambda) = u_j(\frac{\lambda}{2})\,f(\frac{\lambda}{2}) + u_j(\frac{\lambda}{2}+\frac{1}{2})\,f(\frac{\lambda}{2}+\frac{1}{2})\,,$$

où $\lambda \in [0,1]$, $j = 0,1$. On démontre dans [2] le résultat suivant

Théorème 1.1.

i) *Si* $X = (X_n)_{n\in\mathbb{Z}}$ *est un processus stationnaire du second ordre, les processus* T_0X *et* T_1X *sont également stationnaires. Si* X *a une densité spectrale* f, *alors* P_0f *et* P_1f *sont les densités spectrales respectivement de* T_0X *et* T_1X, *et plus généralement chaque processus* $T_{\omega_n}\cdots T_{\omega_1}X$ *admet une densité spectrale égale à* $P_{\omega_n}\cdots P_{\omega_1}f$.

ii) *Soit* Q_0 *le nombre de zéros de* H_0. *On suppose que, pour tout* $p \in \{1,\ldots,Q_0\}$, *pour tout* $k \in \{1,\ldots,2^p-2\}$, *il existe* $\ell \in \{0,\ldots,p\}$, *tel que*

$$H_0\left(\frac{k\,2^\ell}{2^p-1}+\frac{1}{2}\right) \neq 0\,. \tag{3}$$

Si $f \in E$, *alors, pour presque tout* $\omega = (\omega_1,\omega_2,\ldots,\omega_n,\ldots) \in \Omega$ *(au sens de la probabilité produit équiprobable sur* Ω*), la suite de processus* $(T_{\omega_n}\cdots T_{\omega_1}X)_{n\geq 1}$ *"converge" vers un bruit blanc, c'est-à-dire* $(P_{\omega_n}\cdots P_{\omega_1}f)_{n\geq 1}$ *converge vers une constante* $c(f,\omega)$ *quand* $n \to +\infty$.

Le point i) résulte d'un calcul élémentaire, et l'assertion ii) du Théorème de Ionescu-Tulcea et Marinescu [6] et de la Loi des grands nombres. Signalons que la condition (3) est exactement l'hypothèse sur les cycles périodiques invariants donnée dans [2]. Dans ce travail, sous une hypothèse du même type mais un peu plus forte que (3), nous nous proposons, d'une part de généraliser la propriété de convergence du ii) à tout $\omega \in \Omega$, et d'autre part de prouver que la vitesse de convergence est exponentielle quand f est hölderienne.

Théorème 1.2. *Soit* Q *le nombre de zéros de* H_0H_1. *On suppose que, pour tout* $p \in \{1,\ldots,Q\}$, *pour tout* $k \in \{1,\ldots,2^p-2\}$, *il existe* $\ell \in \{0,\ldots,p\}$, *tel que*

$$H_0\left(\frac{k\,2^\ell}{2^p-1}\right) H_1\left(\frac{k\,2^\ell}{2^p-1}\right) \neq 0\,. \tag{4}$$

Soit X *un processus stationnaire du second ordre admettant une densité spectrale* f *continue, et soit* $\omega = (\omega_1,\ldots,\omega_n,\ldots) \in \Omega$ *quelconque. Alors la mesure spectrale* $P_{\omega_n}\cdots P_{\omega_1}f$ *du processus* $T_{\omega_n}\cdots T_{\omega_1}X$ *converge uniformément vers une constante* $c(f,\omega)$ *quand* $n \to +\infty$, *c'est-à-dire,*

$$\lim_{n\to+\infty} \|P_{\omega_n}\cdots P_{\omega_1}f - c(f,\omega)\|_\infty = 0\,. \tag{5}$$

Théorème 1.3. *On conserve les hypothèses et notations du théorème précédent. Si $f \in E^\eta$, alors il existe deux constantes $D > 0$ et $\rho \in [0, 1[$ (indépendantes de f) telles que l'on ait, pour tout $\omega = (\omega_1, \ldots, \omega_n, \ldots) \in \Omega$,*

$$\|P_{\omega_n} \cdots P_{\omega_1} f - c(f, \omega)\|_\eta \leq D\rho^n \|f\|_\eta \,. \tag{6}$$

Remarques. a) La propriété (6) montre que la suite d'opérateurs $(P_{\omega_n} \cdots P_{\omega_1})_{n \geq 1}$ converge en norme dans E^η avec une vitesse exponentielle. L'étude du comportement asymptotique des spectres discrets dans l'arbre de filtrage, que nous n'abordons pas ici, a été traitée dans [2]. Par ailleurs une étude plus précise des itérées de T_0 a été faite dans [5] dans le cadre des filtres polynomiaux non nécessairement QMF.

b) L'hypothèse (3) est une condition nécessaire et suffisante pour que H_0 engendre une analyse multirésolution, et elle assure que la suite $\{P_0^n f,\ n \geq 1\}$ converge uniformément vers $f(0)$ pour tout $f \in E$, voir [2]. Pour $\omega \in \Omega$, on note μ_ω la mesure de probabilité sur le tore définie par

$$\int f \, d\mu_\omega = c(f, \omega) \,, \qquad f \in E \,.$$

Si $\omega = (0, 0, \ldots)$, le résultat ci-dessus montre que $\mu_\omega = \delta_0$ (masse de Dirac en 0). Plus généralement, si ω est de la forme $\omega = (\omega_1, \ldots, \omega_r, 0, 0, \ldots)$, on a $\mu_\omega = \sum_{k=0}^{2^r - 1} u_{\omega_r}(k/2) \cdots u_{\omega_1}(k/2^r) \, \delta_{(k/(2^r))}$.

Cependant ce type de propriété ne s'étend pas à tous les éléments de Ω. Plus précisément, soient ν la mesure produit équiprobable sur Ω, $\mathcal{D}$ l'ensemble des points dyadiques de $[0, 1[$, et soit enfin A le sous-ensemble de Ω formé des ω tels que μ_w soit de la forme

$$\mu_\omega = \sum_{a \in \mathcal{D}} \alpha_a(\omega) \, \delta_a \ ,$$

où les $\alpha_a(\omega)$ sont des réels positifs tels que $\sum_{a \in \mathcal{D}} \alpha_a(\omega) = 1$. Alors $\nu(A) = 0$. En effet, on établit aisément par récurrence que

$$\int_0^1 f(\lambda) \, d\lambda = 2^{-n} \sum_{\omega_1, \ldots, \omega_n \in \{0,1\}} \int_0^1 P_{\omega_n} \cdots P_{\omega_1} f(\lambda) \, d\lambda \,,$$

pour tout $f \in E$ et pour tout $n \geq 1$. D'où, grâce au théorème de convergence dominée sur (Ω, ν),

$$\int_0^1 f(\lambda) \, d\lambda = \int_\Omega c(f, \omega) \, d\nu(\omega) \,.$$

Donc, pour tout borélien B, on a $m(B) = \int_\Omega \mu_\omega(B)\, d\nu(\omega)$, où m est la mesure de Lebesgue sur $[0,1]$. Pour $B = \mathcal{D}$, on obtient $0 = m(\mathcal{D}) \geq \nu(A)$. Donc $\nu(A) = 0$.

Signalons également que, si $\omega = (1,1,\ldots)$, alors μ_ω est une mesure continue: $\mu_\omega(\{x\}) = 0$ pour tout $x \in [0,1[$. En effet supposons que $\mu_\omega(\{x\}) > 0$. On sait que μ_ω est invariante sous l'action de P_1 et Δ, où Δ est la transformation définie par $\Delta\lambda = 2\lambda$ (modulo1), voir [7]. On en déduit aisément que x est nécessairement un point fixe pour une certaine puissance Δ^p de Δ, et que $u_1(\Delta^k x) = 1$ pour $k = 1,\ldots,p$, ce qui contredit l'hypothèse (4). Donc $\mu_\omega(\{x\}) = 0$. De même, si $\omega = (0,1,0,1,\ldots)$, ou plus généralement si ω est cyclique, on peut montrer que μ_ω est une mesure continue.

c) Pour tout $n \geq 1$, l'application A_n définie par

$$f \longmapsto (P_{\omega_n} \cdots P_{\omega_1} f)_{\omega_1,\ldots,\omega_n \in \{0,1\}}$$

est injective de E dans E^{2^n}. Il suffit de le vérifier pour $n = 1$. θ_r, pour λ fixé, on a ${}^t[P_0 f(2\lambda), P_1 f(2\lambda)] = \mathcal{A}(\lambda)\ {}^t[f(\lambda), f(\lambda + 1/2)]$, où $\mathcal{A}(\lambda)$ est une matrice carrée d'ordre 2 qui s'exprime aisément à l'aide de $u_0(\lambda)$ et $u_1(\lambda)$, et dont le déterminant est $D(\lambda) = (u_0(\lambda))^2 - (u_1(\lambda))^2$. L'injectivité de A_1 résulte du fait que D a un nombre fini de zéros.

Par conséquent si deux processus X et Y admettent des densités spectrales f_X et f_Y distinctes, alors $A_n f_X \neq A_n f_Y$ pour tout $n \geq 1$, et en ce sens, la transformée en ondelettes des signaux stationnaires fournit un procédé d'analyse spectrale. Cependant, la propriété d'injectivité de A_n peut se "dégrader" quand $n \to +\infty$. Plus précisément, considérons l'exemple du filtre de Haar, $H_0(\lambda) = (1 + e^{2i\pi\lambda})/2$. Un calcul simple montre que, pour $f(\lambda) = \sin 2\pi\lambda$, on a $P_0 f = f/2$ et $P_1 f = -f/2$, de sorte que $c\,(f,\omega) = 0$ pour tout $\omega \in \Omega$. En d'autres termes, l'application $f \mapsto c\,(f,\cdot)$ associée au filtre de Haar n'est pas injective. Nous ne sommes pas parvenus à étudier, pour H_0 quelconque, l'injectivité de $f \mapsto c\,(f,\cdot)$.

La suite de ce papier est consacrée à la démonstration des théorèmes 1.2 et 1.3 qui repose, d'une part sur la positivité des opérateurs P_0, P_1 et la notion de points périodiques [2], [4] (étude dans E), et d'autre part sur le Théorème de Ionescu-Tulcea et Marinescu et des arguments de compacité (étude dans E^η).

2. Démonstration du Théorème 1.2.

Dans ce paragraphe, nous nous proposons de démontrer le Théorème 1.2. Pour simplifier, on notera, pour tout $n \in \mathbb{N}^*$ et tout $\omega = (\omega_1, \dots, \omega_n, \dots) \in \Omega$,

$$\Pi_n^\omega f(\lambda) = P_{\omega_n} \cdots P_{\omega_1} f(\lambda), \qquad f \in E,\, \lambda \in [0,1]\,.$$

Rappelons que $u_1(\lambda) = u_0(\lambda + 1/2)$. Les opérateurs P_0 et P_1 définis par (2) sont bornés, positifs sur E, et vérifient $P_0 1 = P_1 1 = 1$, où 1 est la fonction identiquement égale à 1. De même chaque opérateur Π_n^ω est positif, borné sur E, et vérifie

$$\Pi_n^\omega 1 = 1, \quad \|\Pi_n^\omega f\|_\infty \le \|f\|_\infty, \qquad \text{pour tout } f \in E\,.$$

On définit, pour $k \in \mathbb{N}^*$, $\mathcal{T}_k = \operatorname{vect}\{e^{-2i\pi(k+1)\lambda}, \dots, e^{2i\pi(k+1)\lambda}\}$. Notons que u_0 et u_1 appartiennent à $\mathcal{T}_N$. Pour de simples raisons de degré, il est clair que, si f est un polynôme trigonométrique, alors les fonctions $\Pi_n^\omega f$ appartiennent à $\mathcal{T}_N$ pour n assez grand. De même, on montre facilement que P_0, P_1, et donc Π_n^ω, opèrent sur $\mathcal{T}_N$.

Soient S_0 et S_1 les applications définies par

$$S_j \lambda = \frac{\lambda + j}{2}\,, \qquad \lambda \in [0,1]\,, j = 0, 1\,.$$

Pour $\lambda \in [0,1]$, $\omega \in \Omega$, et $n \in \mathbb{N}^*$, on note $T_{n,\lambda}^\omega$ l'ensemble des points $\sigma_n \cdots \sigma_1 \lambda$ tels que $\sigma_i \in \{S_0, S_1\}$ et $u_{\omega_n}(\sigma_1\lambda) \cdots u_{\omega_1}(\sigma_n \cdots \sigma_1 \lambda) > 0$. Compte-tenu des identités $u_j(\lambda/2) + u_j(\lambda/2 + 1/2) = 1$, l'ensemble $T_{n,\lambda}^\omega$ n'est jamais vide. On notera $[a]$ la partie entière d'un réel a, et θ le shift défini sur Ω par

$$\theta\omega = \theta(\omega_1, \dots, \omega_n, \dots) = (\omega_2, \dots, \omega_n, \dots)\,.$$

La démonstration du Théorème 1.2 utilise les deux lemmes techniques suivants:

Lemme 2.1. *Soit h une fonction de E à valeurs positives ou nulles, et soient $\omega \in \Omega$, $\lambda \in [0,1]$, $m, \ell \in \mathbb{N}^*$, et enfin $y = \tau_\ell \cdots \tau_1 \lambda \in T_{\ell,\lambda}^{\theta^m\omega}$. Si $\Pi_m^\omega h(y) > 0$, alors on a $\Pi_{\ell+m}^\omega h(\lambda) > 0$.*

Lemme 2.2. *On note $r = [\log_2(2N+1)] + 1 + 2Q$. Soit h une fonction de $\mathcal{T}_N$ à valeurs positives ou nulles, mais non identiquement nulle. Il existe $\delta > 0$ telle que*

$$P_{\mu_n} \cdots P_{\mu_1} P_1 P_{\omega_r} \cdots P_{\omega_1} h(\lambda) \geq \delta\,, \quad \textit{pour tous } \lambda \in [0,1]\,,\ n \geq 1\,,\ \omega_i, \mu_i \in \{0,1\}\,.$$

Preuve du Lemme 2.1. On montre aisément par récurrence que

$$\Pi_m^\omega h(\lambda) = \sum_{\sigma_1,\ldots,\sigma_m \in \{S_0,S_1\}} u_{\omega_m}(\sigma_1\lambda) \cdots u_{\omega_1}(\sigma_m \cdots \sigma_1\lambda)\, h(\sigma_m \cdots \sigma_1 \lambda)\,.$$

Comme $y \in T_{\ell,\lambda}^{\theta^m \omega}$, on a $A = u_{\omega_{m+\ell}}(\tau_1\lambda) \cdots u_{\omega_{m+1}}(\tau_\ell \cdots \tau_1 \lambda) > 0$. Le Lemme 2.1 se déduit alors de l'inégalité

$$\begin{aligned}
&\Pi_{m+\ell}^\omega h(\lambda)\\
&\quad \geq A \sum_{\sigma_{\ell+1},\cdots,\sigma_{\ell+m} \in \{S_0,S_1\}} u_{\omega_m}(\sigma_{\ell+1}y) \cdots u_{\omega_1}(\sigma_{\ell+m} \cdots \sigma_{\ell+1}y) h(\sigma_{\ell+m} \cdots \sigma_{\ell+1} y)\\
&\quad = A\ \Pi_m^\omega h(y)\,.
\end{aligned}$$

Avant de donner la preuve du Lemme 2.2, commençons par faire quelques rappels sur la notion de points périodiques [2], [4], et le lien avec la condition (4). Pour $p \in \mathbb{N}^*$, on dit que $\lambda \in [0,1]$ est un point p-périodique s'il existe p éléments $\sigma_1, \ldots, \sigma_p$ de $\{S_0, S_1\}$ tels que $\sigma_p \cdots \sigma_1 \lambda = \lambda$, et si p est le plus petit entier pour lequel on a une telle relation (de manière équivalente, si $\Delta^p \lambda = \lambda$ et $\Delta^k \lambda \neq \lambda$ pour $k = 1, \ldots, p-1$, où $\Delta x = 2x \bmod 1$). La famille $\{\sigma_1, \ldots, \sigma_p\}$ vérifiant la relation ci-dessus est unique, et l'on note

$$\mathcal{C}_\lambda = \{\sigma_k \cdots \sigma_1 \lambda :\ k = 1, \ldots, p\}\,.$$

Il est clair que les points périodiques d'ordre inférieur ou égal à un entier m, $m \in \mathbb{N}^*$, sont de la forme $k/(2^p - 1)$, où $p \in \{1, \ldots, m\}$ et $k \in \{0, 1, \ldots, 2^p - 1\}$. La condition (4) est équivalente à la suivante: pour tout $\lambda \in]0,1[$, p-périodique, $2 \leq p \leq Q$, il existe $y \in \mathcal{C}_\lambda$, tel que

$$H_0(y) H_1(y) \neq 0\,. \tag{7}$$

Pour $\lambda \in [0,1]$, on note $A_\lambda = \{\sigma_n \cdots \sigma_1 \lambda :\ n \geq 1,\ \sigma_1, \ldots, \sigma_n \in \{S_0, S_1\}\}$. Nous aurons besoin des propriétés suivantes démontrées dans [4]:

Soit $\lambda \in [0,1]$. Si λ est périodique, alors un (et uniquement un) des points $S_0\lambda$ et $S_1\lambda$ est périodique. Si λ n'est pas périodique, les points de A_λ sont distincts deux à deux et ne sont pas périodiques.

Preuve du Lemme 2.2. On note Z l'ensemble des zéros de $u_0 u_1$, et $|E|$ le cardinal d'un ensemble quelconque E. Rappelons que $|Z| = Q$. Les opérateurs P_i étant positifs et tels que $P_i 1 = 1$, il suffit de prouver qu'il existe $\delta > 0$ tel que l'on ait, pour tous $\lambda \in [0,1]$, $\omega_1, \ldots, \omega_r \in \{0,1\}$

$$P_1 P_{\omega_r} \cdots P_{\omega_1} h(\lambda) \geq \delta \,. \tag{8}$$

a) Soit $r_1 = [\log_2(2N+1)] + 1$. Si λ est non périodique et tel que $A_\lambda \cap Z = \varnothing$, alors $\Pi_m^\omega h(\lambda) > 0$, pour tous $m \geq r_1$, $\omega \in \Omega$. En effet, sinon on aurait $h(\sigma_m \cdots \sigma_1 \lambda) = 0$ pour chaque $\sigma_m, \ldots, \sigma_1 \in \{S_0, S_1\}$, ce qui est impossible car h admet au plus $2N+1$ racines.

b) Soit $r_2 = r_1 + Q$. Si λ est non périodique, alors $\Pi_m^\omega h(\lambda) > 0$, pour tous $m \geq r_2$, $\omega \in \Omega$. Pour prouver b), considérons les ensembles

$$A_\lambda^Q = \{\sigma_k \cdots \sigma_1 \lambda :\ 1 \leq k \leq Q,\ \sigma_i \in \{S_0, S_1\}\}\,,$$

$$F_\lambda^Q = \{\sigma_Q \cdots \sigma_1 \lambda :\ \sigma_i \in \{S_0, S_1\}\}\,.$$

On note $p = |A_\lambda^Q \cap Z|$. Soit $\omega' \in \Omega$ quelconque. On a $0 \leq p \leq Q$, et $|F_\lambda^Q - T_{Q,\lambda}^{\omega'}| \leq 2^{Q-1} + 2^{Q-2} + \cdots + 2^{Q-p} = 2^Q - 2^{Q-p}$. Pour prouver cette inégalité, on peut par exemple représenter l'ensemble A_λ^Q sous la forme d'un arbre dyadique de racine λ (admettant pour fils $\lambda/2$ et $\lambda/2 + 1/2$ ... etc ...), et remarquer que le nombre $|F_\lambda^Q - T_{Q,\lambda}^{\omega'}|$ est d'autant plus grand que les éléments de Z sont proches de λ dans l'arbre. Il en résulte que $|T_{Q,\lambda}^{\omega'}| \geq 2^{Q-p} > Q - p$. Les points de A_λ étant distincts deux à deux, on en déduit qu'il existe $y \in T_{Q,\lambda}^{\omega'}$ tel que $A_y \cap Z = \varnothing$. Rappelons que y est nécessairement non périodique. Soient maintenant $m \geq 1$ et $\omega \in \Omega$ quelconques: il existe $y_m \in T_{Q,\lambda}^{\theta^m \omega}$ non périodique tel que $A_{y_m} \cap Z = \varnothing$. Du a), il vient que $\Pi_m^\omega h(y_m) > 0$ pour tout $m \geq r_1$, d'où, d'après le Lemme 2.1, $\Pi_{m+Q}^\omega h(\lambda) > 0$, ce qui prouve b).

c) Soit $r = r_2 + Q$. Si λ est périodique, $\lambda \neq 0$, alors $\Pi_m^\omega h(\lambda) > 0$, pour tous $m \geq r$, $\omega \in \Omega$. Démontrons tout d'abord que, pour tout

$\omega' \in \Omega$, il existe un élément y de $T^{\omega'}_{Q,\lambda}$ non périodique: dans le cas contraire, en vertu des propriétés sur les points périodiques rappelées ci-dessus, l'ensemble $T^{\omega'}_{Q,\lambda}$ serait en fait réduit à un seul élément qui en outre appartiendrait à $\mathcal{C}_\lambda$, d'où $u_0(t)\, u_1(t) = 0$ pour tout $t \in \mathcal{C}_\lambda$, ce qui contredit l'hypothèse (7).

Soit $m \in \mathbb{N}^*$. Il existe donc $y \in T^{\theta^m\omega}_{Q,\lambda}$ non périodique, de sorte qu'on a, pour tout $m \geq r_2$, $\Pi^\omega_m h(y) > 0$ et donc d'après le Lemme 2.1, $\Pi^\omega_{m+Q} h(\lambda) > 0$. Le point c) est prouvé.

On a en particulier démontré que, pour tout $\omega_1, \ldots, \omega_r, \omega_{r+1} \in \{0,1\}$ et tout $\lambda \neq 0$, $P_{\omega_{r+1}} P_{\omega_r} \cdots P_{\omega_1} h(\lambda) > 0$. Remarquons que, si $\omega_{r+1} = \omega_r = \cdots = \omega_2 = 0$ et $\omega_1 = 1$, alors $\Pi^\omega_{r+1} h(0) = u_0(0) \cdots u_0(0)$ $u_1(1/2) h(1/2) = h(1/2)$, ce dernier terme pouvant être nul. Pour $\lambda = 0$, il est donc nécessaire d'avoir $\omega_{r+1} = 1$.

d) (*cas* $\lambda = 0$). On a $P_1 P_{\omega_r} \cdots P_{\omega_1} h(0) > 0$, pour tous $\omega_1, \ldots, \omega_r \in \{0,1\}$. En effet, comme $u_1(0) = 0$, on obtient

$$\begin{aligned} &P_1 P_{\omega_r} \cdots P_{\omega_1} h(0) \\ &\quad = \sum_{\sigma_2,\ldots,\sigma_{r+1} \in \{S_0, S_1\}} u_1(\frac{1}{2}) u_{\omega_r}(\sigma_2 \frac{1}{2}) \cdots u_{\omega_1}(\sigma_{r+1} \cdots \sigma_2 \frac{1}{2}) h(\sigma_{r+1} \cdots \sigma_2 \frac{1}{2}) \\ &\quad = P_{\omega_r} \cdots P_{\omega_1} h(\frac{1}{2}) \end{aligned}$$

ce dernier terme étant positif d'après ce qui précède.

Notons que r est bien indépendant de la fonction h. Nous pouvons maintenant prouver (8). Soient $\omega_1, \ldots, \omega_r \in \{0,1\}$. Il existe une constante $\delta_{(\omega_1,\ldots,\omega_r)} > 0$ ne dépendant que de $(\omega_1, \ldots, \omega_r)$ telle que $P_1 P_{\omega_r} \cdots P_{\omega_1} h \geq \delta_{(\omega_1,\ldots,\omega_r)}$. On en déduit (8) avec $\delta = \min\{\delta_{(\omega_1,\ldots,\omega_r)} : \omega_i \in \{0,1\}\}$.

DÉMONSTRATION DU THÉORÈME 1.2. Soit $\omega \in \Omega$.

1^{er} *cas*: il existe $k \in \mathbb{N}^*$ tel que $\omega_n = 0$ pour tout $n > k$. Alors $\Pi^\omega_n = P_0^{n-k} \Pi^\omega_k$. On déduit de l'étude des itérées de P_0 faite dans [2] que, pour tout $f \in E$, la suite $\{\Pi^\omega_n f : n \geq 1\}$ converge uniformément vers la constante $\Pi^\omega_k f(0)$.

2^{ieme} *cas*: il existe une suite strictement croissante $\{\phi(n)\}_{n \geq 1}$ d'entiers positifs tels que $\omega_{\phi(n)} = 1$. Commençons par supposer que

• $f \in \mathcal{T}_N$: on peut choisir les $\phi(n)$ tels que $\phi(n+1) - \phi(n) > r+1$, où r est l'entier défini dans le Lemme 2.2. Soit $\psi(n) = \phi(n) - r - 1$.

La famille $\{\Pi^{\omega}_{\psi(n)} f\ :\ n \geq 1\}$ est bornée dans l'espace $\mathcal{T}_N$ qui est de dimension finie. On peut donc en extraire une sous-suite $\{\Pi^{\omega}_{\tau(n)} f\ :\ n \geq 1\}$ convergeant uniformément vers une fonction g de $\mathcal{T}_N$.

Nous allons démontrer que g est identiquement égale à

$$c = \inf_{\lambda \in [0,1]} [g(\lambda)] .$$

A cet effet procédons par l'absurde et supposons qu'il existe $\lambda \in [0, 1]$ tel que $g(\lambda) > c$. On a $\Pi^{\omega}_{\tau(n+1)} = R_n \Pi^{\omega}_{\tau(n)}$ où

$$R_n = P_{\omega_{\tau(n+1)}} \cdots P_{\omega_{\tau(n)+r+2}} P_{\omega_{\tau(n)+r+1}} P_{\omega_{\tau(n)+r}} \cdots P_{\omega_{\tau(n)+1}} .$$

Rappelons que par construction $\omega_{\tau(n)+r+1} = 1$. Le Lemme 2.2 appliqué avec $h = g - c$ assure l'existence d'une constante $\delta > 0$ telle que l'on ait $R_n(g - c) \geq \delta$, ou encore $R_n g \geq c + \delta$, pour tout $n \geq 1$. Or on a $\Pi^{\omega}_{\tau(n+1)} f - R_n g = R_n(\Pi^{\omega}_{\tau(n)} f - g)$, d'où $\|\Pi^{\omega}_{\tau(n+1)} f - R_n g\|_{\infty} \leq \|\Pi^{\omega}_{\tau(n)} f - g\|_{\infty}$. La suite $\{R_n g\ :\ n \geq 1\}$ converge donc uniformément vers g. Il en résulte que $\lim_{n \to +\infty} (\inf_{\lambda \in [0,1]} R_n g(\lambda)) = c$, ce qui est impossible d'après l'inégalité ci-dessus. Donc $g = c$. On conclut que $\{\Pi^{\omega}_n f\ :\ n \geq 1\}$ converge uniformément vers c en remarquant que, pour tout $m \geq \tau(n)$,

$$\|\Pi^{\omega}_m f - c\|_{\infty} = \|P_{\omega_m} \cdots P_{\omega_{\tau(n)+1}} (\Pi^{\omega}_{\tau(n)} f - c)\|_{\infty} \leq \|\Pi^{\omega}_{\tau(n)} f - c\|_{\infty} .$$

Passons au cas général

• $f \in E$: il existe une suite $\{f_k\}_{k \geq 1}$ de polynômes trigonométriques convergeant dans E vers f. Soit $\varepsilon > 0$. On a

$$\begin{aligned} \|\Pi^{\omega}_q f - \Pi^{\omega}_p f\|_{\infty} &\leq \|\Pi^{\omega}_q f - \Pi^{\omega}_q f_k\|_{\infty} + \|\Pi^{\omega}_q f_k - \Pi^{\omega}_p f_k\|_{\infty} \\ &\quad + \|\Pi^{\omega}_p f_k - \Pi^{\omega}_p f\|_{\infty} \\ &\leq 2\, \|f_k - f\|_{\infty} + \|\Pi^{\omega}_q f_k - \Pi^{\omega}_p f_k\|_{\infty} . \end{aligned}$$

On fixe k assez grand pour que $\|f_k - f\|_{\infty} \leq \varepsilon/3$. On sait que, pour ℓ assez grand, $\Pi^{\omega}_{\ell} f_k \in \mathcal{T}_N$ (car P_0 et P_1 contractent les degrés). On déduit de ce qui précède que la suite $\{\Pi^{\omega}_n f_k\ :\ n \geq 1\}$ converge dans E, et finalement que $\{\Pi^{\omega}_n f\ :\ n \geq 1\}$ est une suite de Cauchy dans E. Cette dernière suite converge donc vers une fonction $h \in E$, et on a en particulier $\lim_{n \to +\infty} P_1(P_{\omega_{\phi(n)-1}} \cdots P_{\omega_1} f) = P_1 h = h$ (car $\omega_{\phi(n)} = 1$). On conclut en utilisant le fait que, sous l'hypothèse (7) (qui assure que

u_1 n'a pas de cycle périodique invariant), les fonctions P_1-invariantes sont constantes, voir [2].

3. Démonstration du Théorème 1.3.

Nous conservons les notations et hypothèses précédentes (voir début du Paragraphe 2), et nous nous proposons de démontrer le Théorème 1.3, c'est-à-dire l'inégalité (6) pour tout $f \in E^\eta$. Pour simplifier les notations, on suppose que $\eta = 1$ (la démonstration est identique pour $\eta \in]0,1]$ quelconque). Remarquons tout d'abord que P_0, P_1, et donc chaque Π_n^ω, sont des opérateurs bornés sur E^1. Plus précisément, on démontre facilement l'existence de constantes $C, R > 0$ telles que l'on ait, pour $j = 0, 1$ et tout $f \in E^1$,

$$m_1(P_j f) \leq 2^{-1} m_1(f) + C\|f\|_\infty \,, \tag{9}$$

$$\|P_j f\|_1 \leq 2^{-1}\|f\|_1 + R\|f\|_\infty \,. \tag{10}$$

On en déduit que, pour tout $\omega \in \Omega$,

$$\|\Pi_2^\omega f\|_1 \leq \frac{1}{4}\,\|f\|_1 + \frac{3}{2}\,R\,\|f\|_\infty \,,$$

et plus généralement, pour tout $n \in \mathbb{N}^*$,

$$\|\Pi_n^\omega f\|_1 \leq 2^{-n}\|f\|_1 + 2\,R\,\|f\|_\infty \,. \tag{11}$$

D'autre part, pour tout $f \in E$, les suites $\{P_0^n f \,:\, n \geq 1\}$ et $\{P_1^n f \,:\, n \geq 1\}$ convergent dans E. En vertu du Théorème de Ionescu-Tulcea et Marinescu [6], [9], la valeur propre 1 est l'unique valeur spectrale de module 1 pour P_0 et P_1, et on obtient les décompositions suivantes sur E^1 :

Pour $i = 0, 1$, il existe une mesure de probabilité ν_i, P_i-invariante sur le tore, et un opérateur Q_i borné sur E^1, de rayon spectral strictement inférieur à 1, tels que

$$P_i f = \int f\, d\nu_i + Q_i(f)\,, \quad \textit{pour tout } f \in E^1\,, \tag{12}$$

avec, en outre, $Q_i\nu_j = 0$ *et* $\nu_i\nu_j = \nu_j$ *pour* $i, j \in \{0,1\}$ *(on a noté* ν_i *l'opérateur défini sur* E^1 *par* $\nu_i(f) = \int f d\nu_i$*). On en déduit que, pour tout* $f \in E^1$*,*

$$\Pi_n^\omega(f) = \nu_{\omega_1}(f) + \nu_{\omega_2}Q_{\omega_1}f + \cdots + \nu_{\omega_n}Q_{\omega_{n-1}}\cdots Q_{\omega_1}f + Q_{\omega_n}\cdots Q_{\omega_1}f\,. \tag{13}$$

La démonstration du Théorème 1.3 utilise les trois lemmes suivants:

Lemme 3.1. *Il existe une constante* $D > 0$ *telle que pour tous* $f \in E^1$*,* $n \in \mathbb{N}^*$*,* $\omega_1, \dots, \omega_n \in \{0,1\}$*,*

$$\|Q_{\omega_n}\cdots Q_{\omega_1}f\|_1 \le D\,\|f\|_1\,.$$

Lemme 3.2. *On a pour tout* $f \in E^1$ *et tout* $\omega = (\omega_1, \omega_2, \dots, \omega_n, \dots) \in \Omega$

a) $\lim_{n\to+\infty}\|Q_{\omega_n}\cdots Q_{\omega_1}f\|_\infty = 0\,.$

b) $\lim_{n\to+\infty}\|Q_{\omega_n}\cdots Q_{\omega_1}f\|_1 = 0\,.$

Lemme 3.3. *Pour tout réel* $\beta > 0$*, il existe* $M \in \mathbb{N}^*$ *tel que pour tous* $f \in E^1$*,* $n \ge M$*,* $\omega_1, \dots, \omega_n \in \{0,1\}$*,*

$$\|Q_{\omega_n}\cdots Q_{\omega_1}f\|_1 \le \beta\,\|f\|_1\,.$$

Commençons par admettre ces lemmes, et donnons la

Démonstration du Théorème 1.3. Soient β tel que $0 < \beta < 1$, et M l'entier du Lemme 3.3 correspondant à β. Soient $\omega \in \Omega$, et $k \in \mathbb{N}^*$ qu'on écrit sous la forme $k = qM + r$, où $q \in \mathbb{N}$ et $r \in \{0, \dots, M-1\}$. Alors

$$\begin{aligned}\|Q_{\omega_k}\cdots Q_{\omega_1}f\|_1 &\le D\,\beta^q\,\|f\|_1\\ &= D'\,(\beta^{1/M})^k\,\|f\|_1\,.\end{aligned}$$

On pose $\rho = \beta^{1/M}$ et $T_kf = Q_{\omega_k}\cdots Q_{\omega_1}f$. Utilisant (13) et le fait que, pour $g \in E^1$ et $j = 0, 1$, $\|\nu_j(g)\|_1 = |\nu_j(g)| \le \|g\|_\infty \le \|g\|_1$, on montre aisément que

$$\begin{aligned}\|\Pi_{n+p}^\omega f - \Pi_n^\omega f\|_1 &\le \|T_nf\|_1 + \sum_{k=n}^{n+p}\|T_kf\|_1\\ &\le 2D'\Big(\sum_{k=n}^{n+p}\rho^k\Big)\|f\|_1\,.\end{aligned}$$

On conclut grâce au critère de Cauchy, et au Théorème 1.2 qui permet d'identifier la limite à une constante.

Preuve du Lemme 3.1. La démonstration de ce lemme est donnée dans [2]. Nous la reprenons ici car elle met en jeu une inégalité importante dont nous aurons besoin dans la suite. Grâce à (9), (12), et enfin aux relations entre Q_i et ν_j, on obtient les majorations suivantes:

$$\begin{aligned}\|Q_{\omega_n}\cdots Q_{\omega_1}f\|_1 &= \|Q_{\omega_n}P_{\omega_{n-1}}\cdots P_{\omega_1}f\|_1\\ &= \|Q_{\omega_n}P_{\omega_{n-1}}\cdots P_{\omega_1}f\|_\infty + m_1(P_{\omega_n}\cdots P_{\omega_1}f)\\ &\le 2\,\|f\|_\infty + 2^{-1}m_1(P_{\omega_{n-1}}\cdots P_{\omega_1}f) + C\,\|f\|_\infty\,,\end{aligned}$$

et finalement

$$\begin{aligned}\|Q_{\omega_n}\cdots Q_{\omega_1}f\|_1 \le{}& (2 + C + C\,2^{-1} + \cdots + C\,2^{-(n-1)})\,\|f\|_\infty\\ &+ 2^{-n}m_1(f)\,.\end{aligned}\tag{14}$$

Preuve du Lemme 3.2. a) On obtient grâce à (13)

$$\nu_{\omega_{n+1}}\Pi_n^\omega f = \sum_{i=1}^{n+1}\nu_{\omega_i}Q_{\omega_{i-1}}\cdots Q_{\omega_1}f\,.$$

Rappelons que $\{\Pi_n^\omega f\}_{n\ge 1}$ converge uniformément vers une constante $c(f,\omega)$. On a en outre

$$\begin{aligned}\|\nu_{\omega_{n+1}}(\Pi_n^\omega f) - c(f,\omega)\|_\infty &= \|\nu_{\omega_{n+1}}\left[\Pi_n^\omega f - c(f,\omega)\right]\|_\infty\\ &\le \|\Pi_n^\omega f - c(f,\omega)\|_\infty\,,\end{aligned}$$

d'où, d'après le Théorème 1.2, $\lim_{n\to+\infty}\|\nu_{\omega_{n+1}}(\Pi_n^\omega f) - c(f,\omega)\|_\infty = 0$. On en déduit que $\{\sum_{i=1}^n \nu_{\omega_i}Q_{\omega_{i-1}}\cdots Q_{\omega_1}f\,:\ n\ge 1\}$ converge dans E vers $c(f,\omega)$ quand $n\to+\infty$. On utilise à nouveau (13) pour en déduire le a) du lemme.

b) On pose $a_n = m_1(Q_{\omega_n}\cdots Q_{\omega_1}f)$ et $b_n = \|Q_{\omega_n}\cdots Q_{\omega_1}f\|_\infty$. En vertu du a), il reste à prouver que $\lim_{n\to+\infty}a_n = 0$. Or, on a $a_k = m_1(P_{\omega_k}Q_{\omega_{k-1}}\cdots Q_{\omega_1}f)$, d'où d'après (9), $a_k \le 2^{-1}a_{k-1} + Cb_{k-1}$, et pour tout $p\ge 1$,

$$a_{n+p} \le 2^{-p}a_n + C\,(b_{n+p-1} + 2^{-1}b_{n+p-2} + \cdots + 2^{-p+1}b_n)\,.$$

L'assertion b) résulte donc du point a) et du fait que la suite $\{a_n\}_{n\ge 1}$ est bornée (*cf.* Lemme 3.1).

Preuve du Lemme 3.3. On note $\mathcal{S}_1$ la sphère unité de E^1. Par ailleurs on définit sur Ω la distance $d(\omega,\omega') = \sum_{k\geq 1} 2^{-k}|\omega_k - \omega'_k|$. Rappelons que (Ω, d) est compact.

Nous procédons par l'absurde en supposant qu'il existe un réel $\beta > 0$ pour lequel on a la propriété suivante: pour tout $n \in \mathbb{N}^*$, il existe $\psi(n) \geq n$, $\omega_1^{\psi(n)}, \ldots, \omega_{\psi(n)}^{\psi(n)} \in \{0,1\}$, $f_{\psi(n)} \in \mathcal{S}_1$, tels que

$$\|Q_{\omega_{\psi(n)}^{\psi(n)}} \cdots Q_{\omega_1^{\psi(n)}} f_{\psi(n)}\|_1 > \beta\,.$$

On pose $\omega^{\psi(n)} = (\omega_1^{\psi(n)}, \ldots, \omega_{\psi(n)}^{\psi(n)}, 0, 0, \ldots) \in \Omega$. En vertu du Théorème d'Ascoli et de la compacité de Ω, il existe une suite d'entiers positifs $\{\phi(n)\}_{n\geq 1}$ strictement croissante (extraite de $\{\psi(n)\}_{n\geq 1}$) telle que $\{\omega^{\phi(n)}\}_{n\geq 1}$ converge vers $\omega \in \Omega$, et telle que $\{f_{\phi(n)}\}_{n\geq 1}$ converge dans E vers $f \in E^1$, avec $\|f\|_1 \leq 1$. Posant

$$A_n = Q_{\omega_{\phi(n)}^{\phi(n)}} \cdots Q_{\omega_1^{\phi(n)}}\ ,$$

on obtient

$$\beta < \|A_n f_{\phi(n)}\|_1 \leq \|A_n(f_{\phi(n)} - f)\|_1 + \|A_n f\|_1\ .$$

Nous allons démontrer que ces deux derniers termes ont une limite nulle quand $n \to +\infty$, ce qui constituera bien une contradiction:

De (14), il vient que

$$\|A_n(f_{\phi(n)} - f)\|_1 \leq E\ \|f_{\phi(n)} - f\|_\infty + 2^{-\phi(n)} m_1(f_{\phi(n)} - f)\,,$$

avec $E = 2(C+1)$. En outre, on a $m_1(f_{\phi(n)} - f) \leq 2$, d'où

$$\lim_{n\to+\infty} \|A_n(f_{\phi(n)} - f)\|_1 = 0\,.$$

La convergence de $\{\omega^{\phi(n)}\}_{n\geq 1}$ vers $\omega = (\omega_1, \ldots, \omega_n, \ldots)$ entraîne qu'il existe une suite $\{k(n)\}_{n\geq 1}$ d'entiers positifs, avec $\lim_{n\to+\infty} k(n) = +\infty$, telle que $w_k^{\phi(n)} = \omega_k$ pour tout $1 \leq k \leq k(n)$. Si $k(n) \geq \phi(n)$, alors $A_n = Q_{\omega_{\phi(n)}} \cdots Q_{\omega_1}$. Si $k(n) < \phi(n)$, alors

$$A_n = Q_{\omega_{\phi(n)}^{\phi(n)}} \cdots Q_{\omega_{k(n)+1}^{\phi(n)}} (Q_{\omega_{k(n)}} \cdots Q_{\omega_1})\,,$$

d'où $\|A_n f\|_1 \leq D\ \|Q_{\omega_{k(n)}} \cdots Q_{\omega_1} f\|_1$ d'après le Lemme 3.1. On déduit du Lemme 3.2.b) que $\lim_{n\to+\infty} \|A_n f\|_1 = 0$, ce qui achève la démonstration du Lemme 3.3.

References.

[1] Coifman, R., Meyer, Y., Quake, S., Wickerhauser M.V., Signal processing and compression with wave packets. *Proc. of the Conference on Wavelets*, Marseille, Spring 1989.

[2] Conze, J. P., Raugi, A., Fonctions harmoniques pour un opérateur de transition et applications. *Bull. Soc. Math. France* **118** (1990), 273-310.

[3] Esteban, D., Galand, C., Application of Quadrature Mirror Filters to Split-band Voice Coding Schemes. *Proc. of ICASSP*, Hartford, Connecticut, 1977.

[4] Hervé, L., Etude d'opérateurs quasi-compacts et positifs. Applications aux opérateurs de transfert *Ann. Inst. H. Poincaré, Probabilités et Statistique.* **30**, (1994), 437-466.

[5] Hervé, L., Comportement asymptotique dans l'algorithme de transformée en ondelettes. Lien avec la régularité de l'ondelette. *Revista Mat. Iberoamericana* **11** (1995), 431-451.

[6] Ionescu-Tulcéa, C.T., Marinescu, G., Théorie ergodique pour une classe d'opérations non complétement continues. *Ann. of Math.* **52** (1950), 140-147.

[7] Keane, M., Strongly mixing g-measures. *Invent. Math.* **16** (1972), 309-324.

[8] Mallat, S., Multiresolution approximations and wavelet orthonormal bases of $L^2(\mathbb{R})$. *Trans. Amer. Math. Soc.* **315** (1989), 69-88.

[9] Norman, M. F., *Markov processes and learning models.* Academic Press, 1972.

Recibido: 3 de mayo de 1.995
Revisado: 6 de septiembre de 1.995

Loïc Hervé
I.R.M.A.R.-Universitè de Rennes 1
Laboratoire de Probabilités
Campus de Beaulieu, 35042 Rennes Cedex
FRANCE
herve@univ-rennes1.fr

Revista Matemática Iberoamericana
Vol. 12, N.° 3, 1996

Spectral factorization of measurable rectangular matrix functions and the vector-valued Riemann problem

Marek Rakowski and Ilya Spitkovsky

Abstract. We define spectral factorization in L_p (or a generalized Wiener-Hopf factorization) of a measurable singular matrix function on a simple closed rectifiable contour Γ. Such factorization has the same uniqueness properties as in the nonsingular case. We discuss basic properties of the vector valued Riemann problem whose coefficient takes singular values almost everywhere on Γ. In particular, we introduce defect numbers for this problem which agree with the usual defect numbers in the case of a nonsingular coefficient. Based on the Riemann problem, we obtain a necessary and sufficient condition for existence of a spectral factorization in L_p.

1. Introduction.

Let Γ be a simple closed rectifiable contour which is the positively oriented boundary of a finitely connected region $\mathcal{D}_+$, and let $\mathcal{D}_- = \mathbb{C}_\infty \setminus (\mathcal{D}_+ \cup \Gamma)$. Let G and g be functions on Γ. The *Riemann problem* consists in finding functions ϕ_+ and ϕ_- which are analytic in $\mathcal{D}_+$ and $\mathcal{D}_-$, respectively, and whose nontangential boundary limits satisfy equation

$$\phi_+(t) + G(t)\,\phi_-(t) = g(t)\,. \tag{1.1}$$

This problem is also called a *Hilbert problem* [6], or a *barrier problem* [2], in the literature. The name Hilbert problem originates in [7], where the homogeneous version of the problem was considered under the assumptions that Γ is a smooth contour which is a boundary of a simply connected region, G is twice differentiable, and the scalar functions ϕ_+ and ϕ_- are continuous up to Γ.

A classical solution of the Riemann problem in the case where Γ is smooth and bounds a finitely connected region, G and g are Hölder continuous, and G does not vanish, is as follows. Assume $\Gamma = \Gamma_0 \cup \Gamma_1 \cup \cdots \cup \Gamma_N$ where Γ_0 encloses $\Gamma_1 \cup \cdots \cup \Gamma_N$, and consider the homogeneous problem

$$\phi_+(t) = -G(t)\,\phi_-(t)\,. \tag{1.2}$$

Suppose the change of argument of $G(t)$ along the contour Γ_i is $2\pi\lambda_i$, $i = 0, 1, \ldots, N$. Assume $0 \in \mathcal{D}_+$, and pick a point α_i in the hole bordered by Γ_i $(i = 1, 2, \ldots, N)$. Let

$$\pi(z) = (z-\alpha_1)^{\lambda_1}(z-\alpha_2)^{\lambda_2}\cdots(z-\alpha_N)^{\lambda_N}, \tag{1.3}$$

let $\kappa = \lambda_0 + \lambda_1 + \cdots + \lambda_N$, and let

$$G_0(t) = -t^{-\kappa}\pi(t)\,G(t)\,. \tag{1.4}$$

Then $\log G_0(t)$ is continuous on Γ and satisfies the Hölder condition. Consequently, if

$$\gamma(z) = \frac{1}{2\pi i}\int_\Gamma \frac{\log G_0(t)}{t-z}\,dt \tag{1.5}$$

and $\gamma_\pm(z) = \gamma(z)$ for $z \in \mathcal{D}_\pm$, $\gamma_+(t) - \gamma_-(t) = \log G_0(t)$. Hence $e^{\gamma_+(t)} = e^{\gamma_-(t)}G_0(t)$, and

$$\varphi_+(z) = \frac{1}{\pi(z)}\,e^{\gamma(z)} \qquad \text{and} \qquad \varphi_-(z) = z^{-\kappa}e^{\gamma(z)} \tag{1.6}$$

are functions whose nontangential limits to Γ are Hölder continuous and satisfy equation (1.2). Functions φ_+ and φ_- can be used to obtain solution of the nonhomogeneous problem.

Equation (1.4) shows that the Riemann problem can be approached through factorization of its coefficient. Suppose we can find a factorization

$$G(t) = G_+(t)\,\Lambda(t)\,G_-(t)\,, \tag{1.7}$$

where $G_+(t)$ and $1/G_+(t)$ are boundary values of functions analytic in $\mathcal{D}_+$ and continuous up to Γ, $G_-(t)$ and $1/G_-(t)$ are boundary values of functions analytic in $\mathcal{D}_-$ and continuous up to Γ, and $\Lambda(t) = (t-t_+)^\kappa/(t-t_-)^\kappa$ for some points $t_+ \in \mathcal{D}_+$ and $t_- \in \mathcal{D}_-$ and an integer κ. Then (1.1) is equivalent to

$$\frac{\phi_+(t)}{G_+(t)} + \Lambda(t)\,G_-(t)\,\phi_-(t) = \frac{g(t)}{G_+(t)}\,. \tag{1.8}$$

The decomposition $g(t)/G_+(t) = g_+(t) + g_-(t)$, where g_+ (respectively g_-) is a boundary value of a function analytic in $\mathcal{D}_+$ (respectively $\mathcal{D}_-$) and continuous up to Γ, immediately yields all solutions of equation (1.1). We note that factorization (1.7) exists *e.g.* when G is Hölder continuous and does not vanish on Γ [2].

The factorization approach applies naturally to more general versions of the Riemann problem considered in the literature. The problem with $G(t)$ a square nonsingular matrix valued function has been treated in [6]. Factorability of an essentially bounded nonsingular matrix function G and the Riemann problem in L_p were considered in [12]. The case where G is a measurable nonsingular matrix function and ϕ_+ and ϕ_- are in $L_p(\Gamma)$ has been treated in [13] (see also [9]). Below, we extend some of the results presented in [9] to the case where G takes singular values. In particular, we relate the properties of the Riemann problem with a measurable singular matrix valued coefficient with existence of a factorization of the coefficient.

Let G be a continuous nonsingular matrix valued function on a simple closed rectifiable contour Γ. A (*left*) *standard factorization* of G relative to Γ is a factorization $G = G_+\Lambda\,G_-$ where $G_+(z)$ and $G_+(z)^{-1}$ are analytic in $\mathcal{D}_+$ and continuous up to Γ, $G_-(z)$ and $G_-(z)^{-1}$ are analytic in $\mathcal{D}_-$ and continuous up to Γ, and

$$\Lambda(t) = \begin{pmatrix} \left(\dfrac{t-t_+}{t-t_-}\right)^{\kappa_1} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \left(\dfrac{t-t_+}{t-t_-}\right)^{\kappa_n} \end{pmatrix} \tag{1.9}$$

for integers $\kappa_1 \geq \cdots \geq \kappa_n$. This factorization is also called a *Wiener-Hopf factorization* or a *spectral factorization relative to* Γ. The properties of a standard factorization relative to Γ are described in [2].

Let E_{p+} (respectively E_{p-}) be the space of functions f analytic in $\mathcal{D}_+$ (respectively $\mathcal{D}_-$) such that $\{\int_{\Gamma_k} |f|^p\}$ is bounded for some sequence of rectifiable contours Γ_k approaching Γ in $\mathcal{D}_+$ (respectively $\mathcal{D}_-$; see [5]). If the components of $G_+(z)$ and $G_+(z)^{-1}$ are in E_{p+} and E_{q+}, where $1/p + 1/q = 1$, the components of $G_-(z)$ and $G_-(z)^{-1}$ are in E_{q-} and E_{p-}, and Λ is given by (1.9), $G = G_+\Lambda G_-$ is called a (*left*) *factorization in* L_p [9]. We note that factorization with a different Λ has been considered in [14].

A function G may admit a left factorization in L_p although the space of all $g \in L_p(\Gamma)$ for which the problem (1.1) is solvable is not closed. Suppose the contour Γ is such that the operator of singular integration $(\mathcal{S}f)(t) = (1/\pi i)\int_\Gamma f(\tau)/(\tau - t)\, d\tau$ on the space $L_p(\Gamma)$ is bounded. Suppose G and its multiplicative inverse are essentially bounded, and $G = G_+\Lambda G_-$ is a factorization in L_p. Then the set of all $g \in L_p(\Gamma)$ for which problem (1.1) is solvable is a closed subspace of $L_p(\Gamma)$ if and only if the operator $G_+\mathcal{S}G_+^{-1}$ is bounded. If G and G^{-1} are bounded, a factorization $G = G_+\Lambda G_+^{-1}$ in L_p with the operator $G_+\mathcal{S}G_+^{-1}$ bounded is called in [2] a *generalized* (*left*) *standard factorization relative to* Γ.

The definition of a standard factorization relative to a contour has been extended to the singular case in [3] by requiring that G_+ have a left (respectively G_- a right) multiplicative inverse which is analytic in $\mathcal{D}_+$ (respectively in $\mathcal{D}_-$) and continuous up to the boundary, and that Λ be a square nonsingular diagonal matrix function as in (1.9). If G is a rational matrix function, a necessary and sufficient condition for existence of a canonical standard factorization ($\kappa_1 = \cdots = \kappa_k = 0$), together with realization formulas for the factors, has been obtained in [11]. Below, we apply this idea to factorization in L_p of measurable singular matrix valued functions. In addition to allowing functions to take singular matrix values, we make only general assumptions on contours. We assume that the contour Γ is simple, closed, and rectifiable. We do not require that Γ be regular [4] or Smirnov. Thus, the operator of singular integration on the space $L_p(\Gamma)$ is in general unbounded.

The paper is organized as follows. In Section 2 we indicate basic properties of factorization in L_p of singular matrix functions. In Section 3 we discuss the vector valued Riemann problem with singular matrix valued coefficient G. In Section 4 we relate the factorization of the coefficient G with the Riemann problem.

2. Spectral factorization in L_p.

Below, L_p with $p \geq 1$ will denote $L_p(\Gamma)$ (with respect to the usual Lebesgue measure). We will denote by L_{p+} and L_{p-} the closed subspaces of L_p formed by nontangential boundary limits of functions in E_{p+} and E_{p-}, where $E_{p\pm}$ are as defined above and $E_{\infty+}$ (respectively $E_{\infty-}$) is the space of functions analytic and bounded in $\mathcal{D}_+$ (respectively $\mathcal{D}_-$). We will identify L_{p+} and L_{p-} with E_{p+} and E_{p-}. $\dot{L}_{p-}$ will denote functions in E_{p-} which vanish at infinity. If $X \in \{L_p, L_{p+}, L_{p-}, \dot{L}_{p-}\}$, we will denote by $X^{m\times n}$ the space of $m \times n$ matrices over X. To simplify notation, we will write X^n instead of $X^{1\times n}$ or $X^{n\times 1}$.

Definition 2.1. *Let G be an $m \times n$ matrix valued function with measurable entries and let $p > 1$. By a (left) spectral factorization in L_p relative to Γ we will understand a factorization*

$$G = G_+ \Lambda\, G_- \,, \tag{2.1}$$

where

i) *$G_+ \in L_{p+}^{m\times k}$ and there exists $G_+^L \in L_{q+}^{k\times m}$ (with $q = p/(p-1)$) such that $G_+^L(t)\, G_+(t) = I$ almost everywhere on Γ,*

ii) *$G_- \in L_{q-}^{k\times n}$ and there exists $G_-^R \in L_{p-}^{n\times k}$ such that $G_-(t)\, G_-^R(t) = I$ almost everywhere on Γ,*

iii) *the middle factor*

$$\Lambda(t) = \begin{pmatrix} \left(\frac{t-t_+}{t-t_-}\right)^{\kappa_1} & & & \mathbf{0} \\ & \left(\frac{t-t_+}{t-t_-}\right)^{\kappa_2} & & \\ & & \ddots & \\ \mathbf{0} & & & \left(\frac{t-t_+}{t-t_-}\right)^{\kappa_k} \end{pmatrix}, \tag{2.2}$$

where t_+ is a point inside Γ, t_- is a point outside Γ, and $\kappa_1 \geq \kappa_2 \geq \cdots \geq \kappa_k$ are integers.

A right spectral factorization of G relative to Γ is a factorization $G = G_- \Lambda\, G_+$ with Λ as above and $G_- \in L_{p-}^{m\times k}$ and $G_+ \in L_{q+}^{k\times n}$

such that there exist functions $G_-^L \in L_{q-}^{k\times m}$ *and* $G_+^R \in L_{p+}^{k\times n}$ *for which* $G_-^L(t)\,G_-(t) = I$ *and* $G_+(t)\,G_+^R(t) = I$ *almost everywhere on* Γ.

Note that if a function G admits a spectral factorization in L_p relative to Γ, then the rank of G is constant almost everywhere on Γ. Also, since $\Lambda \in L_\infty^{k\times k}$, by Hölder's inequality $G \in L_1^{m\times n}$. To simplify notation, we will assume $0 \in \mathcal{D}_+$ and write

$$\Lambda(t) = \begin{pmatrix} t^{\kappa_1} & & & \mathbf{0} \\ & t^{\kappa_2} & & \\ & & \ddots & \\ \mathbf{0} & & & t^{\kappa_k} \end{pmatrix}. \tag{2.3}$$

We show first that the integers $\kappa_1, \kappa_2, \dots, \kappa_k$ are unique.

Theorem 2.2. *Suppose* $1 < p_1 \le p_2 < \infty$ *and let* $G_{1+}\Lambda_1 G_{1-}$ *and* $G_{2+}\Lambda_2 G_{2-}$ *with*

$$\Lambda_1(t) = \begin{pmatrix} t^{\kappa_1^{(1)}} & & & \mathbf{0} \\ & t^{\kappa_2^{(1)}} & & \\ & & \ddots & \\ \mathbf{0} & & & t^{\kappa_k^{(1)}} \end{pmatrix}$$

and

$$\Lambda_2(t) = \begin{pmatrix} t^{\kappa_1^{(2)}} & & & \mathbf{0} \\ & t^{\kappa_2^{(2)}} & & \\ & & \ddots & \\ \mathbf{0} & & & t^{\kappa_k^{(2)}} \end{pmatrix}$$

be spectral factorizations in L_{p_1} *and* L_{p_2} *of a function* $G \in L_1^{m\times n}$ *relative to a contour* Γ. *Then* $\kappa_j^{(1)} \ge \kappa_j^{(2)}$ *for* $j = 1, 2, \dots, k$.

PROOF. Let $G_{2-}^R \in L_{p_2-}^{n\times k}$ and $G_{1+}^L \in L_{q_1+}^{k\times m}$ be right and left multiplicative inverses of G_{2-} and G_{1+}. Then

$$\Lambda_1 H_- = H_+ \Lambda_2\ , \tag{2.4}$$

where $H_+ = G_{1+}^L G_{2+} \in L_{p+}^{k\times k}$ and $H_- = G_{1-} G_{2-}^R \in L_{p-}^{k\times k}$ with $p = 1/(1/q_1 + 1/p_2) = 1/(1 - 1/p_1 + 1/p_2) \ge 1$. Also, G_{1+} and G_{2+} have the same column span almost everywhere on Γ, so H_+ takes nonsingular values almost everywhere on Γ. Similarly, H_- takes nonsingular values almost everywhere on Γ.

It follows from (2.4) that

$$t^{\kappa_i^{(1)}-\kappa_j^{(2)}} H_-(i,j) = H_+(i,j)\,. \tag{2.5}$$

Since $L_{p+} \cap L_{p-}$ consists of constants, $H_+(i,j) = 0$ if $\kappa_j^{(2)} > \kappa_i^{(1)}$ and $H_+(i,j)$ is a polynomial of degree at most $\kappa_i^{(1)} - \kappa_j^{(2)}$ otherwise. Suppose $\kappa_r^{(2)} > \kappa_r^{(1)}$. Then, for all $j \le r$ and $i \ge r$, $H_+(i,j) = 0$ contradicting nonsingularity of H_+ almost everywhere on Γ.

Corollary 2.3. *The integers* $\kappa_1, \kappa_2, \ldots, \kappa_k$ *in* (2.2) *are unique.*

The integers $\kappa_1, \kappa_2, \ldots, \kappa_k$ in (2.2) or (2.3) are called the *indices* of the factorization, and the sum of all indices is called the *total index* of the factorization. If all the indices of the factorization are equal to 0, the factorization is said to be *canonical*.

The proof of Theorem 2.2 actually gives the nonuniqueness of all the factors in a spectral factorization.

Theorem 2.4. *Suppose* $1 < p_2 \le p_1$,

$$G_{1+} \Lambda\, G_{1-} \tag{2.6}$$

is a spectral factorization in L_{p_1} *of a function* G *relative to a contour* Γ, *and* G *admits spectral factorization in* L_{p_2} *relative to* Γ *with the same total index. Then*

$$G_{2+} \Lambda\, G_{2-} \tag{2.7}$$

is a spectral factorization in L_{p_2} *of* G *relative to* Γ *if and only if*

$$G_{2+} = G_{1+} H_+ \qquad \text{and} \qquad G_{2-} = \Lambda^{-1} H_+^{-1} \Lambda\, G_{1-} \tag{2.8}$$

where H_+ *is a matrix polynomial such that* $\det H_+ \not\equiv 0$ *and*

i) $H_+(i,j) = 0$ *if* $\kappa_i < \kappa_j$,
ii) $H_+(i,j)$ *is a constant if* $\kappa_i = \kappa_j$,
iii) $\deg\ H_+(i,j) \le \kappa_i - \kappa_j$ *if* $\kappa_i > \kappa_j$.

PROOF. Suppose (2.7) is a spectral factorization in L_{p_2} of G. In the notation of the proof of Theorem 2.2,

$$H_+ = G_{1+}^L G_{2+} \tag{2.9}$$

where H_+ is a matrix polynomial whose determinant is not equal to zero identically and which satisfies properties i)-iii). In particular, $H_+ \in L_{\infty+}$. Multiplying both sides of (2.9) by G_{1+} we obtain

$$G_{1+}H_+ = G_{1+}G_{1+}^L G_{2+}\,. \tag{2.10}$$

Since $G_{1+}G_{1+}^L$ is a projection onto the column span of G_{1+}, and the column spans of G_{1+} and G_{2+} coincide almost everywhere on Γ,

$$G_{1+}(z)\,G_{1+}^L(z)\,G_{2+}(z) = G_{2+}(z) \tag{2.11}$$

for almost everywhere $z \in \Gamma$. Since a function analytic in $\mathcal{D}_+$ with nontangential boundary values equal to 0 on a set of positive measure is identically 0, equality (2.11) is valid inside Γ and

$$G_{1+}H_+ = G_{2+}\,. \tag{2.12}$$

Hence

$$G_{1+}\Lambda\,G_{1-} = G_{2+}\Lambda\,G_{2-} = G_{1+}H_+\Lambda\,G_{2-} = G_{1+}\Lambda\Lambda^{-1}H_+\Lambda\,G_{2-}$$

and the second equality in (2.8) holds as well.

Suppose now (2.6) is a spectral factorization of G in L_{p_1} relative to Γ and H_+ with $\det H_+ \not\equiv 0$ satisfies conditions i)-iii) of the theorem, and $G_{2\pm}$ satisfy (2.8). Then $\det H_+$ is a nonzero constant, and $H_+^{\pm 1} \in L_{\infty+}$. Also, $\Lambda^{-1}H_+\Lambda$ is a matrix polynomial in $1/z$ with a nonzero constant determinant, and $(\Lambda^{-1}H_+\Lambda)^{\pm 1} \in L_{\infty-}$. Hence $G_{2+} \in L_{p_1+}^{m\times k}$, $G_{2-} \in L_{q_1-}^{k\times n}$, and G_{2-}^R, a right multiplicative inverse of G_2, is an element of $L_{p_1-}^{k\times n}$. Suppose $\tilde{G}_{2+}\Lambda\tilde{G}_{2-}$ is a factorization of G in L_{p_1} relative to Γ and $\tilde{G}_{2+}^L \in L_{q_2+}^{k\times m}$ is a left multiplicative inverse of G_2. Then

$$\tilde{G}_{2+}^L G_{2+}\Lambda = \Lambda\,\tilde{G}_{2-}G_{2-}^R\ ,$$

or $G_+\Lambda = \Lambda\,G_-$ where $G_+ = \tilde{G}_{2+}^L G_{2+} \in L_{1+}^{k\times k}$ and $G_- = \tilde{G}_{2-}G_{2-}^R \in L_{1-}^{k\times k}$. By the same argument as above, G_+ is a unimodular matrix polynomial and G_{2+} has a left multiplicative inverse in $L_{q_2+}^{k\times m}$.

Since $G_{2+}\Lambda\, G_{2-} = \tilde{G}_{2+}\Lambda\, \tilde{G}_{2-}$ and the functions Λ, Λ^{-1}, and G_+^{-1} are bounded, $G_{2-} \in L_{q_2-}^{k\times n}$. Thus, (2.7) with G_{2+} and G_{2-} given by (2.8) is a spectral factorization of G in L_{p_2} relative to Γ.

In particular, Theorem 2.4 determines possible nonuniqueness of a spectral factorization in L_p of a function G relative to Γ. It also has the following corollary.

Corollary 2.5. *Suppose* $1 < p_1 < p_2$ *and a matrix function* G *admits spectral factorizations in* L_{p_1} *and* L_{p_2} *relative to* Γ *with the same total index. Then*

i) G *admits a spectral factorization in* L_p *relative to* Γ *for every* $p \in [p_1, p_2]$,

ii) *if* $p_0 \in [p_1, p_2]$, *a spectral factorization in* L_{p_0} *of* G *relative to* Γ *is a spectral factorization in* L_p *for all* $p \in [p_1, p_2]$.

A meromorphic matrix function W has a pole at a point $\lambda \in \mathbb{C}$ if it has a nonzero coefficient at a negative power of $z - \lambda$ in the Laurent expansion at λ. Equivalently, W has a pole at λ if at least one of its entries has a pole at λ. The function W has a zero at λ if each meromorphic multiplicative generalized inverse of W has a pole at λ. If the function W is analytic at λ, it has a zero at λ if the rank of $W(z)$ drops at $z = \lambda$. Every rational matrix function without poles or zeros on Γ admits a spectral factorization relative to Γ with all the factors rational (see [2] for the discussion of the regular case, that is, the case where the function is square and takes nonsingular values at all but a finite number of points; the argument in the nonregular case is similar). Later, we will need the following observation.

Proposition 2.6. *If* $G \in L_1^{m\times n}$ *admits a spectral factorization in* L_p *relative to* Γ *and* F *and* H *are rational* $M \times m$ *and* $n \times N$ *matrix functions analytic and with full column respectively row rank on* Γ, *then the function* FGH *also admits a spectral factorization in* L_p *relative to* Γ.

Proof. Let $\tilde{G}_+\tilde{\Lambda}\,\tilde{G}_-$ be a spectral factorization in L_p relative to Γ of the function G. Since F is a rational matrix function, there is a finite set $\{\lambda_1, \lambda_2, \ldots, \lambda_\eta\} \subset \mathcal{D}_+$ which contains all the poles and zeros of F in $\mathcal{D}_+$. Pick $\lambda \in \{\lambda_1, \lambda_2, \ldots, \lambda_\eta\}$. After multiplying $F\tilde{G}_+$ on the right by a unimodular matrix polynomial in $z - \lambda$, we can obtain a matrix function whose columns have linearly independent leading coefficients in the Laurent expansion at λ. Indeed, suppose $F\tilde{G}_+ = (\, f_1 \quad f_2 \quad \ldots \quad f_k\,)$ and the leading coefficients in the Laurent expansions at λ of f_i's are linearly dependent. Then we can replace say f_i by

$$\tilde{f}_i(z) = f_i(z) - \sum_{\substack{j=1 \\ j\neq i}}^{n} c_j\,(z-\lambda)^{\gamma_j} f_j(z)\,, \tag{2.13}$$

with c_j's constants and γ_j's nonnegative integers such that $\tilde{f}_i$ has a pole at λ of a smaller order, or vanishes at λ to a higher order, than f_i. Since the columns of F are linearly independent over the field of scalar rational functions, for every function ϕ analytic and nonzero at λ the order of the zero at λ of the product $F\tilde{G}_+\phi$ is bounded by the largest partial multiplicity of the zero of F at λ. Hence the finite number of operations as in (2.13) can provide a matrix function whose columns have linearly independent leading coefficients in the Laurent expansions at λ. It follows that there exists a square rational matrix function R_1 whose determinant is not identically equal to zero and which has neither poles nor zeros on Γ such that $F\tilde{G}_+R_1 = \widehat{G}_+ \in L_{p+}^{m\times k}$ has full column rank at all points $z\in\mathcal{D}_+$ and $R_1^{-1}\tilde{G}_+^L F^{-1} \in L_{q+}^{k\times m}$.

Similarly, there exists a square rational matrix function R_2 whose determinant is not equal to zero identically and which does not have poles or zeros on Γ such that $R_2\tilde{G}_-H = \widehat{G}_- \in L_{q-}^{k\times n}$ has a right multiplicative inverse in $L_{p-}^{n\times k}$. If $\widehat{R}_1\Lambda\widehat{R}_2$ is a spectral factorization relative to Γ of the rational matrix function $R_1^{-1}\tilde{\Lambda}R_2^{-1}$, $(\widehat{G}_+\widehat{R}_1)\Lambda(\widehat{R}_2\widehat{G}_-)$ is a spectral factorization in L_p relative to Γ of the function G.

We illustrate the concepts of this section with an example.

Example 2.7 Let Γ be the unit circle. Pick a branch of $z^{1/3}$ on $\mathbb{C}\setminus(-\infty,0)$, and let

$$G(t) = \begin{pmatrix} (t^{1/3})^2 \\ (t^{1/3})^5 \end{pmatrix}\,,$$

where the value of $t^{1/3}$ is determined almost everywhere by the selected branch. Let $\Phi(z)$ be a branch of $(z+1)^{2/3}$ which is analytic in $\mathbb{C}\setminus(-\infty,-1]$, and let $\Psi(z)$ be a branch of $(z/(z+1))^{2/3}$ which is analytic in $\mathbb{C}_\infty\setminus[-1,0]$, such that

$$G(t) = \begin{pmatrix} \Phi(t) \\ t\,\Phi(t) \end{pmatrix} (\,\Psi(t)\,) =: G_+(t)\,G_-(t)\,. \tag{2.14}$$

Let $p > 3$. Then $G_+ \in L_{p+}^{2\times 1}$ and $G_- \in L_{q-}$. Also, $G_-^{-1} \in L_{p-}$ and G_+ has a left multiplicative inverse $G_+^L(z) = (\,\Phi(z)^{-1} \quad 0\,) \in L_{q+}^{1\times 2}$. Thus, (2.14) is a canonical spectral factorization of G in L_p relative to the circle.

Suppose $p \in (1,3)$. From (2.14),

$$(2.15)\quad G(t) = \begin{pmatrix} \dfrac{1}{t+1}\,\Phi(t) \\ \dfrac{t}{t+1}\,\Phi(t) \end{pmatrix} (\,t\,) \left(\frac{t+1}{t}\,\Psi(t) \right) =: \widehat{G}_+(t)\,(\,t\,)\,\widehat{G}_-(t)\,.$$

Plainly, $\widehat{G}_+ \in L_{p+}^{2\times 1}$ and $\widehat{G}_- \in L_{q-}$. Also, $\widehat{G}_-^{-1} \in L_{p-}$ and $\widehat{G}_+$ has a left multiplicative inverse $\widehat{G}_+^L(z) = (\,(z+1)/\Phi(z) \quad 0\,) \in L_{q+}^{1\times 2}$. Thus, (2.15) is a spectral factorization of G in L_p relative to the circle.

Suppose G admits a spectral factorization in L_3 relative to the circle. By Theorem 2.2, the total index of the factorization is either 0 or 1. Then, by Corollary 2.5, either (2.14) or (2.15) is a spectral factorization of G in L_3 relative to the circle. Since $G_- \notin L_{3/2-}$ and $\widehat{G}_+ \notin L_{3+}$, this is a contradiction. Thus, G admits a spectral factorization in L_p relative to the circle if and only if $p \in (1,3) \cup (3,\infty)$.

3. Vector-valued Riemann problem with singular coefficient.

Suppose G is a measurable $m \times n$ matrix valued function on a contour Γ, and $p > 1$. The vector-valued Riemann problem consists in finding for a given function $g \in L_p^m$ a pair of functions (ϕ_+, ϕ_-) with $\phi_+ \in L_{p+}^m$ and $\phi_- \in \dot{L}_{p-}^n$ such that

$$(3.1)\qquad \phi_+(t) + G(t)\,\phi_-(t) = g(t)\,.$$

For brevity, we will refer to this problem as the *Riemann problem* with *coefficient* G. The set of all functions $g \in L_p^m$ for which the problem is solvable is called the *image* of the problem. If the image of the Riemann problem is closed, the problem is said to be *normally solvable.* The set of all solutions of the homogeneous problem is called the *kernel* of the problem.

The *dual problem* consists in finding for a given $h \in L_q^n$ a pair of functions $\psi_- \in \dot{L}_{q-}^n$ and $\psi_+ \in L_{q+}^m$ such that

$$(3.2)\qquad \psi_-(t) + G^T(t)\,\psi_+(t) = h(t)\,.$$

Here q is the conjugate exponent to p, that is, $1/p+1/q=1$. Similarly as in the case where G takes nonsingular values almost everywhere on Γ [9], there is a connection between the Riemann problem and its dual. Identify L_q^n with the dual space of L_p^n through the map

$$\langle f,\, g\rangle = \sum_{j=1}^{n} \int_\Gamma f_j(t) g_j(t)\, dt$$

for all $f(t)=\sum_{j=1}^n f_j(t)e_j \in L_p^n$ and all $g(t)=\sum_{j=1}^n g_j(t)e_j \in L_q^n$. If $\mathcal{L}\subset L_p^n$, the *annihilator* of $\mathcal{L}$ is the closed subspace of L_q^n

$$\{\, g\in L_q^n :\ \langle f,\, g\rangle = 0, \quad \text{for all } f\in\mathcal{L}\,\}\,.$$

Proposition 3.1. *The annihilator of the image of the Riemann problem with coefficient G contains the space of "+" components of elements in the kernel of its dual. If $G\in L_\infty^{m\times n}$, the two spaces coincide.*

PROOF. Suppose $\psi_- + G^T\psi_+ = 0$ for some $\psi_-\in \dot{L}_{q-}^n$ and $\psi_+\in L_{q+}^m$. Then $\psi_+^T G = -\psi_- \in \dot{L}_{q-}^n$, and hence

$$\langle \psi_+,\, (\phi_+ + G\phi_-)\rangle \;=\; \langle\psi_+,\,\phi_+\rangle \;-\; \langle\psi_-,\,\phi_-\rangle \;= 0\,,$$

for all $\phi_+\in L_{p+}^m$ and $\phi_-\in\dot{L}_{p-}$. Thus, ψ_+ annihilates the image of the problem.

Suppose $\langle\psi,\,\phi_+ + G\phi_-\rangle \;=0$ for all $\phi_+\in L_{p+}^m$ and all $\phi_-\in\dot{L}_{p-}^n$ such that $G\phi_-\in L_p^m$. Then $\langle\psi,\,\phi_+\rangle \;=\; 0$ for all $\phi_+\in L_{p+}^m$ and $\psi =: \psi_+\in L_{q+}^m$. If $G\in L_\infty^{m\times n}$, $G\phi_-\in L_p^m$ for all $\phi_-\in\dot{L}_{p-}^n$ and so $G^T\psi_+$ annihilates $\dot{L}_{p-}^n$. That is, $G^T\psi_+\in\dot{L}_{q-}^n$ and ψ_+ is the "+" component of an element in the kernel of the dual problem.

If the coefficient G of a Riemann problem takes almost everywhere nonsingular values, the defect numbers of the problem are the dimension α_R of the kernel and the co-dimension β_R of the closure of the image of the problem. If G takes singular values, both α_R and β_R are generically infinite. In view of Proposition 3.1, β_R can be defined as the co-dimension of $\{\psi_+\in L_{q+}^m :\ \psi_+ G = 0\}$ in the annihilator of the image of the problem. This definition discards the generic left kernel of G.

A similar observation holds for the dual problem.

Proposition 3.2. *The annihilator of the image of the dual problem contains the space of "−" components of elements in the kernel of the problem. If $G \in L_\infty^{m\times n}$, the two spaces coincide.*

Suppose G takes nonsingular values almost everywhere on Γ. Then

$$\begin{aligned} \{(\psi_+,\psi_-) \in L^n_{q+} \times \dot{L}^n_{q-} :\ & \psi_- + G^T\psi_+ = 0\} \\ &\cong \{\psi_+ \in L^n_{q+} :\ G^T\psi_+ \in \dot{L}^n_{q-}\} \\ &\cong \{\psi_- \in \dot{L}^n_{q-} :\ \psi_- + G^T\psi_+ = 0, \quad \text{for some } \psi_+ \in L^n_{q+}\}\,. \end{aligned} \tag{3.3}$$

Indeed, if $G^T\psi_+ = 0$, then $\psi_+ = 0$. Hence the map $(\psi_+,\psi_-) \to \psi_-$ is a bijection from the first space in (3.3) to the third one. Plainly, the map $(\psi_+,\psi_-) \to \psi_-$ is a bijection from the first space in (3.3) to the second one. If G takes singular values on Γ, the same $\psi_- \in \dot{L}_{q-}$ may occur in several (in fact, infinitely many) elements in the kernel of the dual problem. Thus, the second congruence in (3.3) does not have to be valid. More precisely,

$$\begin{aligned} \{(\psi_+,\psi_-) \in L^m_{q+} \times \dot{L}^n_{q-} :\ & \psi_- + G^T\psi_+ = 0\} \\ &\cong \{\psi_+ \in L^m_{q+} :\ G^T\psi_+ \in \dot{L}^n_{q-}\} \\ &\cong \{\psi_- \in \dot{L}^n_{q-} :\ \psi_- + G^T\psi_+ = 0 \ \text{ for some } \psi_+ \in L^m_{q+}\} \\ &\quad \dot{+}\ \{\psi_+ \in L^m_{q+} :\ G^T\psi_+ = 0\}\,. \end{aligned}$$

The space on the right hand side of the preceding direct sum represents the generic kernel of G^T. The dimension of the space on the left hand side of this direct sum can be finite when the generic kernel of G^T is infinite dimensional. Similarly,

$$\begin{aligned} \{(\phi_+,\phi_-) \in L^m_{p+} \times \dot{L}^n_{p-} :\ & \phi_+ + G\psi_- = 0\} \\ &\cong \{\phi_- \in \dot{L}^n_{p-} :\ G\phi_- \in L^m_{p+}\} \\ &\cong \{\phi_+ \in L^m_{p+} :\ \phi_+ + G\phi_- = 0 \ \text{ for some } \phi_- \in \dot{L}^n_{p-}\} \\ &\quad \dot{+}\ \{\phi_- \in \dot{L}^n_{p-} :\ G\phi_- = 0\}\,. \end{aligned}$$

The direct summand on the right hand side of the last congruence can be finite dimensional although $\ker G$ is generically infinite dimensional.

Definition 3.3. *The defect numbers of a Riemann problem with coefficient G are the dimension α_R of the space of "+" components of elements in the kernel of the problem, and the co-dimension β_R of*

$$\{\psi_+ \in L^m_{q+} :\ G^T\psi_+ = 0\} \tag{3.4}$$

in the annihilator of its image. If α_R *or* β_R *is finite, the difference* $\alpha_R - \beta_R$ *is called the index of the problem. The defect numbers of the dual problem are the dimension* α_D *of the space of "* $-$ *" components of elements in the kernel of the dual problem, and the co-dimension* β_D *of*

$$\{\phi_- \in \dot{L}^n_{p-} : \ G\phi_- = 0\} \tag{3.5}$$

in the annihilator of the image of the dual problem. If α_D *or* β_D *is finite, the difference* $\alpha_D - \beta_D$ *is called the index of the dual problem.*

Note that if G takes nonsingular values almost everywhere on Γ, the spaces (3.4) and (3.5) are trivial and Definition 3.3 is equivalent to the usual definition of defect numbers. Also note that (3.4) and (3.5) are closed subspaces of L^m_q and $\dot{L}^n_p$. To see that (3.5) is closed, suppose $\phi \in L^n_p$ is such that $G\phi \neq 0$. Without loss of generality assume that G consists of a single row. Let $G^\dagger(t) = G(t)^*$ if $G(t) = 0$, and let

$$G^\dagger(t) = \frac{1}{G(t)G(t)^*}\, G(t)^*$$

otherwise. Then $G^\dagger$ is a measurable matrix function whose values are Moore-Penrose inverses of the values of G. We have

$$\phi = G^\dagger G\phi + (I - G^\dagger G)\phi =: \phi_1 + \phi_2$$

and $\|\phi_1\|_p > 0$. For any $\tilde{\phi} \in L^n_p$ such that $G\tilde{\phi} = 0$,

$$\|\phi - \tilde{\phi}\|_p = \|\phi_1 + (\phi_2 - \tilde{\phi})\|_p \geq \|\phi_1\|_p \ ,$$

and it follows that $\{\phi \in L^n_p : \ G\phi = 0\}$ is a closed subspace of L^n_p. Hence (3.5), the intersection of this space and $\dot{L}^n_{p-}$, is closed. The space (3.4) is closed by a similar argument.

The defect numbers of a Riemann problem and its dual are related as follows.

Proposition 3.4. *If* α_R, β_R, α_D, *and* β_D *are the defect numbers of a Riemann problem and its dual, then*

$$\alpha_R \leq \beta_D \qquad \textit{and} \qquad \alpha_D \leq \beta_R \ . \tag{3.6}$$

Also, inequalities (3.6) *are equalities if the indices of the problem and its dual are finite and opposite or if* $G \in L^{m\times n}_\infty$.

Proof. The space of " + " components of elements in the kernel of the Riemann problem is isomorphic to the quotient space of " − " components of elements in the kernel of the problem modulo $\{\,\phi_- \in \dot{L}^n_{p-} : G\phi_- = 0\,\}$. Hence, by Proposition 3.2, $\alpha_R \leq \beta_D$ with equality if $G \in L^{m\times n}_\infty$. Similarly, by Proposition 3.1, $\alpha_D \leq \beta_R$ with equality if $G \in L^{m\times n}_\infty$.

Suppose the indices of the problem and its dual are finite and opposite. Then

$$\alpha_R - \beta_D = \beta_R - \alpha_D\ .$$

Since by (3.6) $\alpha_R - \beta_D \leq 0$ and $\beta_R - \alpha_D \geq 0$, it follows that $\alpha_R = \beta_D$ and $\alpha_D = \beta_R$.

We discuss now the homogeneous Riemann problem in the case where the coefficient G admits a spectral factorization in L_p relative to Γ.

Proposition 3.5. *Suppose $G_+\Lambda G_-$ is a spectral factorization in L_p relative to Γ of the coefficient G of a Riemann problem, let $G^L_+ \in L^{k\times m}_{q+}$ be a left multiplicative inverse of G_+, and let $G^R_- \in L^{n\times k}_{p-}$ be a right multiplicative inverse of G_-. Then*

i) *(ϕ_+,ϕ_-) is a solution of the homogeneous problem $\phi_+ + G\phi_- = 0$ if and only if*

$$\phi_+ = G_+\rho_+ \qquad \textit{and} \qquad \phi_- = r_- - G^R_-\Lambda^{-1}\rho_+\ , \tag{3.7}$$

where ρ_+ is a vector function with j^{th}-entry a polynomial of degree at most $\kappa_j - 1$ if $\kappa_j > 0$ and zero if $\kappa_j \leq 0$, and $r_- \in \dot{L}^n_{p-}$ is such that $Gr_- = 0$,

ii) *(ψ_+,ψ_-) is a solution of the homogeneous dual problem $\psi_- + G^T\psi_+ = 0$ if and only if*

$$\psi_- = G^T_-\rho_- \qquad \textit{and} \qquad \psi_+ = r_+ - (G^L_+)^T\Lambda^{-1}\rho_-\ ,$$

where ρ_- is a vector function with j^{th} entry zero if $\kappa_j \geq 0$ and a polynomial in z^{-1} of degree at most $-\kappa_j$ which vanishes at infinity if $\kappa_j < 0$, and $r_+ \in L^m_{q+}$ is such that $G^T r_+ = 0$.

Proof. We verify assertion i). Suppose (ϕ_+,ϕ_-) is a solution of the homogeneous problem. Then

$$G^L_+\phi_+ = -\Lambda G_-\phi_- =: \rho_+\ . \tag{3.8}$$

Comparing both sides of the equality (3.8) we find out that ρ_+ is a vector polynomial satisfying the degree requirements. We have

$$G_+ G_+^L \phi_+ = G_+ \rho_+ \ .$$

Since $\phi_+ \in \operatorname{im} G_+$ almost everywhere on Γ, $G_+ G_+^L \phi_+ = \phi_+$ and the first equality in (3.7) holds. By (3.8),

$$G_- \phi_- = -\Lambda^{-1} \rho_+ \ . \tag{3.9}$$

Since $-G_-^R \Lambda^{-1} \rho_+$ is a solution of equation $G_- x = -\Lambda^{-1}\rho_+$ in $\dot{L}_{p-}^n$, $r_- := \phi_- + G_-^R \Lambda^{-1} \rho_+ \in \dot{L}_{p-}^n$ is such that $G r_- = 0$. Thus, the second equality in (3.7) holds.

Conversely, suppose ϕ_+ and ϕ_- satisfy (3.7) with appropriate r_- and ρ_+. Then

$$\phi_+ + G\phi_- = G_+\rho_+ + G r_- - G_+ \rho_+ = 0\,,$$

and (ϕ_+, ϕ_-) is a solution of the homogeneous problem.

It follows from Proposition 3.5 that if the coefficient G in a Riemann problem admits a spectral factorization in L_p relative to Γ, then α_R equals the sum of positive indices of the factorization, and α_D equals the absolute value of the sum of negative indices of the factorization. In fact, a stronger statement is true.

Theorem 3.6. *Suppose the coefficient G in a Riemann problem admits a spectral factorization in L_p relative to Γ with indices $\kappa_1, \kappa_2, \ldots, \kappa_k$. Then*

$$\alpha_R = \beta_D = \sum \{\kappa_i : \ \kappa_i > 0\}$$

and

$$\alpha_D = \beta_R = \sum \{-\kappa_i : \ \kappa_i < 0\}\,.$$

PROOF. We show that β_D is the sum of the positive indices; the argument regarding β_R is similar. Since $\dot{L}_{q-}^n$ is contained in the image of the dual problem, the annihilator of the image of the dual problem is a subspace of $\dot{L}_{p-}^n$. Let $G = G_+ \Lambda G_-$ with Λ as in (2.3) be a spectral factorization in L_p relative to Γ, let j be such that $\kappa_j > 0 \geq \kappa_{j+1}$, and let $G_1, G_2, \ldots, G_j$ be the first j columns of $G_-^R \in L_{p-}^{n\times k}$. We show that the elements of the set

$$\{t^{-i}G_l(t) :\ 1 \le l \le j,\ 1 \le i \le \kappa_l\} \tag{3.10}$$

form a basis for a space which complements the space (3.5) in the annihilator of the image of the dual problem. Since $G_-^R(\infty)$ has linearly independent columns, the elements of the set (3.10) are linearly independent. Using the factorization $G = G_+\Lambda G_-$, we can rewrite the space (3.5) as

$$\{\, \phi_- \in \dot{L}_{p-}^n :\ G_-\phi_- = 0 \,\}\,. \tag{3.11}$$

Since G_l's are the columns of a right multiplicative inverse of G_-, the span of the set (3.10) intersects trivially with the space (3.11). Now members of the set (3.10) annihilate $\dot{L}_{q-}^n$ and

$$t^{-i}G(t)\,G_l(t) \in L_{p+}^n, \qquad 1 \le l \le j, \quad 1 \le i \le \kappa_l\ .$$

Hence the members of the set (3.10) annihilate the image of the dual problem. Finally, consider an arbitrary $\phi_- \in \dot{L}_{p-}^n$ that annihilates the image of the dual problem. Choose f_- in the linear span of (3.10) such that $\Lambda G_-(\phi_- - f_-)(\infty) = (0)$ and let $\hat{\phi}_- = \phi_- - f_-$. Then $\hat{\phi}_- \in \dot{L}_{p-}^n$ and

$$\int_\Gamma \hat{\phi}_-(t)^T G_-(t)^T \Lambda(t)\, G_+(t)^T \psi_+(t)\, dt = 0 \tag{3.12}$$

for all $\psi_+ \in L_{q+}^m$ such that $G_-^T \Lambda G_+^T \psi_+ \in L_q^n$. In particular, (3.12) holds whenever $\psi_+ = (G_+^L)^T p$ with $G_+^L \in L_{q+}^{k\times m}$ a left multiplicative inverse of G_+ and p a vector polynomial. Hence

$$\int_\Gamma \big(\Lambda(t)\, G_-(t)\, \hat{\phi}_-(t)\big)^T p(t)\, dt = 0$$

for each vector polynomial p and $\Lambda G_-\hat{\phi} \in L_{1+}^k$. Since $\Lambda G_-\hat{\phi}_- \in \dot{L}_{1-}^k$, it follows that $\Lambda G_-\hat{\phi}_- = 0$ and $\phi_- = f_- + \hat{\phi}_-$ where f_- is in the span of (3.10) and $\hat{\phi}_-$ is a member of the space (3.11).

Corollary 3.7. *If the coefficient G of a Riemann problem admits a spectral factorization in L_p relative to Γ, then the index of the problem,* *and the opposite of the index of the dual problem, are both equal to the total index of the factorization.*

In particular, if G admits a spectral factorization in L_p, the indices of the Riemann problem and its dual are finite and opposite.

4. Condition for existence of a spectral factorization.

We will need below the following lemma. If G is a meromorphic matrix function defined on a connected domain $\mathcal{D}$, its rank is constant at all but a countable number of points in $\mathcal{D}$. This rank is usually called the *normal rank* of G.

Lemma 4.1. *Suppose* Γ *is a simple closed curve which forms a boundary of a connected domain* $\mathcal{D}_+$*, let* $p > 0$*, and suppose* $G \in L_p^{m\times n}$ *is formed by nontangential boundary values of a matrix function* G_+ *meromorphic in* $\mathcal{D}_+$ *with normal rank* k*. Then* $\operatorname{rank} G = k$ *almost everywhere on* Γ*.*

Proof. If $k < \min\{m,n\}$, let $H(t)$ be any $(k+1)\times(k+1)$ submatrix of $G(t)$ and form H_+ from the corresponding entries of G_+. Then $\det H_+ \equiv 0$ implies $\det H(t) = 0$ almost everywhere on Γ. Thus, $\operatorname{rank} G(t) \le k$ for almost everywhere $t \in \Gamma$.

Choose a point $z_+ \in \mathcal{D}_+$ such that $\operatorname{rank} G_+(z_+) = k$, and pick matrices $A \in \mathbb{C}^{k\times m}$ and $B \in \mathbb{C}^{n\times k}$ such that $\operatorname{rank}(AG_+(z_+)B) = k$. Then $AG_+(z)B$ is a meromorphic $k\times k$ matrix function and $\det(AG_+(z)B) \not\equiv 0$. Hence $\det(AG(t)B) \ne 0$ and consequently $\operatorname{rank} G(t) \ge k$ almost everywhere on Γ. Thus, $\operatorname{rank} G = k$ almost everywhere on Γ.

One can formulate the following necessary and sufficient condition for existence of a canonical spectral factorization in L_p of a function G relative to Γ (*cf.* [14, Theorem 3.2] and [8]). Recall that if G admits a spectral factorization relative to Γ, then the rank of G is constant almost everywhere on Γ.

Theorem 4.2. *If* $G \in L_1^{m\times n}$ *with* $\operatorname{rank} G = k$ *almost everywhere on* Γ*, the following are equivalent:*

i) *there exist collections of linearly independent constant vectors* $\{a_1, a_2, \ldots, a_k\}$ *and* $\{b_1, b_2, \ldots, b_k\}$ *such that the image of the Riemann* *problem with coefficient G contains $\{t^{-1}a_1, t^{-1}a_2, \dots, t^{-1}a_k\}$ and the image of the dual problem contains $\{b_1, b_2, \dots, b_k\}$.*

ii) *the function G admits a canonical spectral factorization in L_p relative to Γ.*

Moreover, if the equivalent conditions i) *and* ii) *are satisfied, the image of either of the problems contains all rational vector functions in its closure.*

PROOF. Suppose first i) holds. Pick $\phi_{j+} \in E^m_{p+}$ and $\phi_{j-} \in \dot{E}^n_{p-}$ such that

$$\phi_{j+}(t) + G(t)\,\phi_{j-}(t) = t^{-1}a_j\,, \qquad j = 1, 2, \dots, k\,, \tag{4.1}$$

and let $\Phi_- = (\phi_{1-}\ \phi_{2-}\ \dots\ \phi_{k-})$. Then $F(t) := t\,G(t)\Phi_-(t) \in L^{m\times k}_{p+}$ and $F(0) = (a_1\ a_2\ \dots\ a_k)$. Similarly, pick $\psi_{j+} \in E^m_{q+}$ and $\psi_- \in \dot{E}^n_{q-}$ such that

$$\psi_{j-}(t) + G^T(t)\,\psi_{j+}(t) = b_j\,, \qquad j = 1, 2, \dots, k\,, \tag{4.2}$$

and let $\Psi_+ = (\psi_{1+}\ \psi_{2+}\ \dots\ \psi_{k+})$. Then $H = G^T\Psi_+ \in E^{n\times k}_{q-}$ and

$$H(\infty) = (b_1\ b_2\ \dots\ b_k)\,.$$

Let $S(t) = t\Psi^T_+(t)G(t)\Phi_-(t)$. Since

$$S(t) = \Psi^T_+(t)F(t) = H^T(t)(t\Phi_-(t)), \tag{4.3}$$

$S(t) \in L^{k\times k}_{1+} \cap L^{k\times k}_{1-}$. Thus, $S(t) = S$ is a constant. Also, $\det S \neq 0$. Indeed, by Lemma 4.1, $F(t)$ has linearly independent columns for almost everywhere $t \in \Gamma$. Since $\operatorname{rank} G = k$ almost everywhere on Γ, the column spans of F and G are equal almost everywhere on Γ. Thus, to prove that S is nonsingular it suffices to show $\operatorname{rank}(\Psi^T_+G) = k$ almost everywhere on Γ. But this follows from Lemma 4.1 and the fact that

$$(G^T\Psi_+)(\infty) = H(\infty) = (b_1\ b_2\ \dots\ b_k)\,.$$

Let

$$G_+(t) = F(t)\,, \qquad G^L_+(t) = S^{-1}\Psi^T_+(t)\,,$$
$$G_-(t) = S^{-1}H^T(t)\,, \qquad G^R_-(t) = t\,\Phi_-(t)\,.$$

Then $G_+ \in L^{m\times k}_{p+}$, $G^L_+ \in L^{k\times m}_{q+}$, $G_- \in L^{k\times n}_{q-}$, and $G^R_- \in L^{n\times k}_{p-}$. By (4.3),

$$G^L_+(t)\,G_+(t) = I \qquad \text{and} \qquad G_-(t)\,G^R_-(t) = I\,.$$

By (4.3) and the definition of F,

$$G_+^L(t)\,G(t)\,G_-^R(t) = I$$

almost everywhere on Γ. Hence

$$G_-^R G_+^L G\, G_-^R G_+^L = G_-^R G_+^L\,,$$

or $G^\times G\, G^\times = G^\times$ where $G^\times = G_-^R G_+^L$. Since $\operatorname{rank} G^\times = \operatorname{rank} G$ almost everywhere on Γ, $G\,G^\times G = G$ (see [1, Theorem 1.5.2]; *cf.* [10, Lemma 3.8]). Thus,

$$G(t) = G(t)\,t\,\Phi_-(t)\,S^{-1}\Psi_+^T(t)\,G(t) = G_+(t)\,G_-(t)$$

almost everywhere on Γ and it follows that G admits a canonical spectral factorization in L_p relative to Γ.

Conversely, suppose ii) holds and let $G = G_+G_-$ be a canonical factorization. Let $G_-^R \in L_{p-}^{n\times k}$ be a right multiplicative inverse of G_-. Then $t^{-1}G_-^R(t) \in \dot{L}_{p-}^{n\times k}$, and

$$G(t)\,(t^{-1}G_-^R(t)) = t^{-1}G_+(t) = t^{-1}G_+(0) + t^{-1}(G_+(t) - G_+(0))\,.$$

Hence the columns of $t^{-1}G_+(0)$ are in the image of the problem. Similarly, if $G_+^L \in L_{q+}^{m\times k}$ is a left multiplicative inverse of G_+, $G^T(G_+^L)^T = G_-^T$ and so the columns of $G_-^T(\infty)$ are in the image of the problem. Thus, ii) implies i) and the conditions are equivalent.

The argument from the last paragraph can be used in a more general situation. Suppose G_+G_- is a canonical spectral factorization in L_p relative to Γ. Let $G_+^L \in L_{q+}^{k\times m}$ and $G_-^R \in L_{p-}^{n\times k}$ be one-sided multiplicative inverses of G_+ and G_-, and let $r \in \dot{L}_{\infty-}^k$ be a rational vector function. Then $G_-^R r \in \dot{L}_{p-}^{n\times k}$, and

$$G(G_-^R r) = G_+ r$$

differs from a rational vector function by an element in L_{p+}^m. Hence $\mathcal{Q}(G_+r)$, where $\mathcal{Q}$ is a canonical projection of $L_{p+}^m \dot{+} \dot{L}_{p-}^m$ onto $\dot{L}_{p-}^m$, is a rational vector function in the image of the problem. We claim that any rational vector function in the intersection of $\dot{L}_{p-}^m$ and the closure of the image of the problem arises in this way. Indeed, let $f_- \in \dot{L}_{\infty-}^m$ be a rational vector function such that

$$f_- \notin \{\mathcal{Q}(G_+r) :\ r \in \dot{L}_{\infty-}^k \text{ is a rational vector function}\}\,. \tag{4.4}$$

We may assume f_- has a single pole, located at $\lambda \in \mathcal{D}_+$. Suppose the leading coefficient in the Laurent expansion of f_- at λ is contained in the image of $G_+(\lambda)$. Then after subtracting from f_- an element in the set on the right hand side of (4.4), we obtain a strictly proper rational vector function analytic in $\mathbb{C}\backslash\{\lambda\}$ with the pole at λ of smaller order. By induction, there exists a strictly proper rational vector function with the only pole at λ whose leading coefficient in the Laurent expansion at λ is not contained in the image of $G_+(\lambda)$. Call this function again f_-.

Consider a problem

$$\phi_+ + G\phi_- = g\,, \tag{4.5}$$

where $\phi_- \in \dot{L}^n_{p-}$ is such that $G\phi_- \in L^m_p$ and $\phi_+ \in L^m_{\infty+}$. The image of the problem (4.5) is contained in the image of the Riemann problem. Since rational functions without poles on Γ are dense in L_p, and the projection $\mathcal{P}$ is bounded on $L_{1+}\dot{+}\dot{L}_{1-}$, $L_{\infty+}$ is dense in L_{p+}. Hence the closures of the images of both problems coincide. Now

$$(I - G_+(t)\,G^L_+(t))\,G(t) = \big(G_+(t) - G_+(t)\,G^L_+(t)\,G_+(t)\big)\,G_-(t) = 0$$

almost everywhere on Γ and, since $I - G_+(\lambda)G^L_+(\lambda)$ is an $m\times m$ matrix of rank $m-k$ whose null space coincides with the image of $G_+(\lambda)$,

$$(I - G_+(z)\,G^L_+(z))f_-(z)$$

has a pole at $z=\lambda$. Consequently, there exists a function $\psi_+ \in L^{1\times m}_{1+}$ such that $\psi^T_+ f_-$ has a simple pole at λ and $\int_\Gamma \psi^T_+ g$ equals zero for all functions g in the image of the problem (4.5). Let X be a subspace of L^m_p spanned by f_- and the image of the problem (4.5). Then

$$x \longrightarrow \int_\Gamma \psi_+(t)^T x(t)\,dt$$

is a continuous linear functional on the space X whose kernel contains the image of the problem (4.5) and which has nonzero value at f_-. By the Hahn-Banach Theorem, there exists a continuous linear functional Ψ on L^m_p which annihilates the image of the problem (4.5) and such that $\Psi(f_-) \neq 0$. Hence f_- is not in the closure of the image of the problem (4.5).

In order to obtain a condition for existence of a spectral factorization of a function G in a non-canonical case, we will need the following lemma.

Lemma 4.3. *Suppose the defect numbers α_R and β_D of the Riemann problem with coefficient G and its dual are finite and positive. Then there exists a square rational matrix function H with a nonzero determinant and without poles or zeros on Γ such that the Riemann problem with coefficient GH and its dual have the corresponding defect numbers smaller by 1. Moreover, the Riemann problem with coefficient G (respectively its dual) contains all rational vectors functions in its closure if and only if the image of the Riemann problem with coefficient GH (respectively its dual) contains all rational vector functions in its closure.*

Proof. Pick $(\varphi_+, \varphi_-) \in L^m_{p+} \dot{+} \dot{L}^n_{p-}$ such that $\varphi_+ \neq 0$ and

$$\varphi_+ + G\varphi_- = 0\,.$$

Then $\varphi_- \notin \{\phi \in \dot{L}^n_{p-} :\ G\phi = 0\}$ and there exists a point $z_0 \in \mathcal{D}_-$ such that $\varphi_-(z_0)$ is not a member of

$$\operatorname{span}\{\phi_-(z_0) :\ \phi_- \in \dot{L}^n_{p-} \text{ and } G\phi_- = 0\}\,. \tag{4.6}$$

After adding to φ a linear combination of functions in $\{\phi_- \in \dot{L}^n_{p-} : G\phi_- = 0\}$, and multiplying G on the right by a nonsingular constant matrix, we may assume $\varphi_-(z_0) = e_1$ and

$$\operatorname{span}\{\phi_-(z_0) :\ \phi_- \in \dot{L}^n_{p-} \text{ and } G\phi_- = 0\} \subset \operatorname{span}\{e_2, e_3, \dots, e_n\}\,.$$

As usual, we assume $0 \in \mathcal{D}_+$. Let

$$H(z) = \begin{pmatrix} \frac{z - z_0}{z} & & & \mathbf{0} \\ & 1 & & \\ & & \ddots & \\ \mathbf{0} & & & 1 \end{pmatrix}.$$

We show that the space of "+" components of the members of the kernel of the problem

$$\phi_+ + GH\phi_- = g \tag{4.7}$$

has dimension one less than the corresponding number for the problem with coefficient G. First, note that φ_+ is not a "+" component of a member of the kernel of problem (4.7). Indeed, suppose there exists $\phi_- \in \dot{L}^n_{p-}$ such that $\varphi_+ + GH\phi_- = 0$, and let $f_- = H\phi_- - \varphi_-$. Then $f_- \in \dot{L}^n_{p-}$, $Gf_- = 0$, and

$$f_-(z_0) \notin \operatorname{span}\{e_2, \ldots, e_n\},$$

a contradiction. Secondly, suppose (ϕ_+, ϕ_-) is in the kernel of the Riemann problem with coefficient G. If $\phi_-(z_0) = (0, *, \ldots, *)$, the element $(\phi_+, H^{-1}\phi_-)$ is in the kernel of the problem (4.7). If $\phi_-(z_0) = (\lambda, *, \ldots, *)$ with $\lambda \neq 0$,

$$\left(\varphi_+ - \frac{1}{\lambda}\phi_+, H^{-1}\left(\varphi_- - \frac{1}{\lambda}\phi_-\right)\right)$$

is contained in the kernel of the problem (4.7). Thus, each "+" component of a member of the kernel of the Riemann problem with coefficient G is a linear combination of φ_+ and a "+" component of a member of the kernel of the problem (4.7). Finally, if (ϕ_+, ϕ_-) belongs to the kernel of the problem (4.7), $(\phi_+, H\phi_-)$ satisfies the homogeneous Riemann problem.

Consider now the problem dual to (4.7),

$$\psi_- + (GH)^T\psi_+ = h. \tag{4.8}$$

After multiplying both sides of (4.8) by H^{-1}, we obtain a new problem

$$H^{-1}\psi_- + G^T\psi_+ = h, \qquad \psi_- \in \dot{L}^n_{q-},\ \psi_+ \in L^m_{q+},\ \text{and } h \in L^n_q. \tag{4.9}$$

Let $\mathcal{W}$ be the image of the problem dual to the Riemann problem with coefficient G. Then the image of the problem (4.9) equals $\mathcal{W} + \operatorname{span}(z - z_0)^{-1}e_1$. Since

$$\int_\Gamma \varphi_-(z)^T(z - z_0)^{-1}e_1\, dz = -2\pi i,$$

by Proposition 3.2 $(z - z_0)^{-1}e_1 \notin \operatorname{cl}\mathcal{W}$. We have $\operatorname{cl}(\mathcal{W} + \operatorname{span}\{(z - z_0)e_1\} = \operatorname{cl}\mathcal{W} + \operatorname{span}\{(z - z_0)e_1\}$. Since multiplication by H is an isomorphism $L^n_p \to L^n_p$, it follows that the closure of the image of problem (4.8) equals

$$H(\operatorname{cl}\mathcal{W}) + \operatorname{span}\{H(z)(z - z_0)^{-1}e_1\} = H(\operatorname{cl}\mathcal{W}) + \operatorname{span}\left\{\frac{1}{z}e_1\right\}. \tag{4.10}$$

Now the space $\{\phi_- \in \dot{L}^n_{p-} : \ G\phi_- = 0\}$ has a finite co-dimension β_D in the annihilator of $\mathcal{W}$. Hence the co-dimension of the space

$$\{H^{-1}\phi_- : \ \phi_- \in \dot{L}^n_{p-} \text{ and } G\phi_- = 0\} \tag{4.11}$$

in the annihilator of $H(\operatorname{cl}\mathcal{W})$ equals β_D. Consequently, the co-dimension of the space (4.11) in the annihilator of (4.10) equals $\beta_D - 1$. Since

$$\{\phi_- \in \dot{L}^n_{p-} : \ GH\phi_- = 0\} = \{H^{-1}\phi_- : \ \phi_- \in \dot{L}^n_{p-} \text{ and } G\phi_- = 0\}\,,$$

the co-dimension of the closure of $\{\phi_- \in \dot{L}^n_{p-} : \ GH\phi_- = 0\}$ in the annihilator of the space (4.10) equals $\beta_D - 1$.

It remains to verify the assertion about the images. First, note that the images of the Riemann problems with coefficients G and GH coincide. Indeed, since $H\dot{L}^n_{p-} \subset \dot{L}^n_{p-}$, the image of the problem with coefficient GH is contained in the image of the problem with coefficient G. Since

$$\phi_+ + G\phi_- = \phi_+ - \lambda\varphi_+ + GH(H^{-1}(\phi_- - \lambda\varphi_-))$$

for any scalar λ, and for each $\phi_- \in \dot{L}^n_p$ there exists λ such that $H^{-1}(\phi_- - \lambda\varphi_-) \in \dot{L}_{p-}$, the image of the problem with coefficient G is contained in the image of the problem with coefficient GH.

Suppose the image of the problem dual to the Riemann problem with coefficient G contains all rational vector functions in its closure, and let a rational vector function f be a member of the set (4.10). Then $H^{-1}(f(z) - z^{-1}e_1) \in \operatorname{cl}\mathcal{W}$, so $H^{-1}(f(z) - z^{-1}e_1) \in \mathcal{W}$ and

$$f \in H(\mathcal{W}) + \operatorname{span}\{H(z)(z-z_0)^{-1}e_1\}\,.$$

Thus, f is a member of the image of problem (4.8). Conversely, suppose the image of the problem (4.8) contains all rational vector functions in its closure, and let $f \in \operatorname{cl}\mathcal{W}$ be a rational vector function. Then $Hf \in H(\operatorname{cl}\mathcal{W}) \subset H(\operatorname{cl}\mathcal{W} + \operatorname{span}\{H(z)(z-z_0)^{-1}e_1\}$, so $Hf \in H\mathcal{W} + \operatorname{span}\{H(z)(z-z_0)^{-1}e_1\}$. Thus, $f \in \mathcal{W} + \operatorname{span}\{(z-z_0)^{-1}e_1\}$. Since $(z-z_0)^{-1}e_1 \notin \operatorname{cl}\mathcal{W}$, $f \in \mathcal{W}$.

In a similar way one can show the following dual version of Lemma 4.3. We omit the details of the proof.

Lemma 4.4. *Suppose the defect numbers α_D and β_R of the Riemann problem with coefficient G and its dual are finite and positive. Then there exists a square rational matrix function F with a nonzero determinant and without poles or zeros on Γ such that the Riemann problem with coefficient FG and its dual have the corresponding defect numbers smaller by 1. Moreover, the image of the Riemann problem with coefficient G (respectively its dual) contains all rational vector functions in its closure if and only if the image of the problem with coefficient FG (respectively its dual) contains all rational vector functions in its closure.*

We can give now a necessary and sufficient condition for existence of a spectral factorization in L_p of a summable singular matrix valued function (*cf.* [13, Theorem 3.1]).

Theorem 4.5. *If $G \in L_1^{m\times n}$ and $\operatorname{rank} G = k$ almost everywhere on Γ, the following are equivalent:*

i) *the indices of the Riemann problem with coefficient G and its dual are finite and opposite, and the image of each of the problems contains all rational vector functions in its closure,*

ii) *G admits a spectral factorization in L_p relative to Γ.*

Proof. Suppose first i) holds. By Proposition 3.4, $\alpha_R = \beta_D$ and $\alpha_D = \beta_R$. Applying Lemmas 4.3 and 4.4 a finite number of times, we can find regular rational matrix functions F and H without poles or zeros on Γ such that

1) the annihilator of the image of the Riemann problem with coefficient $\widehat{G} = FGH$ coincides with $\{\psi_+ \in L_{q+}^m :\ \widehat{G}^T\psi_+ = 0\}$,

2) the annihilator of the image of the dual problem equals $\{\phi_- \in \dot{L}_{p-}^n :\ \widehat{G}\phi_- = 0\}$,

3) the image of either of the problems contains all rational vector functions in its closure.

Let

$$\Omega_+ = \operatorname{span}\{\psi_+(0) :\ \psi_+ \in L_{q+}^m \text{ and } \widehat{G}^T\psi_+ = 0\}.$$

Since $\operatorname{rank}\widehat{G} = k$ almost everywhere on Γ, by Lemma 4.1 dim $\Omega_+ \le m-k$. Hence there exist linearly independent vectors $\{a_1, a_2, \dots, a_k\}$ such that $\omega a_i = 0$ whenever $\omega \in \Omega_+$ and $i = 1, 2, \dots, k$. Suppose

$\psi_+ \in L^m_{q+}$ and $\widehat{G}^T\psi_+ = 0$. Then $\psi_+(t)^T t^{-1} a_i \in L_{q+}$, and so

$$\int_\Gamma \psi_+(t)^T t^{-1} a_i \, dt = 0\,, \qquad\qquad \text{for} \quad i = 1, 2, \ldots, k\,.$$

It follows that the set

$$\left\{ \frac{1}{t}\, a_1, \frac{1}{t}\, a_2, \ldots, \frac{1}{t}\, a_k \right\}$$

is in the closure of the image, and hence in the image, of the Riemann problem with coefficient $\widehat{G}$.

Similarly, let

$$\Omega_- = \operatorname{span}\{\phi_-(\infty) :\ \phi_- \in L^n_{p-} \text{ and } \widehat{G}\phi_- = 0\}$$

and pick linearly independent vectors $\{b_1, b_2, \ldots, b_k\}$ such that $\omega_- b_j = 0$, for $j = 1, 2, \ldots, k$ and all $\omega_- \in \Omega_-$. Suppose $\phi_- \in \dot{L}^n_{p-}$ is such that $\widehat{G}\phi_- = 0$. Then $z\phi_-(z) b_j \in \dot{L}_{p-}$ and hence

$$\int_\Gamma \phi_-(t) b_j \, dt = \int_\Gamma (t\phi_-(t) b_j)\, t^{-1} dt = 0\,,$$

for $j = 1, 2, \ldots, k$. Thus, the set $\{b_1, b_2, \ldots, b_k\}$ is contained in the closure of the image, and consequently in the image, of the problem dual to the Riemann problem with coefficient $\widehat{G}$. Consequently, by Theorem 4.2, the function $\widehat{G} = FGH$ admits a canonical spectral factorization in L_p relative to Γ. Hence, by Proposition 2.6 the function G admits a spectral factorization in L_p relative to Γ.

Conversely, suppose ii) holds. By Theorem 3.6, the indices of the problem and its dual are finite and opposite. Applying Lemmas 4.3 and 4.4 a finite number of times, we can find square rational matrix functions F and H whose determinants are not equal to zero identically and which have neither poles nor zeros on Γ such that the Riemann problem with coefficient FGH and the dual problem have defect numbers

$$\alpha_R = \beta_R = \alpha_D = \beta_D = 0\,.$$

By Proposition 2.6 and Theorem 3.6, the function FGH admits a canonical spectral factorization in L_p relative to Γ. By Theorem 4.2, the image of the Riemann problem with coefficient FGH and the image of the dual problem each contain all rational vector functions in their closures. By Lemmas 4.3 and 4.4, the image of the Riemann problem with coefficient G (respectively image of the dual problem) contains all rational vector functions in its closure.

We note that the part of condition i) in Theorem 4.5 involving rational vector functions cannot be in general omitted. Indeed, suppose Γ is the unit circle, $p = 3$, and let

$$G(t) = \begin{pmatrix} t^{2/3} \\ t^{5/3} \end{pmatrix}$$

be as in Example 2.7. Since G admits a spectral factorization in L_p for p in a deleted neighborhood of 3, by Theorem 3.6 the numbers α_R and α_D are finite when the problem is considered in L_{p_1} or L_{p_2} with $p_1 < 3 < p_2$. Since $L_3 \subset L_{p_1}$ and $L_{3/2} \subset L_{p_2/(p_2-1)}$, α_R and α_D are finite when $p = 3$. Since $G \in L_\infty$, by Proposition 3.4 $\alpha_R = \beta_D$ and $\alpha_D = \beta_R$. Thus, the indices of the problem and its dual are finite and opposite although G does not admit a spectral factorization in L_3 relative to the circle.

References.

[1] Ben-Israel, A., Greville, T. N. E., *Generalized Inverses: Theory and Applications.* John Wiley & Sons, 1974.

[2] Clancey, K., Gohberg, I., *Factorization of Matrix Functions and Singular Integral Operators.* OT **3**, Birkhäuser Verlag, 1981.

[3] Clancey, K., Rakowski, M., Factorization of rectangular matrix functions relative to a contour. Manuscript, 1990.

[4] David, G., Opérateurs intégraux singuliers sur certains courbes du plan complexe. *Ann. Sci. École. Norm. Sup.* **17** (1984), 157-189.

[5] Duren, P., *Theory of H^p spaces.* Academic Press, 1970.

[6] Gohberg, I., Krein, M. G., Systems of integral equations on a half line with kernels depending on the differences of arguments. English translation: *Amer. Math. Soc. Trans. Ser.* **14** (1960), 217-287.

[7] Hilbert, D., *Grundzüge einer allgemeinen Theorie der linearen Intergleichungen*, 2nd edition. Verlag von B. G. Teubner, 1924.

[8] Lax, P. D., On the factorization of matrix-valued functions. *Comm. Pure Appl. Math.* **29** (1976), 683-688.

[9] Litvinchuk, G. S., Spitkovsky, I. M., *Factorization of Measurable Matrix Functions.* English translation: OT **25**, Birkhäuser Verlag, 1987.

[10] Rakowski, M., Generalized pseudoinverses of matrix valued functions. *Integral Equations and Operator Theory* **14** (1991), 564-585.
[11] Rakowski, M., Spectral factorization of rectangular rational matrix functions with application to discrete Wiener-Hopf equations, *J. Funct. Anal.* **10** (1992), 410-433.
[12] Simonenko, I. B., Some general questions in the theory of the Riemann boundary problem. English translation: *Math. USSR - Izvestia* **2** (1968), 1091-1099.
[13] Spitkovsky, I., Factorization of measurable matrix-value functions and its relation to the theory of singular integral equations and the vector Riemann boundary-value problem, I. English translation: *Differential Equations* **17** (1981), 477-485.
[14] Spitkovsky, I., Generalized factorization of matrix-valued functions and the Riemann boundary value problem with infinite partial indices. English translation: *Soviet Math. Dokl.* **33** (1986), 145-149.

Recibido: 2 de junio de 1.995

Marek Rakowski*
Department of Mathematics
The Ohio State University
Columbus, OH 43210
U.S.A.
rakowski@math.ohio-state.edu

and

Ilya Spitkovsky†
Department of Mathematics
The College of William and Mary
Williamsburg, VA 23187-8795
U.S.A.
ilya@math.wm.edu

* The research of the first-named author was partially supported by the National Science Foundation Grant DMS-9302706 and by Consejo Nacional de Ciencia y Tecnologia.

† The research of the second-named author was partially supported by the National Science Foundation Grant DMS-9401848.

Revista Matemática Iberoamericana
Vol. 12, N.° 3, 1996

The boundary absolute continuity of quasiconformal mappings II

Juha Heinonen

Dedicated to Fred Gehring on his seventieth birthday

Abstract. In this paper a quite complete picture is given of the absolute continuity on the boundary of a quasiconformal map $\mathbb{B}^3 \to D$, where $\mathbb{B}^3$ is the unit 3-ball and D is a Jordan domain in $\mathbb{R}^3$ with boundary 2-rectifiable in the sense of geometric measure theory. Moreover, examples are constructed, for each $n \geq 3$, showing that quasiconformal maps from the unit n-ball onto Jordan domains with boundary $(n-1)$-rectifiable need not have absolutely continuous boundary values.

1. Introduction.

Suppose that f is a quasiconformal homeomorphism of the open unit ball $\mathbb{B}^n$ of $\mathbb{R}^n$ onto a bounded domain D in $\mathbb{R}^n$. Then f extends homeomorphically to the boundary $\partial\mathbb{B}^n$ if and only if D is bounded by a topological $(n-1)$-sphere [V1, p. 61]. Should such an extension exist, we denote it by f as well, and call D a *Jordan domain* or a *quasiconformal Jordan ball.* Suppose now that the boundary of D has finite Hausdorff $\mathcal{H}_{n-1}$-measure. We say that f is *absolutely continuous* on the boundary if f carries sets of $\mathcal{H}_{n-1}$-measure zero on $\partial\mathbb{B}^n$ to sets of $\mathcal{H}_{n-1}$-measure zero on ∂D. If $n = 2$ and f is conformal, the boundary correspondence is absolutely continuous according to the classical theorem of F. and M. Riesz [R]; but if f is merely quasiconformal, it is well known, and first observed by Beurling and Ahlfors [BA], that the boundary correspondence need not be absolutely continuous even when f is a self-homeomorphism of a disk. The situation is quite different in higher dimensions. For instance, if f is a quasiconformal self-homeomorphism of $\mathbb{B}^n$, the boundary map is a quasiconformal map of $\mathbb{S}^{n-1} = \partial\mathbb{B}^n$ onto itself, and hence preserves sets of $(n-1)$-measure zero, provided $n - 1 \geq 2$. It is therefore natural to ask what conditions on ∂D are needed in order to have the absolute continuity of the boundary map $f : \partial\mathbb{B}^n \to \partial D$ when $n \geq 3$. For instance, is it sufficient that ∂D be of finite $\mathcal{H}_{n-1}$-measure? In the present paper, which is a sequel to [H], a rather complete solution to this problem will be provided in dimension $n = 3$ in the case when the boundary of D is 2-rectifiable in the sense of geometric measure theory. It will also be shown that a direct analog of the F. and M. Riesz theorem is false for quasiconformal mappings in all dimensions. For the record, we shall only be dealing with the absolute continuity of the map $f : \partial\mathbb{B}^n \to \partial D$. It still remains widely open under what conditions the map $f^{-1} : \partial D \to \partial\mathbb{B}^n$ is absolutely continuous. Further open problems are listed in the end of the paper in Section 6.

Before proceeding, let us review the prior results in this area. So assume that f is a quasiconformal mapping of $\mathbb{B}^n$ onto a Jordan domain D whose boundary has finite $\mathcal{H}_{n-1}$-measure, and assume that $n \geq 3$. Gehring showed in [G2] that the boundary correspondence $f : \partial\mathbb{B}^n \to \partial D$ is absolutely continuous if f has a quasiconformal extension to a neighborhood of $\partial\mathbb{B}^n$. Väisälä [V2] arrived at the same positive conclusion under the less restrictive assumption that f be quasisymmetric on $\overline{\mathbb{B}}^n$. (Recall that quasisymmetry is a global condition as opposed to quasiconformality which is local; see (3.13) below for the definition of quasisymmetry.) In [H] it was shown that the answer is likewise affirmative if $\mathcal{H}_{n-1}$-almost every point on ∂D is a "two sided cone point", and if $n \neq 4$. To make this supposition more precise, we next fix some notation. Let L be a line in $\mathbb{R}^n$ through a point a and let $0 < s < 1$. Set

$$\mathcal{C}(a, L, s) = \{x \in \mathbb{R}^n : \ \mathrm{dist}(x, L) < s\,|a - x|\}\,.$$

The point a divides the line L into two pieces, which we shall call L^+ and L^-. The orientation of the line plays no role in our arguments, so this choice is arbitrary. We write

$$\mathcal{C}(a, L^+, s) = \{x \in \mathbb{R}^n : \ \mathrm{dist}(x, L^+) < s\,|a - x|\}\,,$$

and similarly for $\mathcal{C}(a, L^-, s)$. Thus $\mathcal{C}(a, L, s)$ is the union of the two infinite open cones $\mathcal{C}(a, L^+, s)$ and $\mathcal{C}(a, L^-, s)$ with s determining the angle opening. We also use the notation

$$\mathcal{C}(a, r, L, s) = \mathcal{C}(a, L, s) \cap B(a, r)\,,$$

$$\mathcal{C}(a, r, L^\pm, s) = \mathcal{C}(a, L^\pm, s) \cap B(a, r)\,.$$

Here and throughout $B(z, t)$ will denote the open n-ball which is centered at z and has radius $t > 0$.

We say that a set $E \subset \mathbb{R}^n$ has a *double cone* at a point $a \in E$, or that a is *a double cone point* of E, if there are L, s, and r, possibly depending on a, such that $E \cap \mathcal{C}(a, r, L, s) = \varnothing$.

The following theorem was proved in [H].

Theorem 1.1. *Suppose that $n = 3, 5, 6, \ldots$ and that f is a quasiconformal mapping of $\mathbb{B}^n$ onto a Jordan domain D. Let $\mathcal{C}_D$ denote the set of double cone points of ∂D. Then for any set $A \subset \mathcal{C}_D$ we have that $\mathcal{H}_{n-1}(A) = 0$ if and only if $\mathcal{H}_{n-1}(f^{-1}(A)) = 0$. In particular, if $\mathcal{H}_{n-1}$-almost every point of ∂D is a double cone point of ∂D, then the boundary map $f : \partial\mathbb{B}^n \to \partial D$ is absolutely continuous.*

It follows from Theorem 1.1 in particular that if ∂D admits a tangent plane at $\mathcal{H}_{n-1}$ almost every point, then the boundary correspondence of f is absolutely continuous. If f is quasisymmetric, it follows from the results in [V2] that ∂D admits tangents almost everywhere, if it has finite $\mathcal{H}_{n-1}$ measure. Hence Theorem 1.1 contains the aforementioned results of Gehring and Väisälä in dimensions $n \neq 4$. The proof in [H] works in all dimensions $n \geq 3$ for mappings that are bi-Lipschitz in the quasihyperbolic metric; by the aid of the Sullivan-Tukia-Väisälä approximation theorem the general quasiconformal case can be reduced to this case in dimensions different from four. Unfortunately, I have not been able to dispense with this reduction, and consequently, there is no proof of Theorem 1.1 in dimension $n = 4$ (see Added in Proof at the end of the paper).

Geometric measure theory has taught us that the right concept of rectifiability is expressed in terms of "approximate tangents". If $E \subset \mathbb{R}^n$, we say that an $(n-1)$-plane V in $\mathbb{R}^n$ is an *approximate tangent plane* for E at a if a is a point of $\mathcal{H}_{n-1}$ density of E and if

$$\lim_{r \to 0} \frac{\mathcal{H}_{n-1}(E \cap B(a, r) \setminus \mathcal{C}(a, V, s))}{r^{n-1}} = 0\,,$$

for all $0 < s < 1$, where

$$\mathcal{C}(a, V, s) = \{x \in \mathbb{R}^n : \operatorname{dist}(x, V) < s\,|a - x|\}\,.$$

If such a plane V exists, it is unique and we denote it by $\operatorname{apTan}(E, a)$. Intuitively, $\operatorname{apTan}(E, a)$ is a plane that approximates E near a except for some leftover part which has zero $\mathcal{H}_{n-1}$-density along each cone with vertex at a and axis perpendicular to the plane $\operatorname{apTan}(E, a)$.

In this paper we shall call a set $(n-1)$*-rectifiable* (or sometimes simply *rectifiable* if there is no danger of misunderstanding the dimension of the set) if it has finite $\mathcal{H}_{n-1}$-measure and if it admits an approximate tangent plane at $\mathcal{H}_{n-1}$-almost all of its points. We refer to [F] and [M] for more information about rectifiable sets. (Warning: the terminology in both [F] and [M] is slightly different.) It suffices to mention here that a set E of finite $\mathcal{H}_{n-1}$-measure in $\mathbb{R}^n$ is $(n-1)$-rectifiable if and only if it is contained in a countable union of Lipschitz images of $\mathbb{R}^{n-1}$ inside $\mathbb{R}^n$. Moreover, every set of finite $\mathcal{H}_{n-1}$-measure can be decomposed into a rectifiable and a purely unrectifiable part, the latter being a set whose intersection with any rectifiable set in $\mathbb{R}^n$ has zero $\mathcal{H}_{n-1}$-measure.

Next, we say that a boundary point a of a domain D is an *inner cone point* if there are L, s and r such that $\mathcal{C}(a, r, L^+, s)$ lies in D. Moreover, we say that a is an *inner tangent point* of D if there is a line L with the following property: for each $s < 1$ there is $r > 0$ such that $\mathcal{C}(a, r, L^+, s)$ lies in D. In this case the half line L^+ can be called an *interior normal line* to ∂D at a. Naturally, L^+ need not be unique.

The following theorem is the first main result of this paper.

Theorem 1.2. *Suppose that f is a quasiconformal mapping of $\mathbb{B}^3$ onto a Jordan domain D with 2-rectifiable boundary. Then we have a decomposition of ∂D into three disjoint sets,*

$$\partial D = E_0 \cup E_1 \cup E_2\,,$$

where E_0 has $\mathcal{H}_2$-measure zero, E_1 consists of points of inner tangency of D, and E_2 consists of points of 3-density of $\mathbb{R}^3 \setminus D$. The Hausdorff dimension of $f^{-1}(E_2)$ is zero, and for a set $A \subset E_1$ the preimage $f^{-1}(A)$ has $\mathcal{H}_2$-measure zero if and only if A has $\mathcal{H}_2$-measure zero.

In other words, if D is a Jordan domain in $\mathbb{R}^3$ with 2-rectifiable boundary and if f maps $\mathbb{B}^3$ quasiconformally onto D, then, apart from an $\mathcal{H}_2$-null set, the boundary ∂D consists of the “good part”, where f

and f^{-1} both are absolutely continuous, and the "bad part", which is easily detected and which is responsible for the possible failure of the absolute continuity of $f|_{\partial\mathbb{B}^3}$. Thus the only way the absolute continuity can fail for domains with rectifiable boundary is to have a situation where the bad part E_2 is non-empty and has positive $\mathcal{H}_2$-measure. The next theorem says that such situations can occur.

Theorem 1.3. *For each $n \geq 3$ there is a Jordan domain D in $\mathbb{R}^n$ such that D is quasiconformally equivalent to $\mathbb{B}^n$, that ∂D is $(n-1)$-rectifiable, and the set*

$$E_2 = \{a \in \partial D : a \text{ is a point of } n\text{-density of } \mathbb{R}^n \setminus D\} \tag{1.4}$$

has positive $\mathcal{H}_{n-1}$-measure. Moreover, the preimage $f^{-1}(E_2)$ under any quasiconformal map f from $\mathbb{B}^n$ onto D has Hausdorff dimension zero.

Theorem 1.3 answers negatively to an inquiry of Baernstein and Manfredi [BM, p. 846]. It also shows that Theorem 1.1 is quite sharp. (Note that if ∂D has finite $\mathcal{H}_{n-1}$-measure and if it admits double cones at $\mathcal{H}_{n-1}$-almost everywhere, then it is $(n-1)$-rectifiable; see [M, Lemma 15.13]). Of course, it is easy to construct Jordan domains with the measure theoretic properties as in Theorem 1.3; the nontrivial part is to show that some of them can be mapped quasiconformally onto a ball.

The fact that $f^{-1}(E_2)$ has Hausdorff dimension zero in Theorems 1.2 and 1.3 is a recent result of Koskela and Rohde [KR]. They prove, among other things, that the preimage of the set E_2 as described in (1.4) has zero Hausdorff dimension *always*; that is, in all dimensions and for all quasiconformal mappings $f : \mathbb{B}^n \to D$ (with boundary values properly interpreted if D is not Jordan). In our situation, it would be much easier to show that $f^{-1}(E_2)$ has $\mathcal{H}_{n-1}$-measure zero. In fact, the method described in this paper shows that one can construct a domain D as in Theorem 1.3 such that $\mathcal{H}_{n-1}(E_2)$ is positive and that $f^{-1}(E_2)$ has zero Hausdorff $\mathcal{H}_h$-measure for any prescribed Hausdorff measure function h. The construction of the domain is based on the ideas of Väisälä in [V4], where he constructed a quasiconformal Jordan ball whose boundary has positive n-measure. The elaboration of Väisälä's method presented here leads to a general "tree and pipeline" procedure to build quasiconformal balls and may be of independent interest.

One may ask whether the assumption in Theorem 1.2 that ∂D be 2-rectifiable can be relaxed to the assumption that $\mathcal{H}_2(\partial D)$ be finite. I do not know the answer. An example can be constructed to show that the assumptions in Theorem 1.2 cannot be relaxed to "D is Jordan and ∂D has σ-finite Hausdorff $\mathcal{H}_{n-1}$-measure".

I conjecture that Theorem 1.2 is true in all dimensions $n \geq 3$. In the present paper, the argument for Theorem 1.2 relies in a crucial way on the following local description of the boundary of a quasiconformal Jordan ball (see Added in Proof at the end of the paper).

Theorem 1.5. *Suppose that D is a Jordan domain in $\mathbb{R}^3$ which is homeomorphic to $\mathbb{B}^3$ via a K-quasiconformal map. Then for each $x \in D$ we have the estimate*

$$\mathcal{H}_2(B(x, 2\,\mathrm{dist}(x, \partial D)) \cap \partial D) \;\geq\; C(K)\;\mathrm{dist}(x, \partial D)^2\,. \tag{1.6}$$

Theorem 1.5 is interesting in its own right. It quantifies the fact that the boundary of a quasiconformal ball cannot have lower dimensional parts protruding inwards. It has also led Jussi Väisälä to make general conjectures about isodiametric inequalities for sets that satisfy certain connectivity conditions; see (6.1) below. I make the following conjecture involving quasiconformal mappings.

1.7. Wall Conjecture for Quasiconformal Balls. *If D is a domain in $\mathbb{R}^n$ that is homeomorphic to $\mathbb{B}^n$ via a K-quasiconformal map, then for each $x \in D$ we have the estimate*

$$\mathcal{H}_{n-1}(B(x, 2\,\mathrm{dist}(x, \partial D)) \cap \partial D) \;\geq\; C(n, K)\;\mathrm{dist}(x, \partial D)^{n-1}\,. \tag{1.8}$$

Note that the conjecture is true for $n = 2$ for quite trivial reasons; namely, there is a big connected piece of the boundary inside $B(x, 2\,\mathrm{dist}(x, \partial D))$. The conjecture is also true for $n = 1$, when properly interpreted. Despite some effort, I have not been able to prove the conjecture for $n \geq 4$. Assuming that it is true even in the weaker form where the constant $C(n, K)$ in (1.8) is allowed to depend on D, the proof for Theorem 1.2 will work *mutatis mutandis* for all $n \geq 5$. Dimension $n = 4$ has to be excluded for the same reason it is excluded in [H]: at some point in the proof we need to resort to the fact that in dimensions $n \neq 4$ quasiconformal maps of $\mathbb{B}^n$, say, can be replaced by locally bi-Lipschitz quasiconformal maps without changing the boundary values.

It follows from Theorem 1.5 and standard capacity estimates that every quasiconformal Jordan ball in $\mathbb{R}^3$ is regular for the Dirichlet problem for the Laplacian; in fact, it is regular for the p-Laplace equation for all $p > 1$. It is not true that an arbitrary Jordan domain in $\mathbb{R}^3$ is regular for the Laplacian as the well known Lebesgue's spine demonstrates. If the Wall Conjecture 1.7 is true, then quasiconformal Jordan balls are regular for the p-Laplacian in all dimensions and for all $p > 1$. I thank Pekka Koskela for pointing out this application.

Finally, I wish to point out the recent interesting paper by Hanson [Ha], where rectifiability (Hanson uses a weaker notion here) of the boundary of a quasiconformal Jordan ball is tied up with the behavior of the *average derivative* a_f in the classical spirit. Recall that the rectifiability of a Jordan curve Γ in the plane is equivalent to the membership of f' in the Hardy class H^1 for any conformal map f from the unit disk onto the domain bounded by Γ. Hanson proves in [Ha] that among all quasiconformal Jordan balls D in $\mathbb{R}^n$, $n \geq 3$, that are also so-called *uniform domains*, the finiteness of $\mathcal{H}_{n-1}(\partial D)$ is equivalent to the membership of the average derivative a_f in a "Hardy space" H^{n-1}, if f maps $\mathbb{B}^n$ quasiconformally onto D. (We refer to [Ha] for a precise definition for these concepts.) Many relations between a_f, rectifiability and absolute continuity remain to be sorted out. In particular, it is plausible that Hanson's theorem indeed requires some extra assumptions on D. Hanson [Ha, (5.8) p. 140–141] also advances a conjecture about quasiconformal mappings that is similar to the wall conjecture 1.7. It is not clear what the relationship between these two conjectures are.

2. Outline of Proof for Theorem 1.2.

In this section the main points in the proof for Theorem 1.2 are sketched for the expert's convenience.

Assume that the Wall Conjecture 1.7 is true. We know that $\mathcal{H}_{n-1}$-almost every point on ∂D has an approximate tangent plane. We let E_0 be the exceptional set, and E_2 the set consisting of the points of n-density of the complement of D. For the first part of the theorem, it suffices to show that each point in $E_1 = \partial D \setminus E_0 \cup E_2$ is a point of inner tangency for D. If this is not the case, there is a point $a \in E_1$ and a line L which is perpendicular to $\mathrm{apTan}(\partial D, a)$ and so oriented that the cone $\mathcal{C}(a, r, L^+, s)$ intersects the boundary ∂D for arbitrarily small $r > 0$ and for some fixed s. By (1.8), to each Whitney cube of the open set $D_r = D \cap \mathcal{C}(a, r, L^+, s)$ there corresponds a substantial piece of the boundary lying in a cone $\mathcal{C}(a, r, L^+, s')$ with somewhat bigger opening $s' < 1$. Because a is not a point of density of $\mathbb{R}^n \setminus D$, the set D_r is indeed nonempty and has n-measure comparable to r^n. Now the boundary pieces are essentially disjoint, and their $\mathcal{H}_{n-1}$-measures add up to something which is comparable to r^{n-1}. This contradicts the fact that ∂D admits an approximate tangent plane at a.

The second assertion of Theorem 1.2 is an improvement to Theorem 1.1 which says that f and f^{-1} preserve sets of zero $\mathcal{H}_{n-1}$-measure on double cone points. The proof given in [H] requires double cone points, but, below in Section 3, I give a sharpening of that argument which only needs interior cones, in the presence of approximate tangents. The technical argument of [H] can be shortened somewhat, but the basic idea is still the same. Suppose, for instance, that there is a subset A of interior cone points of ∂D of positive $\mathcal{H}_{n-1}$-measure such that $f^{-1}(A)$ has zero $\mathcal{H}_{n-1}$-measure. After a standard reduction, we may assume that A lies on the boundary of a bi-Lipschitz ball contained in D, hence we may assume without loss of generality that A lies on the boundary of a round ball B contained in D. Then we use the assumption that $n \neq 4$ and replace f by a locally bi-Lipschitz quasiconformal homeomorphism F which agree with f on the boundary. The technical point, as in [H], is to show that $F^{-1}(B) = \Omega$ is a uniform domain with "nice" boundary in $\mathbb{B}^n$; the niceness is defined in terms of the following Ahlfors-David regularity condition:

$$C^{-1} R^{n-1} \le \mathcal{H}_{n-1}(B(x, R) \cap \partial\Omega) \le C R^{n-1}$$

for each $x \in \partial\Omega$ and $0 < R < \operatorname{diam}\Omega$. This condition and known results on quasisymmetric maps onto regular surfaces guarantee that $F|_{\partial\Omega}$ is absolutely continuous, contradicting the hypothesis. In establishing this technical point, we use a Hayman-Wu type "spotting" technique and a Carleson measure argument; the main difference from [H] is that now we have to make use of the approximate tangent planes in place of the exterior cones. More details will follow in the next section.

3. Proof of Theorem 1.2.

The ensuing proof works in all dimensions $n \ge 3$, under right assumptions. Thus, assume that f is a quasiconformal mapping from

$\mathbb{B}^n$ onto a Jordan domain D with rectifiable boundary. Also assume that D satisfies (1.8) for some constant C, possibly depending on D. By Theorem 1.5 this is always true in dimension $n = 3$. Then the conclusion is that the boundary ∂D decomposes as in Theorem 1.2 with $f^{-1}(E_2)$ having Hausdorff dimension zero. If in addition $n \neq 4$, then the absolute continuity of $f|_{f^{-1}(E_1)}$ and $f^{-1}|_{E_1}$ is also true as in Theorem 1.2.

To begin the proof, let E_0 denote the set on ∂D where ∂D does not admit approximate tangent planes. Then E_0 has $\mathcal{H}_{n-1}$-measure zero. We divide $\partial D \setminus E_0$ into two subsets E_1 and E_2, where E_2 consists of the points of n-density of the complement of D in $\mathbb{R}^n$, and E_1 is what remains. Our first task will be to show that every point in E_1 is a point of inner tangency for D.

3.1. Inner tangency of points in E_1.

Pick a point $a \in E_1$. Let L be the line through a which is perpendicular to the approximate tangent plane for ∂D at a. Fix $0 < s < 1$. We need to show that there is $r > 0$ such that one of the two components of the double cone $\mathcal{C}(a, r, L, s)$ is contained in D. Because a is not a point of n-density for the complement of D, and because ∂D has finite $\mathcal{H}_{n-1}$-measure, we can assume, by making s larger if necessary and by choosing an appropriate orientation for L, that

$$\limsup_{r\to 0} \frac{\mathcal{H}_n\big(\mathcal{C}(a, r, L^+, s) \cap D\big)}{r^n} > 0\,. \tag{3.2}$$

Next, we suppose that

$$\partial D \cap \mathcal{C}(a, r, L^+, s) \neq \varnothing$$

for all $r > 0$ and then show that this leads to a contradiction with the fact that

$$\lim_{r\to 0} \frac{\mathcal{H}_{n-1}(\partial D \cap \mathcal{C}(a, r, L^+, s'))}{r^{n-1}} = 0 \tag{3.3}$$

for all $0 < s' < 1$.

To this end, let $D_r = \mathcal{C}(a, r, L^+, s) \cap D$ and observe that $D_r \neq \varnothing$ by (3.2). Suppose first that for each $x \in D_r$ the ball $B_x = B(x, \operatorname{dist}(x, \partial D))$ satisfies

$$\operatorname{diam} B_x \le \varepsilon \operatorname{dist}(B_x, a)\,, \tag{3.4}$$

where $\varepsilon = \varepsilon(s) > 0$ is so small that (3.4) implies

$$2B_x \subset \mathcal{C}(a, 2r, L^+, (1+s)/2)\,.$$

By standard covering theorems (see [M, Chapter 2]), we can choose a countable collection $\{B_i : i = 1, 2, \dots\}$ of balls of the form B_x such that

$$D_r \subset \bigcup_i 2B_i$$

and that

$$\sum_i \chi_{2B_i}(x) \;\leq\; C(n)\,.$$

The latter condition simply says that no point in $\mathbb{R}^n$ belongs to more than $C(n)$ balls of the form $2B_i$. Therefore, by assumption (1.8),

$$\begin{aligned}
\mathcal{H}_{n-1}(\partial D \cap \mathcal{C}(a, 2r, L^+, (1+s)/2)) \;&\geq\; C^{-1} \sum_i \mathcal{H}_{n-1}(\partial D \cap 2B_i) \\
&\geq\; C^{-1} \sum_i (\operatorname{diam} B_i)^{n-1} \\
&\geq\; C^{-1} \Big(\sum_i (\operatorname{diam} B_i)^n\Big)^{(n-1)/n} \\
&\geq\; C^{-1} (\mathcal{H}_n(D_r))^{(n-1)/n}\,.
\end{aligned}$$

Because the constant $C \geq 1$ above is independent of $r > 0$, we contradict (3.3) with the aid of (3.2).

We may thus assume that

$$\operatorname{diam} B_x > \varepsilon \operatorname{dist}(B_x, a)$$

for some $x \in D_r$ and $B_x = B(x, \operatorname{dist}(x, \partial D))$. In this case a simple geometric argument proves the existence of a point $y \in D_r$ and a ball $B_y = B(y, \operatorname{dist}(y, \partial D))$ that belongs to $\mathcal{C}(a, r, L^+, s)$ and satisfies both,

$$2B_y \subset \mathcal{C}(a, 2r, L^+, (1+s)/2)$$

and

$$C^{-1} \operatorname{diam} B_y \;\leq\; \operatorname{dist}(B_y, a) \;\leq\; C \operatorname{diam} B_y\,,$$

for some $C \geq 1$ depending only on n and s. Thus we deduce that, for some $r' < r$,

$$\mathcal{H}_{n-1}(\partial D \cap \mathcal{C}(a, 2r', L^+, (1+s)/2)) \;\geq\; \mathcal{H}_{n-1}(\partial D \cap 2B_y) \;\geq\; C^{-1} {r'}^{n-1}$$

by assumption (1.8). Moreover, $C \geq 1$ is independent of r and r'. This again contradicts (3.3) and we have shown that a is a point of inner tangency of D.

3.5. Absolute continuity in the inner tangency set E_1.

Recall that the fact that $f^{-1}(E_2)$ has Hausdorff dimension zero is due to Koskela and Rohde [KR]. To complete the proof of the theorem, it thus remains to show that f and f^{-1} are absolutely continuous in the sets $f^{-1}(E_1) \subset \partial\mathbb{B}^n$ and $E_1 \subset \partial D$. The proof here has the same idea as in [H]. In that paper, however, the absolute continuity was proved in the set of double cone points, and the existence of an exterior cone was also essentially used there. In the present situation we only have an interior cone to rely on. The supporting role of the exterior cone is taken here by the approximate tangent plane, which exists at each point in E_1. This change forces us to make some technical modifications to the proof in [H]. It would be unreasonable to repeat here all the details of [H], and I apologetically ask the reader to consult that paper whenever necessary. The good news is that the most technical part of the proof of [H, Lemma 3.1] has now been simplified somewhat.

Let us begin with the following lemma.

Lemma 3.6. *Let a be a point in E_1 and denote by T_a the approximate tangent plane* $\operatorname{apTan}(\partial D, a)$*. Then*

$$\limsup_{r\to 0} \inf_{v\in T_a, |v-a|=r} \frac{\operatorname{dist}(v,\partial D)}{|v-a|} = 0\,.$$

PROOF. This lemma looks trivial but a little thinking shows that it need not be true if we replace ∂D by an arbitrary $(n-1)$-rectifiable set. In any event, the ensuing proof is quite easy.

Suppose on the contrary that there is $\delta > 0$ and a sequence of radii (r_i), $r_i \to 0$ as $i \to \infty$, such that

$$\frac{\operatorname{dist}(v,\partial D)}{|v-a|} > \delta\,,$$

for all $v \in T_a$ with $|v-a| = r_i$. Then the δr_i-neighborhood U_i of the $(n-2)$-sphere $\partial B(a, r_i) \cap T_a$ does not meet ∂D. Because a is a point of inner tangency of D, it follows that $U_i \subset D$ for all large i. Let L^+ be the interior normal line at a, perpendicular to T_a. Fix s so close to 1 that

U_i meets $\mathcal{C}(a, L^+, s)$ for all large i, and then choose $r_i > 0$ so that the cone $\mathcal{C}(a, r_i, L^+, s)$ is contained in D. Clearly $\mathcal{C}(a, r_i, L^-, s)$ cannot be contained in D, for otherwise the connected open set $U_i \cup \mathcal{C}(a, r_i, L, s)$ is contained in D for all large i and separates the point a from the part of the boundary that lies outside $B(a, r_i)$. Thus, for arbitrary small $r_i > 0$ we have that

$$\mathcal{C}(a, r_i, L^-, s) \cap \partial D \neq \varnothing\,.$$

Using assumption (1.8) on the thickness of the boundary, this leads to a contradiction with an argument similar to that in the end of the proof in (3.1). (Note: the analog of (3.1) in the present case is guaranteed by the size of U_i.) The lemma follows.

3.7. Reduction to a ball.

Suppose now that $A \subset E_1$ has positive $\mathcal{H}_{n-1}$ measure. We need to show that $f^{-1}(A)$ has positive $\mathcal{H}_{n-1}$-measure as well. And this is in fact all that needs to be shown in detail, for the case

$$A \subset E_1 \text{ and } \mathcal{H}_{n-1}(A) = 0 \quad \text{implies} \quad \mathcal{H}_{n-1}(f^{-1}(A)) = 0$$

is treated similarly.

A standard measure theoretic trick guarantees that there is a subset $A_0 \subset A$ of positive $\mathcal{H}_{n-1}$-measure which lies on the boundary of a bounded starshaped subdomain $\Omega_0 \subset D$. The domain Ω_0 can be mapped onto a ball by a bi-Lipschitz self-map of $\mathbb{R}^n$. Because bi-Lipschitz maps preserve rectifiability and sets of positive Hausdorff measure, we can assume, originally, that A lies on the boundary of a ball B_0 contained in D. See [H, Proof of Theorem 4.3] for more details here.

Next we form a Stolz domain Ω in B_0, associated with A the usual way. That is, Ω consists of all the open rays with one end point in $B_0/2$ and the other in A. Then Ω is a bi-Lipschitz ball contained in D and containing A on its boundary. (Note that the round ball B_0 already satisfies these conditions and it would be nice if we could manage with B_0 alone. It is the proof below in (3.14) that needs a domain like Ω which is safely inside B_0.)

3.8. Bi-Lipschitz maps in the quasihyperbolic metric.

Now we use the assumption $n \neq 4$. The Sullivan-Tukia-Väisälä approximation theorem [TV2, 7.12] provides us with a quasiconformal map $F : \mathbb{B}^n \to D$ such that

$$k_D(f(x), F(x)) \leq 1 \tag{3.9}$$

and that

$$C^{-1} k_D(F(x), F(y)) \leq k_{\mathbb{B}^n}(x, y) \leq C\, k_D(F(x), F(y)), \tag{3.10}$$

for all x and y in $\mathbb{B}^n$ and for some $C = C(n, f) \geq 1$. Here k_G denotes the quasihyperbolic metric in a domain G, defined by the metric density $\operatorname{dist}(x, \partial G)^{-1}|dx|$.

Condition (3.9) guarantees that f and F have the same boundary values and (3.10) says that F is bi-Lipschitz in the quasihyperbolic metrics. We deduce that there is no loss of generality in assuming, originally, that the mapping f satisfies (3.10).

3.11. Regular surfaces and subinvariance.

Write $g = f^{-1}$. The main bulk of the proof consists of showing that the boundary $\partial g(\Omega)$ is an Ahlfors-David $(n-1)$-regular set; that is, there is a constant $C \geq 1$ such that

$$C^{-1} R^{n-1} \leq \mathcal{H}_{n-1}(B(x, R) \cap \partial g(\Omega)) \leq C R^{n-1}, \tag{3.12}$$

for all $x \in \partial g(\Omega)$ and $0 < R < \operatorname{diam} g(\Omega)$.

Suppose for a moment that this has been accomplished. The proof is then finished as follows. The *subinvariance principle* for quasiconformal maps guarantees that $g|_\Omega : \Omega \to g(\Omega)$ is a quasisymmetric map, which means that

$$|x - y| \leq t\,|x - z| \quad \text{implies} \quad |g(x) - g(y)| \leq \eta(t)\,|g(x) - g(z)| \tag{3.13}$$

for all points $x, y, z \in \Omega$ and for some homeomorphism $\eta : [0, \infty) \to [0, \infty)$. (See [FHM, p. 120-121] and [V3, Theorem 5.6]). Clearly (3.13) will continue to hold for all points in the closure $\overline{\Omega}$, so that $g|_{\partial\Omega} : \partial\Omega \to \partial\Omega$ is quasisymmetric as well. Because $\partial\Omega$ is a bi-Lipschitz $(n-1)$-sphere and because $\partial g(\Omega)$ is $(n-1)$-regular in the sense of (3.12), we can invoke known results about quasisymmetric maps in such situations to conclude that $\mathcal{H}_{n-1}(E) > 0$ if and only if $\mathcal{H}_{n-1}(g(E)) > 0$ for $E \subset \partial\Omega$. (See [S, 3.4] or [H, 2.7]).

We conclude, therefore, that it remains to prove the regularity (3.12) of $\partial g(\Omega)$.

3.14. Proof of regularity of $\partial g(\Omega)$.

We begin by making two more reductions. Fix ε positive and small, to be determined later. By Lemma 3.6 there is, for each $a \in A$, a positive number r_a such that

$$\operatorname{dist}(v_r, \partial D) < \varepsilon\, |v_r - a|, \tag{3.15}$$

for some $v_r \in T_a \cap \partial B(a, r)$ and for all $0 < r \le r_a$. Then

$$A = \bigcup_{j=1}^{\infty} \{a \in A : r_a > 1/j\},$$

and we may assume that there is $\delta > 0$ such that $r_a > \delta > 0$ for all $a \in A$. We assume further that each point in A is a point of $\mathcal{H}_{n-1}$-density on ∂B_0 (recall that A lies in the smooth hypersurface ∂B_0 and has positive $\mathcal{H}_{n-1}$-measure).

The left inequality in (3.12) follows from the quasisymmetry of g in $\partial\Omega$ by a result of Väisälä [V2, 5.2]. The right inequality in (3.12) follows by standard arguments using (3.10) from the following lemma (for the details, see [H, p. 1564-65]).

Lemma 3.16. *Suppose that (x_i) is a hyperbolically separated sequence of points on $\partial\Omega$; this means that there is $\tau > 0$ such that*

$$B(x_i, \tau \operatorname{dist}(x_i, \partial D)) \cap B(x_j, \tau \operatorname{dist}(x_j, \partial D)) = \varnothing \tag{3.17}$$

whenever $i \neq j$. Then there is $C \ge 1$ such that

$$\sum_{g(x_i) \in B} (1 - |g(x_i)|)^{n-1} \le C (\operatorname{diam} B)^{n-1}, \tag{3.18}$$

for all n-balls B centered on $\partial\mathbb{B}^n$.

One should notice that in (3.18) the constant $C \geq 1$ can, and usually will, depend on everything else but on B.

Thus, fix a ball B, centered at some point on $\partial\mathbb{B}^n$. Henceforth C will denote any positive constant that is independent of B and also of index i. The goal is to find for each x_i in $g^{-1}(B) = f(B \cap \mathbb{B}^n)$ its own spot S_i on ∂D such that the following three conditions hold:

$$\sum_i \chi_{S_i}(x) \leq C, \tag{3.19}$$

that is, no point in $\mathbb{R}^n$ belongs to more that C spots S_i;

$$g(S_i) \subset CB \cap \partial\mathbb{B}^n, \tag{3.20}$$

that is, the image of each spot S_i will not land far from B under the map g; and

$$(1 - |g(x_i)|)^{n-1} \leq C\,\mathcal{H}_{n-1}(g(S_i)), \tag{3.21}$$

that is, the Hausdorff measure of the image $g(S_i) \subset \partial\mathbb{B}^n$ essentially dominates the term $(1 - |g(x_i)|)^{n-1}$ of the sum in (3.18).

It is clear that (3.18) follows from (3.19)-(3.21).

Before we start describing the spots S_i with desired properties, we make two observations.

3.22. Hyperbolic freedom.

There is never any harm in replacing any of the points x_i by a point $\tilde{x}_i$ for which

$$k_D(x_i, \tilde{x}_i) \leq C, \tag{3.23}$$

because it is easily seen that (3.23) implies

$$(1 - |g(x_i)|) \leq C\,(1 - |g(\tilde{x}_i)|). \tag{3.24}$$

The replacement may cause us to diminish the value of τ in (3.17) a little bit, but such adjustments are left to the reader.

3.25. Generational gaps.

Upon dividing (x_i) into generations $\mathcal{G}_\nu$,

$$x_i \in \mathcal{G}_\nu \quad \text{if and only if} \quad \operatorname{dist}(x_i, \partial B_0) \in (2^{-\nu-1}, 2^{-\nu}]\,, \quad \nu \in \mathbf{Z}\,,$$

we can assume that

$$\mathcal{G}_\nu = \varnothing \quad \text{for} \quad \nu \leq C\,, \tag{3.26}$$

and that

$$\mathcal{G}_{\nu_i} \neq \varnothing \neq \mathcal{G}_{\nu_j} \quad \text{implies} \quad \nu_i = \nu_j \text{ or } |\nu_i - \nu_j| \geq C\,. \tag{3.27}$$

Above, C should be thought of as a large constant, to be adjusted later. Condition (3.26) means that we only have to worry about those points x_i that lie near the boundary of B_0, and (3.27) says that we can assume that there are large generational gaps. In short, we assume that $\mathcal{G}_\nu$ is nonempty only if ν is positive and a constant multiple of a large integer. We shall construct the spots S_i in such a way that S_i and S_j are disjoint whenever they correspond to points in different generations, and that the finite overlap condition (3.19) holds for spots S_i corresponding to points from the same generation.

3.28. Determining points z_i.

We shall associate to each point x_i in our sequence two more points, w_i and z_i, of which the latter will play a more important role. To get a mental picture,

$$x_i \quad \rightsquigarrow \quad w_i \in \partial B_0 \quad \rightsquigarrow \quad z_i \in \partial D\,.$$

The point w_i is simply the closest point to x_i on ∂B_0, and z_i is a closest point to w_i on ∂D. Of course, it may happen that $w_i = z_i$. Before we fix these, however, we need to make some adjustments to the sequence (x_i) in the spirit of (3.22).

Thus, pick a point x_i. Let w_i be the closest point to x_i on ∂B_0, and let a_i be the closest point to x_i on A. Because each point in A is assumed to be a point of $\mathcal{H}_{n-1}$ -density, it is clear that the approximate tangent plane T_{a_i} is also tangent to ∂B_0 at a_i. By choosing the constant C in (3.26) large enough, we may assume that

$$\operatorname{dist}(w_i, T_{a_i}) < \varepsilon\, |w_i - a_i|\,, \tag{3.29}$$

where $\varepsilon > 0$ is as in (3.15).

Let w_i' be the point where the ray emanating from the center of B_0 and passing through x_i meets T_{a_i}, and let

$$r_i' = |a_i - w_i'| .$$

Again, by making the constant C in (3.26) large enough, we may assume that $r_i' < \delta$ for all i, where δ is defined just after (3.15). Thus we can find a point $v_i \in T_{a_i} \cap \partial B(a_i, r_i')$ such that

$$\operatorname{dist}(v_i, \partial D) < \varepsilon\, |v_i - a_i| .$$

Now let $\tilde{x}_i$ be the point on the line segment from v_i to the center of B_0 such that

$$\operatorname{dist}(\tilde{x}_i, \partial B_0) = \operatorname{dist}(x_i, \partial B_0) .$$

It is easy to see that

$$k_D(x_i, \tilde{x}_i) \,\le\, C .$$

Therefore, by the discussion 3.22, we may assume, originally, that

$$\operatorname{dist}(w_i', \partial D) < \varepsilon\, |w_i' - a_i|$$

and hence that

$$\operatorname{dist}(w_i, \partial D) \,\le\, |w_i - w_i'| + \varepsilon\, |w_i' - a_i| \,\le\, 2\varepsilon\, |w_i - a_i| ,$$

provided that C in (3.26) is large enough, depending on ε. Next, let z_i be a point on ∂D such that

$$|z_i - w_i| = \operatorname{dist}(w_i, \partial D)$$

and observe that

$$|z_i - w_i| \,\le\, 2\varepsilon\, |w_i - a_i| . \tag{3.30}$$

At this point we could invoke the argument in [H, Main Lemma 3.1] which applies in the present situation. The double cone condition there was used only to guarantee the existence of the points z_i satisfying (3.30). For the reader's convenience, however, I shall sketch below a somewhat different and perhaps easier argument for the rest of the proof.

Towards this end, we require the following lemma which is proved in [HK, 6.6].

Lemma 3.31. *Let $x \in D$. There is a constant $C \geq 1$, depending only on n and on the dilatation of f, such that*

$$\mathcal{H}_{n-1}\big(g(B(x, C\operatorname{dist}(x,\partial D)) \cap \partial D) \cap \Delta_{g(x)}\big) \;\geq\; \frac{1}{2}\,\mathcal{H}_{n-1}(\Delta_{g(x)})\,,$$

where $\Delta_{g(x)}$ is the surface cap,

$$\Delta_{g(x)} = B(g(x), 3(1-|g(x)|)) \cap \partial\mathbb{B}^n\,.$$

Now fix $0<\lambda<1$ and let u_i be the point $(1-\lambda)z_0 + \lambda w_i$ in B_0, where z_0 is the center of B_0. We have

$$\operatorname{dist}(u_i,\partial D) \;\leq\; |u_i - w_i| + |w_i - z_i| \;=\; (1-\lambda) + 2\varepsilon\,|w_i - a_i|\,.$$

By choosing $\lambda = \lambda_i$ such that

$$(1-\lambda_i) \;=\; \varepsilon\operatorname{dist}(x_i,\partial B_0)\,,$$

we find that

$$\operatorname{dist}(u_i,\partial D) \;\leq\; \varepsilon\operatorname{dist}(x_i,\partial B_0) \;+\; 2\varepsilon\,|w_i - a_i| \;\leq\; 10\,\varepsilon\,|w_i - a_i|\,. \tag{3.32}$$

Finally, define S_i by

$$g(S_i) = g\big(B(u_i, C\operatorname{dist}(u_i,\partial D)) \cap \partial D\big) \cap \Delta_{g(u_i)}\,,$$

where C is as in Lemma 3.31. We easily infer by choosing $\varepsilon > 0$ small enough, by observing the generational gap (3.27), Lemma 3.31, formula (3.32), and the geometry of Ω that this choice of S_i will satisfy (3.19)-(3.21). Of course, we need to observe here that

$$k_D(u_i, x_i) \;\leq\; C\,,$$

as well as the assumed hyperbolic separation (3.17) of the points x_i. The details are left to the interested reader.

This completes the proof of Theorem 1.2.

4. A class of quasiconformal balls – proof of Theorem 1.3.

In this section I exhibit a general method to build quasiconformal balls. As mentioned before, the method described below is essentially due to Väisälä who constructed a single interesting example in [V4]; the main idea of blowing up towers with moderate dilatation is of course old and goes back to the early articles of Gehring and Väisälä [GV]. My contribution is simply to axiomatize the construction done in [V4], and then point out how one obtains this way examples that are relevant to the boundary absolute continuity problem.

4.1. Admissible trees.

An admissible tree in $\mathbb{R}^n$ is a tree around which one can build a quasiconformal ball. A precise definition follows shortly. In the ensuing discussion, all line segments are assumed to be finite and closed. We shall work in $\mathbb{R}^n$ for any n bigger than one, although the Riemann mapping theorem trivializes the discussion for $n = 2$.

Let L_0 be a line segment in $\mathbb{R}^n$ and fix $\alpha \in (0, \pi/2]$. Set $\mathcal{J}_0 = \{L_0\}$. Suppose next that a finite collection $\mathcal{J}_i$ of line segments has been determined for all $i = 0, \ldots, k$. Let $L \in \mathcal{J}_k$ be a line segment. Attach a finite number of line segments $L_1, \ldots, L_p$ to L in such a way that

1) exactly one of the end points of each L_i lies on $L \setminus \{$the end points of $L\}$;

2) the angle between each L_i and L is at least $\alpha > 0$;

3) all line segments L_i are mutually disjoint and none of them meets any line segment from $\mathcal{J}_0 \cup \cdots \cup \mathcal{J}_k$ except their *parent* L at one end point.

We further stipulate that all the children of all line segments from $\mathcal{J}_k$ as described above are mutually disjoint; they form the collection $\mathcal{J}_{k+1}$.

We call the set

$$T_\alpha = T = \bigcup_{k=1}^{\infty} \mathcal{J}_k$$

an *admissible tree* with *branching angle* α if it is a bounded set, constructed by the above rules 1)-3), and has the additional property that each line segment L from T retains a positive distance (depending on L) to all other line segments from T, save its immediate family (that is, its parent and children); more formally,

$$\text{dist}(L, T \setminus \{\text{the parent of } L \text{ and the children of } L\}) > 0 \tag{4.2}$$

for each $L \in T$. We understand that L_0 has no parent and that there can be childless line segments.

Next, denote by F_T the set of all points that lie "behind infinitely many branches". More precisely, x is in F_T if x is a cluster point of infinitely many line segments from T.

Theorem 4.3. *Given an admissible tree $T_\alpha = T$ in $\mathbb{R}^n$, there is a domain D in $\mathbb{R}^n$ such that*

$$T \subset \overline{D}, \tag{4.4}$$

that

$$F_T \subset \partial D, \tag{4.5}$$

and that D is quasiconformally equivalent to $\mathbb{B}^n$ by a K-quasiconformal map with K depending only on n and α.

Even more can be said.

Theorem 4.6. *Given any admissible tree $T_\alpha = T$ in $\mathbb{R}^n$, any continuous nondecreasing function $h : [0,1) \to [0,1)$, $h(t) \to 0$ as $t \to 0$, and any $\varepsilon > 0$, there is a domain D in $\mathbb{R}^n$ satisfying (4.5) and (4.6), and there is a quasiconformal map f from $\mathbb{B}^n$ onto D such that*

$$\mathcal{H}_h(f^{-1}(F_T)) = 0 \tag{4.7}$$

and that

$$\mathcal{H}_{n-1}(\partial D \setminus F_T) < \varepsilon. \tag{4.8}$$

Moreover, one can choose f such that its dilatation depends only on n and α.

Above, $\mathcal{H}_h$ denotes the Hausdorff measure obtained from the measure function h; see [F, 2.10].

Accepting Theorems 4.3 and 4.6, it is easy to construct examples as in Theorem 1.3. For instance, one can take a totally disconnected compact set F in $\mathbb{R}^{n-1}$ with positive $\mathcal{H}_{n-1}$-measure, and then form an admissible tree having branches in the upper half space $\mathbb{R}^n_+ = \{x \in \mathbb{R}^n : x_n > 0\}$ such that the end points of these branches capture each point in F. It follows from the construction below that the boundary of the associated domain D is a rectifiable $(n-1)$-sphere, and that one can arrange each point on F to be a point of n-density for the complement of D.

Many other interesting examples of quasiconformal balls can be exhibited by the aid of the above theorems. For instance, the existence of quasiconformal Jordan balls with boundary having positive n-measure is ascertained by the existence of pertinent admissible trees. Väisälä's goal in [V4] was exactly to construct one such domain. Väisälä was partly motivated by the following consequence of his construction: there are mappings in the Sobolev space $W^{1,n}_{loc}(\mathbb{R}^n;\mathbb{R}^n)$ that do not preserve sets of n-measure zero. Theorem 4.6 can be used to show that mappings in $W^{1,n}$ can blow up quite a miniscule set to a set of positive n-measure. This is done by "folding" a mapping promised in Theorem 4.6, *cf.* [Re], [V4, p. 206]. For a general discussion on this topic, see [MM].

We shall only prove Theorem 4.3. It should be clear how the details need to be changed in order to achieve Theorem 4.6. Overall, we shall rely on the carefully detailed argument in [V4].

4.9. Tower maps.

For $h > 0$ define the *straight tower*

$$T(h) = \Delta \cup (\overline{\mathbb{B}}^{n-1} \times [0,h]) \subset \mathbb{R}^n ,$$

where $\Delta = (-e_n)\overline{\mathbb{B}}^{n-1}$ is the *join* of $-e_n = (0,\dots,0,-1)$ and the closed unit ball of $\mathbb{R}^{n-1}$. If $\alpha \in (0,\pi/2]$, a *leaning tower* $T(h,\alpha)$ is obtained from the straight tower $T(h)$ by keeping the base Δ fixed and tilting the upper part $\overline{\mathbb{B}}^{n-1} \times [0,h]$ so that it makes angle α with the hyperplane $\mathbb{R}^{n-1}$. We call Δ the *basement*, and $\overline{\mathbb{B}}^{n-1}$ the *floor*, of the tower $T(h,\alpha)$. The terms *wall* and *roof* of $T(h,\alpha)$ are selfexplanatory when we make the convention that both these sets consist only of points where $\partial T(h,\alpha)$ is smooth, *i.e.* we ignore the corners.

A *tower map* is a quasiconformal map

$$g : \Delta \to T(h,\alpha) \tag{4.10}$$

such that g is the identity on the part of $\partial\Delta$ that does not include $\overline{\mathbb{B}}^{n-1}$. Strictly speaking, g is quasiconformal only in the interior of the basement Δ, but it extends so as to map Δ homeomorphically onto $T(h,\alpha)$.

The existence of such a map is clear; what is crucial is that it can be chosen so that its dilatation only depends on n and α_0, if $\alpha \geq \alpha_0 > 0$. In particular – and this is the main point – the dilatation does not depend on the height h of the (leaning) tower $T(h,\alpha)$. Moreover, we can choose g such that it is a diffeomorphism at every point in the preimage of the wall of the tower. For an explicit construction of the map g, see [V4, Section 3].

4.11. Flattening of walls and germs of similarity.

Suppose that a leaning tower $T(h,\alpha)$ is given and that $\{a_1,\ldots,a_p\}$ is a finite subset of the wall of $T(h,\alpha)$. One can modify both the tower and the tower map in (4.10) so that it becomes a similarity in small neighborhoods of the points $a_i' = g^{-1}(a_i) \in \overline{\mathbb{B}}^{n-1}$. This is done as follows. First one flattens out a small piece of the slightly curved wall surface near each point a_i. This does not cost much in terms of the dilatation. Then, using the language of Väisälä, one can *plant a germ of similarity* on g near each point a_i'. This means that one can modify the map g so that it becomes a similarity (in particular, conformal) in a neighborhood of a_i'. Moreover, the planting can be done in such a way that the cost in dilatation only depends on n and the dilatation of the original map, that is, on n and α only in our case.

In sum, we can assume that given a tower as above and a finite number of points on its wall, we have a tower map

$$g : \Delta \to T'(h,\alpha)\,, \tag{4.12}$$

where the new tower $T'(h,\alpha)$ is being slightly flattened around the given points. (We could call $T'(h,\alpha)$ a *tilted pajupilli.*) Moreover, g is a similarity near those points and its dilatation only depends on n and a lower bound for the tilt angle of the tower. On the part of the boundary of the basement that lies in the lower half space, the map g is still the identity.

The planting procedure is being described in detail in [V4, Section 2].

4.13. Proof of Theorem 4.3.

Once we have the tower map (4.12) at our disposal, it is rather clear how to continue the proof. Suppose that we are given an admissible tree $T_\alpha = T$. First we map the unit ball under a quasiconformal map f_0 onto a thin cylinder $\mathcal{C}_0$ about T_0 such that the height of the cylinder is the length of T_0 and that T_0 is its axis. The dilatation of f_0 only depends on n, and not on the height. We choose the cylinder $\mathcal{C}_0$ so thin that all the children of L_0 in $\mathcal{J}_1$ stick out of it a good proportion of their length, and that all the other descendants remain at a positive distance from L_0; this is possible by (4.2). We reiterate that $\mathcal{C}_0$ can be made as thin as we please with no extra cost at the dilatation of f_0. Consequently, the surface area of $\mathcal{C}_0$ can be made as small as we please; this observation is needed for Theorem 4.6.

Next, at the points a_i, where the children L_i of L_0 leave the cylinder $\mathcal{C}_0$, we flatten the wall of L_0 and assume, as we may by the discussion in 4.11, that f_0^{-1} is a similarity in a neighborhood U_i of each point a_i. We place small similarity copies Δ_i of Δ in all those neighborhoods U_i such that the origin in Δ corresponds to a_i in Δ_i. Usually the child L_i leaves the cylinder $\mathcal{C}_0$ in a tilt, and we place a thin leaning tower on each Δ_i such that L_i is the axis of the tower and that the other end point of L_i lies on the roof of the tower. Any such tower is a similarity copy of a tower of the form $T(h, \alpha)$ described above in 4.9. We choose these towers so thin that they do not meet other descendants but their immediate children; again this is possible by (4.2).

Each base Δ_i can be mapped quasiconformally onto the leaning tower above it. For this we use the tower map g in (4.10) and appropriate similarities. By declaring each such map to be the identity elsewhere in $\mathcal{C}_0$, we get a map

$$f_1 : \mathbb{B}^n \to \mathcal{C}_1 ,$$

where $\mathcal{C}_1$ is $\mathcal{C}_0$ plus all the new towers placed above each Δ_i. The map f_1 is simply f_0 followed by all those little tower maps. Because f_0 was a similarity on $f_0^{-1}(U_i)$, and because the bases Δ_i are located in U_i, the dilatation of f_1 only depends on n and α. In other words, we did not increase the dilatation by this composition because the only nontrivial contribution came from where f_0 was conformal.

Now we continue in a similar fashion. The walls of all the little towers in $\mathcal{C}_1$ are flattened near the points where the children (the grandchildren of T_0) leave $\mathcal{C}_1$, and f_1 is modified so as to become a similarity near those points. This modification increases the dilatation but there is no accumulation because the increase only occurs at places where f_0 was conformal. Then we blow up new (possibly leaning) towers from those newly created similarity neighborhoods. Thus the dilatation of the map $f_2 : \mathbb{B}^n \to \mathcal{C}_2$ will not grow, where, naturally, f_2 is f_1 followed by the new even littler tower maps, declared to be the identity outside the bases, and $\mathcal{C}_2$ is the union of $\mathcal{C}_1$ and the new towers.

The final map f is the limit of the maps $f_0, f_1, f_2, \dots$ constructed in this manner. Its dilatation in $\mathbb{B}^n$ only depends on n and α, and it maps $\overline{\mathbb{B}}^n$ onto $\overline{D}$, where D is the interior of the union $\mathcal{C}_0 \cup \mathcal{C}_1 \cup \mathcal{C}_2 \cup \cdots$. If the tree is properly arranged, f will be a homeomorphism of the closed unit ball onto $\overline{D}$. It is also clear by construction that the set F_T lies on the boundary of D, and that we can always arrange the boundary ∂D minus, possibly, the set F_T, to be of finite Hausdorff $\mathcal{H}_{n-1}$-measure.

This completes the proof of Theorem 4.3.

5. Proof of the Wall Conjecture in dimension $n = 3$.

Soon after Jussi Väisälä heard about the Wall Conjecture, he devised a simple argument in dimension $n = 3$ which also proves the following more general theorem.

Theorem 5.1. (Väisälä, [V6]) *Suppose that G is an open set in $\mathbb{R}^n$, $n \geq 3$, such that $\check{H}^1(\mathbb{R}^n \cup \{\infty\} \setminus G) = 0$ and that $\mathbb{R}^n \setminus G$ satisfies the condition c-LLC$_2$. Then*

$$\mathcal{H}_2\big(B(x, 2\,\mathrm{dist}(x, \partial G)) \cap \partial G\big) \;\geq\; \frac{\pi}{16\,c}\,(\mathrm{dist}(x, \partial G))^2$$

for each $x \in G$.

In the theorem, $\check{H}^1$ denotes the first Čech cohomology group with integer coefficients. The c-LLC$_2$ condition means that for every x in $\mathbb{R}^n \setminus G$ and $r > 0$ points in $(\mathbb{R}^n \setminus G) \setminus \overline{B}(x, r)$ can be joined in $(\mathbb{R}^n \setminus G) \setminus \overline{B}(x, r/c)$, where $c \geq 1$ is a constant independent of x and r.

If D is a $(K-)$quasiconformal Jordan ball in $\mathbb{R}^3$, then $\check{H}^1(\mathbb{R}^3 \cup \{\infty\} \setminus D) = 0$ by Alexander duality, and $\mathbb{R}^3 \setminus D$ is c-LLC$_2$ for some $c \geq 1$ depending only on K by a theorem of Gehring and Väisälä [GV]. The letters LLC stand for *linear local connectivity*. It is also true, and proved by Gehring and Väisälä [GV], that $\mathbb{R}^3 \setminus D$ satisfies the following

c-LLC$_1$ condition, which is dual to c-LLC$_2$: for every $x \in \mathbb{R}^n \setminus G$ and $r > 0$ points in $(\mathbb{R}^n \setminus G) \cap B(x, r)$ can be joined in $(\mathbb{R}^n \setminus G) \cap B(x, cr)$.

Therefore, Theorem 1.5 follows from Väisälä's Theorem 5.1.

I shall next sketch another proof for Theorem 1.5, but the reader should bear in mind that it is not as elegant as Väisälä's argument and it will not generalize so as to cover Theorem 5.1. But even this proof as such has nothing to do with quasiconformal maps; we shall only employ the LLC condition for the complement. In Problem 5 below in Section 6 we formulate a general conjecture along the lines "quantitative topological conditions imply mass bounds". This type of results have recently been popular in Riemannian geometry; see [GP].

PROOF OF THEOREM 1.5. We can normalize the situation so that $x = 0$ and $\operatorname{dist}(x, \partial D) = 1$. It is an easy exercise to check that it is enough to find constants $C_1 = C_1(K) \geq 2$ and $C_2 = C_2(K) > 0$ such that

$$\mathcal{H}_2(B(0, C_1) \cap \partial D) \geq C_2 . \tag{5.2}$$

Next we invoke a lemma which is due to Gehring [G1, Lemma 1]. In the lemma, we denote by K_1 the decomposition of $\mathbb{R}^3$ into closed cubes with vertices in $\mathbb{Z}^3$; then write $K_s = sK_1$ for $s > 0$, and denote by K_s^1 the 1-skeleton of K_s.

Lemma 5.3. *Suppose that a compact set A in $\mathbb{R}^3$ satisfies*

$$\mathcal{H}_2(A) < \frac{s^2}{64} < 1 \tag{5.4}$$

for some $s > 0$. Then some translate $A - y = \{a - y : a \in A\}$, $y \in \mathbb{R}^3$, does not meet the 1-skeleton K_s^1.

Now choose the constant $C_1 > 2$ in (5.2) very large and $s > 0$ very small (both depending on the constant c in the linear local connectivity condition, hence on K only) and assume that (5.4) holds for $A = \partial D \cap \overline{B}(0, C_1)$. Then the part of the (translated) 1-skeleton K_s^1 that lies in $B(0, C_1)$ does not meet $\mathbb{R}^3 \setminus D$, because it does not meet ∂D, it is connected, and it meets D near the point 0. This will lead to a contradiction as follows. One first selects a curve γ_1 in $\mathbb{R}^3 \setminus D$ that joins some point w on ∂D with $|w| = 1$ to a point in $\partial B(0, C_1)$. Then, by using the LLC$_2$ condition, one selects another curve γ_2 joining the same points in $\mathbb{R}^3 \setminus D$, but in such a way that the union $\gamma = \gamma_1 \cup \gamma_2$ will link one of the polygonal circles forming the (translated) 1-skeleton K_s^1. Although at the first glance it seems clear that such a curve γ_2 exists, the selection is not totally trivial; it can be done however.

This linking contradicts the fact that any circle in D is contractible in the complement of γ, because γ lies in $\mathbb{R}^3 \setminus D$. The theorem follows.

6. Open problems.

Problem 1. Prove Theorem 1.1 in all dimensions $n \geq 3$. This can be accomplished if the next question admits a positive answer. Similarly, in that case one can replace double cone points with inner cone points.

Problem 2. Suppose that F is a compact set in $\mathbb{R}^n$, $n \geq 2$, and that $\varphi : F \to \varphi(F)$ is a quasisymmetric embedding of F into $\mathbb{R}^n$. Is it true that the n-measure of $\varphi(F)$ is zero if the n-measure of F is zero?

The proof in [H] of Theorem 1.1 would not only work in all dimensions $n \geq 3$ but it would also tremendously simplify, should the answer to this question be *yes*. In particular, no Sullivan theory of Lipschitz approximations is needed. Note that the answer to the question is *no* if $n = 1$.

Quasisymmetric maps are defined in (3.13), and their basic theory can be found in [TV1], [V2].

Problem 3. What version, if any, of Theorem 1.2 remains true if we only assume that ∂D be of finite $\mathcal{H}_2$-measure? By using the tree construction amended by a certain bubble blowing procedure, it is not hard to construct a quasiconformal Jordan domain D in $\mathbb{R}^n$, $n \geq 3$, such that the boundary ∂D has a purely unrectifiable piece of positive $\mathcal{H}_{n-1}$-measure that transforms onto a set of Hausdorff dimension zero under a quasiconformal map $f : D \to \mathbb{B}^n$, and that the complement $\mathbb{R}^n \setminus D$ has no points of n-density on ∂D. However, I have only been able to construct D in such a way that its boundary has σ-finite Hausdorff $\mathcal{H}_{n-1}$-measure.

Problem 4. Let f be a quasiconformal map of $\mathbb{B}^n$ onto a Jordan domain D in $\mathbb{R}^n$, $n \geq 3$, and suppose that the boundary of D has finite $\mathcal{H}_{n-1}$-measure. When is $f^{-1} : \partial D \to \partial\mathbb{B}^n$ absolutely continuous? The best known result to the author is the case when ∂D is $(n-1)$-regular as defined in (3.12). Then f is not only absolutely continuous, but it induces a measure that is A_∞ related to $\mathcal{H}_{n-1}$. This result is essentially due to Gehring. (See [S, 3.4] or [H, 2.7]). From the point of view of boundary behavior, regularity is a strong assumption. It does not cover, for instance, maps that can be extended to global quasiconformal maps of $\mathbb{R}^n$.

PROBLEM 5. Prove the Wall Conjecture in all dimensions. Related to this, Jussi Väisälä has proposed the following *generalized Wall Conjecture*, abbreviated $\mathrm{WC}(n,p)$, for all integers $n \geq 2$ and $1 \leq p \leq n-2$. Suppose that G is a homologically trivial open set in $\mathbb{R}^n$, $n \geq 2$, and suppose that $\mathbb{R}^n \cup \{\infty\} \setminus G$ is *inner* (k,c)*-joinable* for all $0 \leq k \leq p-1$. Then the conjecture $\mathrm{WC}(n,p)$ states that

$$\mathcal{H}_{p+1}\big(B(x, 2\,\mathrm{dist}(x,\partial G)) \cap \partial G\big) \;\geq\; C(c,n)\,\mathrm{dist}(x,\partial G)^{p+1} \tag{6.1}$$

for $x \in G$. The notion of inner joinability was introduced by Väisälä in [V5], where we refer the reader for a precise definition. It suffices to say here that the inner $(0,c)$-joinability is precisely the c-LLC_2 condition. Thus Väisälä's Theorem 5.1 implies that $\mathrm{WC}(n,1)$ is true. It is also not hard to see that $\mathrm{WC}(n,0)$ is true; note that in this case the second requirement about joinability becomes empty, while the first requirement about G being homologically trivial implies the connectivity of the complement of G.

All other cases of $\mathrm{W}C(n,p)$ are open. The wall conjecture as stated in (1.8) would follow from $\mathrm{W}C(n,n-2)$, because Väisälä has shown in [V5] that the complement of a quasiconformal ball is (k,c)-joinable for all $0 \leq k \leq n-3$.

Acknowledgements. I wish to thank Fred Gehring, Pekka Koskela, Pertti Mattila, Steffen Rohde, Stephen Semmes, and Jussi Väisälä for helpful and inspiring discussions on the topics related to this paper. I also thank the referee for a careful reading of the manuscript and useful comments.

ADDED IN PROOF. After this paper was submitted, two relevant developments took place. First, Semmes (Semmes, Quasisymmetry, measure and a question of Heinonen, this issue) solved Problem 2 above;

its consequences are discussed in (Heinonen, A Theorem of Semmes and boundary absolute continuity in all dimensions, this issue). In particular, Problem 1 is now solved as well. Second, Väisälä (The Wall Conjecture on Domains in Euclidean Spaces, Preprint, University of Helsinki, 1996) solved the generalized Wall Conjecture as in Problem 5 above. As a joint consequence of the results of Semmes and Väisälä, Theorem 1.2 is true in all dimensions $n \geq 3$, verifying the conjecture made on page 6 before Theorem 1.5. Namely, assuming the Wall Conjecture, the case $n \neq 4$ is already proved in the present paper, and the case $n = 4$ can be handled by the aid of the aforementioned result of Semmes as in (Heinonen, this issue).

References.

[BM] Baernstein, A., Manfredi, J. J., Topics in quasiconformal mappings, Topics in modern harmonic analysis. *Proc. Seminar held in Torino and Milano* **II** (1982), 849-862.

[BA] Beurling, A., Ahlfors, L. V., The boundary correspondence under quasiconformal mappings. *Acta Math.* **96** (1956), 125-142.

[F] Federer, H., *Geometric Measure Theory.* Springer, 1969.

[FHM] Fernández, J. L., Heinonen, J., Martio, O., Quasilines and conformal mappings. *J. Analyse Math.* **52** (1989), 117-132.

[G1] Gehring, F. W., *The Hausdorff measure of sets which link in Euclidean space.* Contributions to Analysis: A Collection of Papers Dedicated to Lipman Bers. Academic Press, 1974.

[G2] Gehring, F. W., Lower dimensional absolute continuity properties of quasiconformal mappings. *Math. Proc. Camb. Phil. Soc.* **78** (1975), 81-93.

[GP] Greene, R. E., Petersen V. P., Little topology, big volume. *Duke Math. J.* **67** (1992), 273-290.

[GV] Gehring, F. W., Väisälä , J., The coefficients of quasiconformality of domains in space. *Acta Math.* **114** (1965), 1-70.

[Ha] Hanson, B., Quasiconformal analogues of a theorem of Smirnov. *Math. Scand.* **75** (1994), 133-149.

[H] Heinonen, J., The boundary absolute continuity of quasiconformal mappings. *Amer. J. Math.* **116** (1994), 1545-1567.

[HK] Heinonen, J., Koskela, P., The boundary distortion of a quasiconformal mapping. *Pacific J. Math.* **165** (1994), 93-114.

[KR] Koskela, P., Rohde, S., Hausdorff dimension and mean porosity. To appear in *Math. Ann.*

[MM] Malý, J., Martio, O., Lusin's condition (N) and mappings of the class $W^{1,n}$. *J. reine angew. Math.* **458** (1995), 19-36.

[M] Mattila, P., *Geometry of Sets and Measures in Euclidean Spaces.* Cambridge Studies in Advanced Mathematics **44**. Cambridge University Press, 1995.

[Re] Reshetnyak, Yu. G., Some geometric properties of functions and mappings with generalized derivatives (Russian). *Sibirsk. Mat. Zh.* **7** (1966), 886-919.

[R] Riesz, F., Riesz, M., Über Randwerte einer analytischen Funktionen. *Quatrième Congrès des Math. Scand. Stockholm* (1916), 27-44.

[S] Semmes, S., Bilipschitz mappings and strong A_∞-weights. *Ann. Acad. Sci. Fenn. Series A I Math.* **18** (1993), 211-248.

[TV1] Tukia, P., Väisälä, J., Quasisymmetric embeddings of metric spaces. *Ann. Acad. Sci. Fenn. Ser. A I Math.* **5** (1980), 97-114.

[TV2] Tukia, P., Väisälä, J., Lipschitz and quasiconformal approximation and extension. *Ann. Acad. Sci. Fenn. Ser. A I Math.* **6** (1981), 303-342.

[V1] Väisälä, J., *Lectures on n-dimensional quasiconformal mappings.* Lecture Notes in Math. **229**, Springer-Verlag, 1971.

[V2] Väisälä, J., Quasi-symmetric embeddings in Euclidean spaces. *Trans. Amer. Math. Soc.* **264** (1981), 191-204.

[V3] Väisälä, J., Quasimöbius maps. *J. Analyse Math.* **44** (1984/85), 218-234.

[V4] Väisälä, J., Quasiconformal maps and positive boundary measure. *Analysis* **9** (1989), 205-216.

[V5] Väisälä, J., Metric duality in Euclidean spaces. To appear in *Math. Scand.*

[V6] Väisälä, J., Unpublished notes. 1995.

Recibido: 25 de agosto de 1.995

Juha Heinonen*
University of Michigan
Department of Mathematics
Ann Arbor, MI 48109, U.S.A.
juha@math.lsa.umich.edu

* Supported in part by NSF and the Academy of Finland. The author is a Sloan fellow.

Revista Matemática Iberoamericana
Vol. 12, N.° 3, 1996

Quasisymmetry, measure and a question of Heinonen

Stephen Semmes

Abstract. In this paper we resolve in the affirmative a question of Heinonen on the absolute continuity of quasisymmetric mappings defined on subsets of Euclidean spaces. The main ingredients in the proof are extension results for quasisymmetric mappings and metric doubling measures.

1. Introduction.

If F is a subset of $\mathbb{R}^n$ and $g : F \to \mathbb{R}^n$ is a mapping, then we say that g is *quasisymmetric* if it is not constant and if there exists a homeomorphism $\eta : [0, \infty) \to [0, \infty)$ such that

$$(1.1)\quad |x - y| \le t\,|x - z| \quad \text{implies} \quad |g(x) - g(y)| \le \eta(t)\,|g(x) - g(z)|\,,$$

whenever $x, y, z \in F$. We shall sometimes say that g is η-quasisymmetric to be explicit, or we shall refer to η as *the function that governs the quasisymmetry of* g when we want to be specific but not explicit.

This condition is a little bit hard to digest at first, but it means that the mapping approximately preserves relative distances, even if it may distort distances in an unbounded manner. In other words, if x is a lot closer to y than to z, then the corresponding property for $g(x)$, $g(y)$, and $g(z)$ should also hold, even though the distances themselves may change dramatically. For instance, the mapping defined by $g(x) = ax$

is η-quasisymmetric with $\eta(t) \equiv t$ for all positive numbers a, but this mapping distorts distances strongly when a is very large or very small.

See [TV] for basic facts about quasisymmetric mappings.

In the case of mappings defined on all of $\mathbb{R}^n$ the quasisymmetry condition is equivalent to the more famous quasiconformal condition, which is an infinitesimal version of the same idea. It turns out that quasisymmetric mappings on $\mathbb{R}^n$ send sets of measure zero to sets of measure zero when $n > 1$, see [V1]. This is not true when $n = 1$, because of an example in [BA].

Problem 1.2. (Juha Heinonen.) *If F is a compact subset of $\mathbb{R}^n$, $n > 1$, and $g : F \to \mathbb{R}^n$ is quasisymmetric, is it true that $g(F)$ has Lebesgue measure zero if F has Lebesgue measure zero?*

We shall see that the answer is yes. The proof will not give a new approach to the result for global quasisymmetric mappings, instead it will work by reducing to a method of Gehring [G] for the global case. Note however that quasisymmetric maps defined on subsets of $\mathbb{R}^n$ need not extend to global quasisymmetric mappings, so that the most obvious path to reducing to the global case is not available to us.

It will be more convenient to use the following reformulation of this problem.

Theorem 1.3. *Let F be a compact subset of $\mathbb{R}^n$, $n > 1$, and suppose that $g : F \to \mathbb{R}^n$ is quasisymmetric. Then $g(F)$ has positive Lebesgue measure if F has positive Lebesgue measure.*

Let us check that this resolves Problem 1.2.

Lemma 1.4. *If $g : F \to \mathbb{R}^n$ is quasisymmetric, then $g^{-1} : g(F) \to \mathbb{R}^n$ makes sense and is quasisymmetric.*

This is well-known and easy, but let us go quickly through the proof for the sake of completeness. Our mapping g is injective if it is quasisymmetric, so that its inverse is well-defined. From (1.1) we have that

$$(1.5)\quad \eta(t)\,|g(x) - g(z)| < |g(x) - g(y)| \quad \text{implies} \quad t\,|x - z| < |x - y|.$$

One can sort this out to see that g^{-1} is quasisymmetric, but with $\eta(t)$ replaced by $(\eta^{-1}(1/t))^{-1}$. This proves the lemma.

To see that Theorem 1.3 implies a positive answer to Problem 1.2 one need only switch from g to g^{-1} using the lemma.

In order to prove Theorem 1.3 we shall make some modifications to g and F. It would be simpler if we could just extend g to a mapping on all of $\mathbb{R}^n$, but this is not possible in general. Our plan will be to replace g with a map which lives on a thick set, and then to show that the pull-back of Lebesgue measure under this mapping behaves well.

The modifications of g will proceed in steps. Basically we want to progressively thicken the domain F of g. We begin with a definition.

Definition 1.6. *Let F_0, F be subsets of $\mathbb{R}^n$, with $F_0 \subseteq F$. We shall say that F_0 is a serious subset of F if there exists a constant $C > 0$ so that if $x \in F_0$ and $0 < t < \operatorname{diam} F_0$, then there is a point $y \in F$ such that*

$$C^{-1} t \leq |x - y| \leq t \,. \tag{1.7}$$

We say that F is serious if it is serious as a subset of itself.

This is a mild nondegeneracy condition which forbids isolated islands in a quantitative and uniform way. This is useful for the quasisymmetry condition (1.1), which provides information only about *relative* distances.

The property of a set being serious has been considered before under various names (unknown to the author until it was too late) such as "uniformly perfect" and "homogeneously dense", and it is a special case of the thickness conditions discussed in [VVW]. It may be that the relative property for subsets was not considered before.

We are going to be working with serious sets, and it would be nice if we could find a serious set of positive measure inside any given set of positive measure. Unfortunately this turns out not to be true, Pertti Mattila tells me that there are counterexamples. The following simple observation will suffice for our purposes.

Lemma 1.8. *Let F be a compact subset of $\mathbb{R}^n$ with positive measure. Then for each $\varepsilon > 0$ there is a compact subset F_0 of F such that F_0 is a serious subset of F and $|F_0| > |F| - \varepsilon$.*

We do not give bounds on the seriousness constant here.

To prove this we use points of density and Egoroff's theorem. From Lebesgue's theorem we know that almost every element of F is a point of density of F. That is,

$$\lim_{j\to\infty} \frac{|F \cap B(x, 2^{-j})|}{|B(x, 2^{-j})|} = 1$$

for almost all $x \in F$. Let $\varepsilon > 0$ be given. By Egoroff's theorem we can find a measurable subset F_0 of F on which we have uniform convergence for this limit and $|F_0| > |F| - \varepsilon$. We can take F_0 to be compact because we can always replace it, if necessary, with a compact subset with almost the same measure.

Uniform convergence implies that there is a $\delta > 0$ such that

$$\frac{|F \cap B(x, 2^{-j})|}{|B(x, 2^{-j})|} \geq \frac{1}{2} \tag{1.9}$$

when $x \in F_0$ and $0 < 2^{-j} < \delta$. It is not hard to see that this implies that F_0 is a serious subset of F, but with a horrible constant which depends on δ. (At scales finer than δ the constant is bounded. In other words, we could control the seriousness constant if we were willing to give up control on the measure.) This proves Lemma 1.8.

Of course (1.9) is much stronger than seriousness, but seriousness is a more natural condition for most of what we shall do.

Given a quasisymmetric mapping defined on some set we would like to modify it to get a mapping which is defined on a thicker set. The next result will be the first step of such a process, and then we shall go another step afterwards.

Proposition 1.10. *Suppose that F is a closed subset of $\mathbb{R}^n$, that $g : F \to \mathbb{R}^n$ is quasisymmetric, and that F_0 is a closed serious subset of F. Then we can find a serious closed set F^* in $\mathbb{R}^n$ which contains F_0 (but need not be contained in F) and a quasisymmetric mapping $g^* : F^* \to \mathbb{R}^n$ such that $g^* = g$ on F_0. The seriousness constant for F^* and the function η^* which controls the quasisymmetry of g^* are controlled in terms of the dimension, the seriousness constant for (F_0, F), and the function η that controls the quasisymmetry of g.*

The point here is that F^* is serious as a set unto itself, not as a subset of something else.

Before stating the next thickening result we need another definition.

Definition 1.11. *A closed set E of $\mathbb{R}^n$ is said to be a strong set if there is a constant $C > 0$ so that for each $x \in \mathbb{R}^n \backslash E$ there is a $y \in E$ such that*

$$|x - y| \le C \operatorname{dist}(x, E) \tag{1.12}$$

and

$$\operatorname{dist}(y, \mathbb{R}^n \backslash E) \ge C^{-1} \operatorname{dist}(x, E)\,. \tag{1.13}$$

In other words, a strong set is always approximately at least as big as its complement.

Proposition 1.14. *Suppose that F is a serious closed subset of $\mathbb{R}^n$ and that $g : F \to \mathbb{R}^n$ is quasisymmetric. Then there is a strong set $S \subseteq \mathbb{R}^n$ such that $S \supseteq F$ and g admits an extension to a quasisymmetric mapping $G : S \to \mathbb{R}^n$. The strongness constant for S and the function which governs the quasisymmetry of G can be chosen to depend only on the function that governs the quasisymmetry of g, the seriousness constant for F, and the dimension n.*

We shall need to know that the image is a strong set, and there is a general result to this effect.

Proposition 1.15. *If $G : S \to \mathbb{R}^n$ is quasisymmetric and S is a strong subset of $\mathbb{R}^n$, then so is $G(S)$, with a constant that depends only on the dimension, the strongness constant for S, and the function which governs the quasisymmetry of G.*

It may not be clear that these statements reflect progress, but the they do, and this is manifested in part by the following fact, which says that strong sets are large measure-theoretically.

Proposition 1.16. *If S is a strong subset of $\mathbb{R}^n$, then there is a constant $C > 0$ so that*

$$|S \cap B(x, r)| \ge C^{-1} r^n\,, \tag{1.17}$$

for all $x \in S$ and $r > 0$. C depends only on the dimension and the strongness constant of S.

Here $|A|$ denotes the Lebesgue measure of a set A.

The next point is to convert from mappings to measures. We begin with some definitions.

Definition 1.18. *Let E be a closed subset of $\mathbb{R}^n$, and let μ be a Borel measure with support equal to E.*

a) *We say that μ is doubling on E if there is a constant $C > 0$ so that*

$$\mu(B(x,2r)) \leq C\,\mu(B(x,r))\,, \tag{1.19}$$

for all $x \in E$ and $0 < r < \operatorname{diam} E$.

b) *Define $\delta(x,y) = \delta_\mu(x,y)$ for $x, y \in E$ by*

$$\delta(x,y) = \big(\mu(B(x,|x-y|) \cup B(y,|x-y|))\big)^{1/n}\,. \tag{1.20}$$

We say that μ is a metric doubling measure on E if μ is doubling on E and if there is a true metric $d(x,y)$ on E –i.e., a symmetric nonnegative function which vanishes exactly on the diagonal and which satisfies the triangle inequality– and a constant $C > 0$ such that

$$C^{-1}\,d(x,y) \leq \delta(x,y) \leq C\,d(x,y)\,, \tag{1.21}$$

for all $x, y \in E$.

These are good classes of measures for studying quasisymmetric mappings. The notion of metric doubling measures comes from [DS], in a slightly different form, see also [S1].

Proposition 1.22. *If $G : S \to \mathbb{R}^n$ is quasisymmetric and S is a strong subset of $\mathbb{R}^n$, then the measure μ on $\mathbb{R}^n$ defined by $\mu(A) = |G(A \cap S)|$ is a metric doubling measure on S, with constants that depend only on n, the strongness constant for S, and the function which governs the quasisymmetry of G.*

This is exactly the measure that we are interested in for Theorem 1.3. The question now is what more we can say about it.

Proposition 1.23. *If S is a strong subset of $\mathbb{R}^n$ and μ is a metric doubling measure on S, then there is a metric doubling measure ν on $\mathbb{R}^n$ which agrees with μ on subsets of S. The metric doubling constants* *for ν are controlled in terms of the corresponding constants for μ, the strongness constant for S, and the dimension n.*

This is what we want because of the following absolute continuity result.

Theorem 1.24. *If μ is a metric doubling measure on $\mathbb{R}^n$ and $n > 1$, then μ and Lebesgue measure are absolutely continuous with respect to each other.*

This result was basically proved by Gehring [G]. He did not state it this way, but his argument gives this result with little extra effort. This extension of Gehring's result was observed in [DS]. See Proposition 3.4 of [S1] for a detailed argument for this form of the result.

If μ is a metric doubling measure on $\mathbb{R}^n$, $n > 1$, then the density of μ is an "A_∞ weight", which gives a uniform and scale-invariant version of absolute continuity. In other words Theorem 1.24 comes with quantitative estimates.

The original point of Gehring's argument was to get information about the jacobian of a global quasisymmetric mapping on $\mathbb{R}^n$. We are doing roughly the same thing here, except that we are exploiting some flexibility in metric doubling measures that quasisymmetric mappings do not enjoy. Specifically, in Proposition 1.23 we have an extension result which does not have a counterpart for quasisymmetric mappings. There are no topological obstructions to building extensions of metric doubling measures.

Not all metric doubling measures on $\mathbb{R}^n$ arise from global quasisymmetric mappings in the manner described above. See [S2] for counterexamples.

Let us now summarize some of the main conclusions of these propositions.

Theorem 1.25. *Suppose that F is a closed subset of $\mathbb{R}^n$, that $g : F \to \mathbb{R}^n$ is quasisymmetric, and that F_0 is a closed serious subset of F. Then there is a metric doubling measure ν on $\mathbb{R}^n$ such that $\nu(A) = |g(A)|$ for all Borel subsets of F_0. In particular $|g(A)| = 0$ if and only if $|A| = 0$ when $A \subset F_0$, by Theorem 1.24. The metric doubling constants for ν depend only on n, the seriousness constant for (F_0, F), and the function which governs the quasisymmetry of g.*

Indeed, under these conditions we can use Proposition 1.10 to extend the restriction of g to F_0 to a quasisymmetric mapping on a serious set, and then we can use Proposition 1.14 to extend to a strong set. This permits us to reduce to the case where F_0 is a strong set. We then use Propositions 1.22 and 1.23 to get a metric doubling measure, first on the strong set, and then on all of $\mathbb{R}^n$. This proves Theorem 1.25, modulo the previous propositions.

Theorem 1.3 is an immediate consequence of Lemma 1.8 and Theorem 1.25. Thus we need only prove the various propositions. They are slightly messy, but all pretty straightforward, and largely implicit in the literature, if not explicitly stated in the form that we need. For the sake of readability we shall often provide more detail than needed for experts in the area, and we shall sometimes treat issues with bare hands instead of sending the reader to the literature for lemmata.

Related papers concerning quasisymmetric mappings include [TV], [V2], and [V3].

Although Propositions 1.10 and 1.14 look very similar, they really aren't, in the sense that Proposition 1.10 is much closer to the definitions, whereas the proof of Proposition 1.14 relies on the structure of Euclidean space.

2. The proof of Proposition 1.10.

This is quite straightforward. We are going to take F_0, take a reasonably dense but scattered subset of $F \backslash F_0$, replace g by something simple on little disks centered at points in this scattered subset, and that will do the job. Our first task is to find this reasonably dense but scattered subset. We shall employ this well-known construction again in the next section.

Lemma 2.1. *Let E be a closed subset of $\mathbb{R}^n$, and let H be a subset of $\mathbb{R}^n \backslash E$. Then we can find a subset I of H such that*

(2.2) *for every $x \in H$ there is a point $u \in I$ such that*

$$|x - u| \leq \frac{1}{2} \operatorname{dist}(x, E),$$

and

(2.3) *for every y, $z \in I$ we have that* $|y - z| \geq \dfrac{1}{3} \operatorname{dist}(y, E)$.

Thus I is reasonably dense in H and also reasonably scattered.

Let E and H be given, and let I be a maximal subset of H which satisfies (2.3). It is not hard to find such a maximal subset. For instance, one can write $\mathbb{R}^n \backslash E$ as the increasing union of compact sets K_j, one can build sets I_j recursively by taking I_{j+1} to be the maximal subset of $H \cap K_{j+1}$ which satisfies (2.3) and contains I_j, and then take I to be the union of the I_j's. In each compact part the maximal subset has to be finite, which makes it easier to verify its existence, and then the pieces nest together properly to give maximality for the union.

Thus we can take I to be a maximal subset of H which satisfies (2.3). Let $x \in H$ be given. Either $x \in I$ already, or it is not, in which case $I \cup \{x\}$ will not satisfy (2.3). This means that there is a point $u \in I$ such that

$$\text{(2.4)}\quad \text{either} \quad |x-u| < \frac{1}{3}\,\text{dist}\,(x,E) \quad \text{or} \quad |x-u| < \frac{1}{3}\,\text{dist}\,(u,E)\,.$$

In the first case we get (2.2) directly. In the second case we compute that

$$\text{(2.5)}\quad \text{dist}\,(u,E) \le |x-u| + \text{dist}\,(x,E) \le \frac{1}{3}\,\text{dist}\,(u,E) + \text{dist}\,(x,E)$$

to conclude that $\text{dist}\,(u,E) \le 3\,\text{dist}\,(x,E)/2$, and hence that (2.2) holds. This proves Lemma 2.1.

Lemma 2.6. *Let E, H, and I be as in Lemma 2.1, and set $B(x) = \overline{B}(x, 20^{-1}\text{dist}\,(x,E))$ when $x \in I$. If $x \in I$ and $y \in 2\,B(x)$, then*

$$\text{(2.7)}\qquad \frac{9}{10}\,\text{dist}\,(x,E) \le \text{dist}\,(y,E) \le \frac{11}{10}\,\text{dist}\,(x,E)\,.$$

If $x, z \in I$ and $x \neq z$ then

$$\text{(2.8)}\qquad 2\,B(x) \cap 2\,B(z) = \varnothing\,.$$

Indeed, if $x \in I$ and $y \in 2\,B(x)$, then

$$\text{(2.9)}\qquad |\text{dist}\,(y,E) - \text{dist}\,(x,E)| \le 10^{-1}\text{dist}\,(x,E)\,.$$

This implies (2.7).

Now suppose that $x, z \in I$ and $x \neq z$, but that (2.8) fails to hold, so that there is a point y in the intersection. Then

$$(2.10)\qquad |x-z| \leq |x-y| + |y-z| \leq \frac{1}{10}\operatorname{dist}(x,E) + \frac{1}{10}\operatorname{dist}(z,E)$$

and

$$(2.11)\qquad \operatorname{dist}(z,E) \leq \frac{10}{9}\operatorname{dist}(y,E) \leq \frac{11}{9}\operatorname{dist}(x,E)\,,$$

because of (2.7) (applied to both x and z). Combining these we get that

$$(2.12)\qquad |x-z| < \frac{1}{3}\operatorname{dist}(x,E),$$

in contradiction to (2.3). This proves (2.8), and the lemma follows.

Let us now prove Proposition 1.10. Let g, F, F_0 be as given there, and apply Lemma 2.1 with $E = F_0$ and $H = F\backslash F_0$. We get a subset I of F.

Define F^* by

$$(2.13)\qquad F^* = F_0 \cup \Big(\bigcup_{x\in I} B(x)\Big)\,,$$

where $B(x)$ is as in Lemma 2.6, with $E = F_0$. We shall define g^* a little later. Let us first verify some simple properties of F^*.

Lemma 2.14. *F^* is closed.*

Let $\{z_j\}$ be a sequence of points in F^* which converges to some point $z \in \mathbb{R}^n$. We have to show that $z \in F^*$. If there is a subsequence of $\{z_j\}$ which is contained in F_0, then $z \in F_0$, and $z \in F^*$. If $\{z_j\}$ has a subsequence which is contained in any one of the $B(x)$'s, then z lies in the same $B(x)$, and hence in F^*. The remaining possibility is that there is a subsequence of $\{z_j\}$ such that each term lies in a different $B(x)$. Since $\{z_j\}$ converges and hence is bounded, we must have that the elements of this subsequence accumulate on F_0, because of the way that we defined the $B(x)$'s. In this case we conclude that $z \in F_0$ and hence $z \in F^*$. This proves the lemma.

Lemma 2.15. $\operatorname{diam} F^* \leq 2 \operatorname{diam} F$.

If $p \in F^*$, then either $p \in F_0 \subseteq F$, or $p \in B(x)$ for some $x \in I$. In the latter case we have that

$$\operatorname{dist}(p, F) \leq |p - x| \leq 10^{-1} \operatorname{dist}(x, F_0) \leq 10^{-1} \operatorname{diam} F\,, \tag{2.16}$$

since $x \in F$. This implies the desired bound for $\operatorname{diam} F^*$.

Lemma 2.17. *For each point $x \in F$ there is a point $u \in F^*$ such that $|x - u| \leq \operatorname{dist}(x, F_0)/2$.*

This is trivial. Either $x \in F_0$, in which case we take $u = x$, or not, in which case we take $u \in I$ as in (2.2) (with $E = F_0$). This gives the lemma.

Lemma 2.18. *F^* is serious.*

Let $p \in F^*$ and $0 < t < \operatorname{diam} F^*$ be given, and let us try to find a point $q \in F^*$ with

$$C^{-1}\, t \leq |p - q| \leq C\, t \tag{2.19}$$

for a suitable constant C. We may as well assume that $t \leq \operatorname{diam} F$, since otherwise we can use Lemma 2.15 to reduce the problem to the definition of $\operatorname{diam} F^*$.

Suppose first that $p \in F_0$. The we can use the assumption that F_0 is a serious subset of F to find a point $x \in F$ such that $C^{-1}\, t \leq |p-x| \leq t$. Lemma 2.17 provides a point $u \in F^*$ such that $|x-u| \leq \operatorname{dist}(x, F_0)/2 \leq |p - x|/2$. Thus $|p - u| \leq |p - x| + |x - u| \leq 2\, t$, which gives the upper bound in (2.19) (with $q = u$). For the lower bound we have that

$$|p - x| \leq |p - u| + |u - x| \leq |p - u| + \frac{1}{2}\, |x - p|\,, \tag{2.20}$$

and hence $|p - x|/2 \leq |p - u|$. This gives the lower bound in (2.19).

Now suppose that $p \in B(z)$ for some $z \in I$. If $t \leq \operatorname{dist}(z, F_0)$, then we can find the required q inside $B(z)$. If $t > \operatorname{dist}(z, F_0)$, then let y be a point in F_0 such that $|y - z| = \operatorname{dist}(z, F_0)$. Choose $x \in F$ so that $C^{-1}\, t \leq |y-x| \leq t$, as we can do because of the seriousness of F_0 inside of F. Let $u \in F^*$ be associated to x as in Lemma 2.17. Then

$$\begin{aligned} |p - u| &\leq |p - z| + |z - y| + |y - x| + |x - u| \\ &\leq t + t + t + \operatorname{dist}(x, F_0) \\ &\leq 3\, t + |x - y| \leq 4\, t\,. \end{aligned} \tag{2.21}$$

This gives the upper bound that we want for (2.19) (with $q = u$). For the lower bound we observe that

$$
\begin{aligned}
|y-x| &\le |p-u| + |p-y| + |x-u| \\
&\le |p-u| + (|p-z| + |z-y|) + \frac{1}{2}\operatorname{dist}(x, F_0) \\
&\le |p-u| + \operatorname{dist}(z, F_0) + \operatorname{dist}(z, F_0) + \frac{1}{2}|x-y| \\
&\le |p-u| + 2\operatorname{dist}(z, F_0) + \frac{1}{2}|x-y|\,.
\end{aligned} \tag{2.22}
$$

Thus $C^{-1}\,t \le |y-x| \le 2\,|p-u| + 4\operatorname{dist}(z, F_0)$. If t is much larger that $\operatorname{dist}(z, F_0)$ then this implies the lower bound in (2.19). If not, then again we simply take a suitable q in $B(z)$. This proves Lemma 2.18.

Let us now define $g^* : F^* \to \mathbb{R}^n$. Of course we set $g^* = g$ on F_0, and we define g^* on each $B(x)$ as follows. Given $x \in I$ choose a point $\pi(x) \in F_0$ so that

$$
|x - \pi(x)| = \operatorname{dist}(x, F_0)\,. \tag{2.23}
$$

Define g^* on $B(x)$ by

$$
g^*(w) = g(x) + a\,\frac{|g(x) - g(\pi(x))|}{|x - \pi(x)|}\,(w - x)\,, \tag{2.24}
$$

for all $w \in B(x)$. Here a is a small positive number to be chosen in the next lemma. Thus on the ball $B(x)$ we have taken g^* to be a similarity with the same value as g at the center and whose distortion ratio is approximately the same as that of g at that location and scale.

Let $\beta(x)$, $x \in I$, denote the ball which is the image of $B(x)$ under g^*. Thus

$$
\beta(x) = B(g(x), 20^{-1}\,a\,|g(x) - g(\pi(x))|), \tag{2.25}
$$

by the definition of g^* and $B(x)$ (in Lemma 2.6).

Lemma 2.26. *If a is small enough, depending only on the function which governs the quasisymmetry of g, then the balls $2\,\beta(x)$, $x \in I$, are pairwise disjoint and each is disjoint from $g^*(F_0) = g(F_0)$.*

This is just a question of the quasisymmetry condition. Suppose that $y, z \in I$, $y \neq z$. Using (2.3) we get that $|y - \pi(y)| \leq 3\,|y - z|$, and similarly we have that $|z - \pi(z)| \leq 3\,|y - z|$. Quasisymmetry then implies that

$$|g(y) - g(\pi(y))| + |g(z) - g(\pi(z))| \leq C\,|g(y) - g(z)|\,. \tag{2.27}$$

This implies that $2\,\beta(y)$ and $2\,\beta(z)$ are disjoint if a is small enough.

Now suppose that $x \in I$ and $w \in F_0$. We want to show that $g(w) \notin 2\,\beta(x)$. We have that $|x - \pi(x)| = \operatorname{dist}(x, F_0) \leq |x - w|$, by definition of $\pi(x)$, and so

$$|g(x) - g(\pi(x))| \leq C\,|g(x) - g(w)|\,, \tag{2.28}$$

by quasisymmetry. This implies that $g(w) \notin 2\,\beta(x)$ if a is small enough.

This proves Lemma 2.26. Fix now a choice of a as above, depending only on the function that governs the quasisymmetry of g.

It remains to prove that g^* is quasisymmetric. The argument for this has some generality, and we shall need it again later, and so we formulate it in more general terms than required for the present circumstances.

Lemma 2.29. *Let A be a closed subset of $\mathbb{R}^n$, and let $\{B_i\}_{i\in I}$ and $\{\beta_i\}_{i\in I}$ be collections of closed balls in $\mathbb{R}^n$. Set $A^* = A \cup \bigcup_{i\in I} B_i$ and let A' denote the union of A and the set of centers of the balls B_i, $i \in I$.*

Suppose that $H : A^* \to \mathbb{R}^n$ has the property that the restriction of H to A' is quasisymmetric, that $H(B_i) = \beta_i$ for each $i \in I$, and that the restriction of H to each B_i is a quasisymmetric mapping with a function governing the quasisymmetry that can be taken to be independent of i. Suppose also that the balls $2\,B_i$, $i \in I$, are pairwise disjoint and are disjoint from A, that the balls $2\,\beta_i$, $i \in I$, are pairwise disjoint and disjoint from $H(A)$, and that there is a constant $C > 0$ so that

$$C^{-1}\operatorname{dist}(B_i, A) \leq \operatorname{radius} B_i \leq \operatorname{dist}(B_i, A), \tag{2.30.a}$$

and

$$C^{-1}\operatorname{dist}(\beta_i, H(A)) \leq \operatorname{radius} \beta_i \leq \operatorname{dist}(\beta_i, H(A)) \tag{2.30.b}$$

for all $i \in I$. (Note that the upper bounds follow from the disjointness of the $2\,B_i$'s from A, the $2\,\beta_i$'s from $H(A)$.)

Then $H : A^* \to \mathbb{R}^n$ is quasisymmetric, with bounds which depend only on a uniform choice of a function which governs the quasisymmetry of the various restrictions of H mentioned above, and on the constants in (2.30).

If we can prove this lemma then we get that g^* is quasisymmetric, because our balls have the correct disjointness properties and satisfy the analogue of (2.30) (by their definitions), because the restrictions of g^* to the various $B(x)$'s are trivially quasisymmetric, with uniform bounds, and because the restriction of g^* to $A' = F_0 \cup I$ agrees with g and hence is quasisymmetric.

Thus Proposition 1.10 will follow once we prove Lemma 2.29.

Beware of the small changes in notation from the previous situation to the lemma, $B(x)$ to B_i, etc.

The lemma is a straightforward but unpleasant exercise, a matter of checking cases. Let A, H, etc. be as above.

Let us first record a small observation.

Sublemma 2.31. *Suppose that $p, q \in B_i$ and $w \in A^* \backslash B_i$. Then*

$$C^{-1}\,|q-w| \le |p-w| \le C\,|q-w|\,, \tag{2.32}$$

$$C^{-1}\,|H(q)-H(w)| \le |H(p)-H(w)| \le C\,|H(q)-H(w)|\,, \tag{2.33}$$

$$|p-w| \ge C^{-1}\mathrm{diam}\,B_i \tag{2.34}$$

and

$$|H(p)-H(w)| \ge C^{-1}\mathrm{diam}\,\beta_i\,, \tag{2.35}$$

for a suitable constant C.

This follows from the our assumptions, which ensure that $2\,B_i$ is disjoint from $A^* \backslash B_i$, and that $2\,\beta_i$ is disjoint from $H(A^* \backslash B_i)$.

Suppose that we are given $x, y, z \in A^*$ and $t > 0$ which satisfy $|x-y| \le t\,|x-z|$. We want to show that $|H(x)-H(y)| \le \theta(t)\,|H(x)-H(z)|$ for some $\theta(t)$ which tends to 0 when $t \to 0$ and which is bounded on finite intervals. (See Lemma 2.42 below for a small technical point here.)

If x, y, z all lie in A', or all lie in some B_i, then we get the desired bound from our hypotheses.

If no two of x, y, z lie in the same B_i, then we can reduce to the previous case where x, y, z all lie in A', by using Sublemma 2.31 to switch from a point in some B_i to the center of B_i. That is, such a change will not affect any of the distances involved by more than a bounded factor. (Remember that A' consists exactly of A and the various centers of the B_i's.)

Thus we may assume that exactly two of x, y, z lie in some B_i, and that the remaining point lies in $A^* \backslash B_i$.

We may as well assume that x is one of the two points that lies in B_i. For if it is not, then we can use Sublemma 2.31 to reduce to the case where y and z are both equal to the center of B_i, and where x either lies in A or is the center of some other B_j. Again these changes will not affect the relevant distances by more than a bounded factor. After these changes all three points would lie in A', which is already covered by our assumptions.

Thus we may assume that x lies in B_i, and that exactly one of y and z do too. We may also assume that the remaining point lies in A', because Sublemma 2.31 again permits us to make the substitution without affecting the quantities involved by more than a bounded factor.

In order to deal with this remaining situation we make another small observation.

Sublemma 2.36. *For each $i \in I$ let c_i denote the center of B_i, and choose $\xi_i \in A$ such that $|c_i - \xi_i| = \operatorname{dist}(c_i, A)$. Then*

$$C^{-1} \operatorname{diam} B_i \leq |c_i - \xi_i| \leq C \operatorname{diam} B_i \tag{2.37}$$

and

$$C^{-1} \operatorname{diam} \beta_i \leq |H(c_i) - H(\xi_i)| \leq C \operatorname{diam} \beta_i \tag{2.38}$$

for each i and a suitable constant C.

The bounds (2.37) follow from (2.30) and the definitions of c_i and ξ_i. (See also Sublemma 2.31.)

As for (2.38), notice that $|H(c_i) - H(\xi_i)|$ is comparable in size to $\operatorname{dist}(H(c_i), H(A))$, because of quasisymmetry and the fact that $|c_i - \xi_i| = \operatorname{dist}(c_i, A)$. This implies (2.38), because of (2.30) again.

This proves Sublemma 2.36.

Let us come back to our original problem of the quasisymmetry of H. We have our three points x, y, z with $|x - y| \le t\,|x - z|$, and we want to prove something like $|H(x) - H(y)| \le \theta(t)\,|H(x) - H(z)|$. We have already reduced to the case where x and exactly one of y and z lies in some B_i, and where the remaining point lies in $A' \backslash B_i$.

This last situation is slightly obnoxious because it is really a combination of two cases. For the sake of explanation suppose that it is y which lies in B_i. Then we could have that $|x - y|$ is very small compared to the radius of B_i, and that $\operatorname{dist}(z, B_i)$ is large compared to the radius of B_i. In order to establish quasisymmetry we should show that such a circumstance leads to something similar in the image. It is more convenient however to do this in two steps, first to compare $|x-y|$ with the radius of B_i and make a similar comparison in the image, and then to compare $\operatorname{dist}(z, B_i)$ with the radius of B_i and to make a similar comparison in the image. Our final estimate will be obtained as a product of estimates from these two parts.

Assume first that $y \in B_i$, so that $z \in A' \backslash B_i$. Set

$$\begin{aligned} &r = \frac{|x - y|}{|c_i - \xi_i|}, && s = \frac{|c_i - \xi_i|}{|x - z|}, \\ &R = \frac{|H(x) - H(y)|}{|H(c_i) - H(\xi_i)|}, && S = \frac{|H(c_i) - H(\xi_i)|}{|H(x) - H(z)|}. \end{aligned} \tag{2.39}$$

By assumption we have that $r\,s \le t$, and we want to bound $R\,S$ by a function of t which tends to 0 as $t \to 0$.

Sublemma 2.40. *$r, s, R, S \le C$ for some constant C.*

For r and R this follows from Sublemma 2.36 and the fact that $x, y \in B_i$, $H(x), H(y) \in \beta_i$. For s and S we observe that $z \notin B_i$, $H(z) \notin \beta_i$, so that Sublemma 2.31 can be applied. With this observation the bounds for s and S follow from Sublemma 2.36 also. This proves Sublemma 2.40.

Since $r\,s \le t$ we get that one of r and s is $\le \sqrt{t}$. Our quasisymmetry hypotheses imply that the corresponding R or S is bounded by a good function of $\sqrt{t}$. (For r we have to use (2.37) to get to the quasisymmetry of H on B_i.) We conclude that $R\,S$ is bounded by a good function of $\sqrt{t}$, since they are each bounded separately. This is the bound that we need.

Assume now that $z \in B_i$, so that $y \in A' \backslash B_i$. Define r, s, R, S as above. Again we have $r\,s \leq t$ by assumption, and we want to control $R\,S$.

Sublemma 2.41. *$r, s, R, S \geq C^{-1}$ for some constant C.*

This is practically the same as Sublemma 2.40, but with the roles of y and z reversed.

In this case we can conclude that each of r and s is bounded by a constant multiple of t. Our quasisymmetry hypotheses then imply that each of R and S is bounded by a function of t, and so the product is too.

This completes the proof of Lemma 2.29. Note that we have not given the most efficient estimates in the argument.

For the record, let us mention a small lemma which we have used implicitly.

Lemma 2.42. *Suppose that $\theta : [0,\infty) \to [0,\infty)$ satisfies $\theta(0) = 0$, $\theta(t)$ is continuous at 0, and θ is bounded on bounded sets. Then there is a homeomorphism $\Theta : [0,\infty) \to [0,\infty)$ such that $\theta(t) \leq \Theta(t)$ for all t.*

Indeed, following Väisälä we set $\Theta(t) = t + \sup_{0 \leq s \leq 2t} \theta(t)$ when $t = 2^n$, $n \in \mathbb{Z}$, and use affine interpolation to define Θ on the rest. (Thanks to Alestalo for pointing out the author's stupidity for the first version.)

3. The proof of Proposition 1.14.

The argument will parallel the proof of Proposition 1.10 in the previous section, except for one piece of information that we shall have to obtain for ourselves.

Let g and F be given, as in Proposition 1.14. Let I be as in Lemma 2.1, applied with $E = F$ and $H = \mathbb{R}^n \backslash F$. Let $B(x)$, $x \in I$, be defined as in Lemma 2.6 (with $E = F$). Define the set S by

$$S = F \cup \Big(\bigcup_{x \in I} B(x) \Big) , \tag{3.1}$$

as in (2.13).

In the next two lemmas we give basic properties of S. At this stage we do not use the assumption that F is serious, only that it is closed. The seriousness will not be used until we start to work with our quasisymmetric mapping.

Lemma 3.2. *S is closed.*

This is the same as Lemma 2.14, with only cosmetic changes.

Lemma 3.3. *S is a strong set.*

Let $x \in \mathbb{R}^n \backslash S$ be given, as in Definition 1.11. Thus $x \in \mathbb{R}^n \backslash F$. The point is that x must be reasonably close to $B(u)$ for some $u \in I$, but it is helpful to distinguish between the cases where x is very close to some $B(u)$ or never too close. Actually our threshold will be sufficiently generous that the latter never happens.

Suppose first that

$$\operatorname{dist}(x, S) \le \frac{1}{2} \operatorname{dist}(x, F) . \tag{3.4}$$

Choose $z \in S$ so that $|x - z| = \operatorname{dist}(x, S)$. Then $z \notin F$, and so $z \in B(u)$ for some $u \in I$. Because $|x - z| \le \operatorname{dist}(x, F)/2$ we get that

$$\frac{1}{2} \operatorname{dist}(x, F) \le \operatorname{dist}(z, F) \le \frac{3}{2} \operatorname{dist}(x, F) . \tag{3.5}$$

This means that $\operatorname{dist}(z, F)$ is comparable in size to the radius of $B(u)$, because of (2.7) in Lemma 2.6 and the definition of $B(u)$. Since $|x - z| = \operatorname{dist}(x, S)$ and $z \in B(u)$ it is easy to see that we can find a point y of the type required in Definition 1.11, inside $B(u)$ (and not just in S).

Now suppose that

$$\operatorname{dist}(x, S) > \frac{1}{2} \operatorname{dist}(x, F) . \tag{3.6}$$

In fact this cannot happen. Indeed, apply Lemma 2.1 to get a point $u \in I$ such that $|x-u| \le \operatorname{dist}(x,F)/2$, as in (2.2). Then

$$\operatorname{dist}(x,S) < |x-u| \le \frac{1}{2}\operatorname{dist}(x,F) < \operatorname{dist}(x,S)\,, \tag{3.7}$$

a contradiction.

This completes the proof of Lemma 3.3.

To prove Proposition 1.14 we need to build a quasisymmetric extension G of g. We would like to do this in the same way as in Section 2 (around (2.24)), but in the present situation we have the problem that g is not yet defined at the elements of I. The main point of the argument that follows will be to extend g quasisymmetrically to I. Once we do that we can proceed as in Section 2 (using Lemma 2.29).

The elements of I basically represent holes in F, large puddles of its complement. We need to show that these holes correspond to holes in the complement of $g(F)$ in a reasonable manner. The next couple of lemmas will enable us to do that.

Lemma 3.8. *Let a homeomorphism $\eta : [0,\infty) \to [0,\infty)$ and a dimension n be given. For each $\varepsilon > 0$ there exist $\delta > 0$ and $R > 1$, depending on ε, η, and n, with the following properties. Let E be a subset of $\mathbb{R}^n$ and $h : E \to \mathbb{R}^n$ be an η-quasisymmetric mapping which satisfy the normalizations*

$$0, u \in E \quad \text{and} \quad h(0) = 0,\ h(u) = u, \tag{3.9}$$

where $u = (1,0,\dots,0)$. Suppose that E is δ-thick in $B(0,R)$, in the sense that

$$\operatorname{dist}(x,E) \le \delta \quad \text{whenever } x \in B(0,R). \tag{3.10}$$

Then $h(E \cap B(0,R))$ is ε-thick in $B(0,1)$, so that

$$\operatorname{dist}(z,h(E)) \le \varepsilon \quad \text{whenever } z \in B(0,1). \tag{3.11}$$

This is a weaker version of [V3, Theorem 3.1], weaker by dint of having estimates which depend on the dimension and which are obtained through very nonconstructive means. For the reader's convenience we include a proof by compactness which is mentioned in the introduction of [V3] (and attributed to Tukia).

Suppose that Lemma 3.8 is not true. Then there exist η, n, and ε as above, a sequence $\{E_k\}$ of subsets of $\mathbb{R}^n$, and a sequence $\{h_k\}$ of η-quasisymmetric mappings from E_k into $\mathbb{R}^n$, such that $\{E_k\}$ and $\{h_k\}$ satisfy the analogues of the normalizations (3.9), each E_k is $1/k$-thick inside $B(0,k)$, but each $h_k(E_k \cap B(0,k))$ fails to be ε-thick inside $B(0,1)$.

Claim 3.12. *There is a sequence of integers k_j and an η-quasisymmetric mapping $H : \mathbb{R}^n \to \mathbb{R}^n$ such that the h_{k_j}'s converge to H "uniformly on compact sets" in the sense that*

$$\lim_{j\to\infty} \sup_{x\in B\cap E_{k_j}} |h_{k_j}(x) - H(x)| = 0\,, \tag{3.13}$$

for every ball B in $\mathbb{R}^n$.

This is pretty standard, but let us be careful.

The first step is to show that we have equicontinuity of the h_k's on compact sets. That is, for each ball B there exists a function $\omega_B : [0,\infty) \to [0,\infty)$ such that $\omega_B(0) = 0$, ω_B is continuous at 0, ω_B is bounded on finite intervals, and

$$|h_k(x) - h_k(y)| \le \omega_B(|x-y|)\,, \tag{3.14}$$

for all x, $y \in B \cap E_k$ and all k. This follows from the uniform quasisymmetry hypotheses and the normalizations.

Once we have this equicontinuity condition we can conclude that there is continuous mapping $H : \mathbb{R}^n \to \mathbb{R}^n$ and a subsequence $\{h_{k_j}\}$ of $\{h_k\}$ which converges to h in the sense of (3.13). This is not hard to prove, using an Arzela-Ascoli argument. Here is one way to do it from scratch. Let $\{p_m\}$ be a countable dense subset of $\mathbb{R}^n$. For each p_m choose a sequence of points $\{p_{m,k}\}_{k=1}^{\infty}$ such that $p_{m,k} \in E_k$ for each k and $|p_{m,k} - p_m| \le 1/k$ when $p_m \in B(0,k)$. We can do this because of our thickness hypotheses. Next choose the subsequence $\{h_{k_j}\}$ of $\{h_k\}$ in such a way that $\lim_{j\to\infty} h_{k_j}(p_{m,k_j})$ exists for each m, and call the result $H(p_m)$. We can find such a subsequence because of the usual Cantor diagonalization argument. We are also using our normalizations and the equicontinuity property (3.14) to know that $\{h_k(p_{m,k})\}_k$ is a bounded sequence for each m. Once one has $\{h_{k_j}\}$ with this property it is not hard to show that H must have a continuous extension to all of $\mathbb{R}^n$, and that we have convergence in the sense of (3.13), using the equicontinuity property (3.14).

It is easy to derive the η-quasisymmetry of H from the corresponding property of the h_k's.

This completes the proof of Claim 3.12.

Let us now finish the proof of Lemma 3.8. Let H be as in the claim. The main point now is that H must be surjective, $H(\mathbb{R}^n) = \mathbb{R}^n$. This is well-known (a consequence of invariance of domain and the connectedness of $\mathbb{R}^n$. One does not really need H to be quasisymmetric here, it is enough for H to be proper). On the other hand we are assuming that $h_k(E_k)$ fails to be ε-thick inside $B(0,1)$ for each k. It is not hard to derive a contradiction to this assumption. Indeed, let k be large, to be chosen soon, and suppose that $z_k \in B(0,1)$ satisfies

$$\operatorname{dist}(z_k, h_k(E_k \cap B(0,k))) \geq \varepsilon\,. \tag{3.15}$$

Because H is a surjection there is a point $x_k \in \mathbb{R}^n$ such that $H(x_k) = z_k$. In fact we have that $x_k \in B(0,L)$ for some large L and all k, because H is η-quasisymmetric, and because $z_k \in B(0,1)$ for all k. In particular we have that $x_k \in B(0,k)$ for large enough k. For sufficiently large k we can find a point $y_k \in E_k \cap B(0, L+1)$ such that $|x_k - y_k| \leq 1/k$, because of the thickness property. If k is large and among the k_j's then

$$\begin{aligned} |z_k - h_k(y_k)| &= |H(x_k) - h_k(y_k)| \\ &\leq |H(x_k) - H(y_k)| + |H(y_k) - h_k(y_k)| < \varepsilon\,, \end{aligned} \tag{3.16}$$

because of the uniform continuity of H on $B(0, L+1)$ and the uniform convergence (3.13). This contradicts (3.15), and Lemma 3.8 follows.

For our purposes the following reformulation of Lemma 3.8 will be more convenient.

Lemma 3.17. *Let a homeomorphism $\eta : [0,\infty) \to [0,\infty)$, a dimension n, and a number $A > 1$ be given. Suppose that X is a subset of $\mathbb{R}^n$ and that $f : X \to \mathbb{R}^n$ is η-quasisymmetric. Suppose also that we have $x, y \in X$, $x \neq y$, and $z \in \mathbb{R}^n \backslash X$ such that*

$$|z - x| \leq |x - y| \quad \textit{and} \quad \operatorname{dist}(z, X) \geq A^{-1}\,|x - y|\,. \tag{3.18}$$

Then there is a point $w \in \mathbb{R}^n \backslash f(X)$ such that

$$|w - f(x)| \leq M\,|f(x) - f(y)| \tag{3.19.a}$$

and

$$\text{(3.19.b)} \qquad \operatorname{dist}(w, f(X)) \geq M^{-1} |f(x) - f(y)|,$$

where $M > 0$ depends on $\eta, n,$ and A, but not on anything else.

Roughly speaking, this says that holes in the complement of X correspond under f to holes in the complement of $f(X)$ in a nice way.

This is an easy consequence of Lemma 3.8. We may as well assume that $x = 0$, $y = u = (1, 0, \dots, 0)$, $f(0) = 0$, and $f(u) = u$, because we can reduce to that case using affine similarities. We apply Lemma 3.8 with $h = f^{-1}$ and $E = f(X)$. (Lemma 1.4 is relevant here.) More precisely, we argue by contradiction. Suppose that there is no point w as in (3.19), so that $E = f(X)$ is $1/M$-thick in $B(0, M)$. If M is large enough, then we can apply Lemma 3.8 to conclude that $h(E) = X$ is ε-thick in $B(0, 1)$ with $\varepsilon = 1/(2A)$, for instance. This contradicts our assumption (3.18), and Lemma 3.17 follows.

Let us return now to our earlier story of F, g, and I. We want to take points in I and associate to them points in the complement of $g(F)$.

Let us decompose I into $I_0 \cup I_1$, where $I_0 = \{u \in I : \operatorname{dist}(u, F) \leq b \operatorname{diam} F\}$, $I_1 = \{u \in I : \operatorname{dist}(u, F) > b \operatorname{diam} F\}$, and $b \in (0, 1)$ is a small constant that will be chosen in a moment. I_0 is the more interesting one, I_1 can be handled practically without thinking. We shall concern ourselves with only I_0 for the time being. Note that $I = I_0$ and $I_1 = \varnothing$ when F is unbounded.

Given $u \in I_0$, choose points $\pi(u), \rho(u) \in F$ such that

$$\text{(3.20)} \qquad \begin{aligned} &|u - \pi(u)| = \operatorname{dist}(u, F), \\ &\operatorname{dist}(u, F) \leq |\pi(u) - \rho(u)| \leq C \operatorname{dist}(u, F). \end{aligned}$$

To get $\rho(u)$ we are using our assumption that F is serious. It is here that we choose the constant b, once and for all, depending only on the seriousness constant of F; we can find such a $\rho(u)$ so long as $\operatorname{dist}(u, F) \leq b \operatorname{diam} F$ and b is small enough. These points $\pi(u), \rho(u)$ are not unique or canonical or anything like that, we simply choose them without worrying about it.

Lemma 3.21. *For each $u \in I_0$ there is a point $\phi(u) \in \mathbb{R}^n \backslash g(F)$ such that*

$$\text{(3.22)} \qquad |\phi(u) - g(\pi(u))| \leq C\, |g(\pi(u)) - g(\rho(u))|$$

and

$$\operatorname{dist}(\phi(u), g(F)) \geq C^{-1} |g(\pi(u)) - g(\rho(u))| \tag{3.23}$$

for a suitable constant $C > 0$.

This follows from Lemma 3.17, applied with $f = g$, $X = F$, $x = \pi(u)$, $y = \rho(u)$, $z = u$, and with $\phi(u)$ taken to be w.

For $u \in I_1$ we can behave more stupidly.

Lemma 3.24. *For each* $u \in I_1$ *we can find a point* $\phi(u) \in \mathbb{R}^n \backslash g(F)$ *such that*

$$\operatorname{dist}(\phi(u), g(F)) = \frac{\operatorname{diam} g(F)}{\operatorname{diam} F} \operatorname{dist}(u, F). \tag{3.25}$$

Keep in mind that $\operatorname{dist}(u, F) \geq b \operatorname{diam} F$ this time. This means that there really is no point in choosing something like $\pi(u)$, an element of F closest to u, because they are all about the same. It is just a question of the distance to F.

The lemma is easy to prove, and we leave it as an exercise. For instance one can find a closed half-space which contains F and which touches F at the boundary, and then choose $\phi(u)$ on the ray which emanates from that point in the direction orthogonal to the hyperplane and away from F.

Thus we have now chosen points $\phi(u) \in \mathbb{R}^n \backslash g(F)$ for all $u \in I$. We need to modify them slightly to keep them from getting too close to each other.

Lemma 3.26. *For each* $u \in I$ *we can find a point* $\psi(u) \in \mathbb{R}^n \backslash g(F)$ *with the following properties.*

i) *If* $u \in I_0$, *then*

$$|\psi(u) - g(\pi(u))| \leq C |g(\pi(u)) - g(\rho(u))| \tag{3.27}$$

and

$$\operatorname{dist}(\psi(u), g(F)) \geq C^{-1} |g(\pi(u)) - g(\rho(u))|. \tag{3.28}$$

ii) *If $u \in I_1$, then*

$$C^{-1}\,\frac{\operatorname{diam} g(F)}{\operatorname{diam} F}\operatorname{dist}(u,F) \le \operatorname{dist}(\psi(u), g(F)) \le C\,\frac{\operatorname{diam} g(F)}{\operatorname{diam} F}\operatorname{dist}(u,F)\,. \tag{3.29}$$

iii) *There is a number $c \in (0, 1/10)$ such that the balls*

$$\beta(u) = \overline{B}(\psi(u), c \operatorname{dist}(\psi(u), g(F)))\,, \qquad u \in I\,,$$

have disjoint doubles.

The constants C and c depend only on n, the seriousness constant for F, and the function which governs the quasisymmetry of g.

To prove this we basically want to take the $\psi(u)$'s to be the same as the $\phi(u)$'s, but with some small perturbation to get the disjointness condition iii). This will require a small coding argument, and first we need to control some multiplicities.

Sublemma 3.30. *For each $u \in I$ there are at most a bounded number of $v \in I$ with*

$$|\phi(v) - \phi(u)| \le \frac{1}{2}\operatorname{dist}(\phi(u), g(F))\,. \tag{3.31}$$

To prove this we need the following.

Claim 3.32. *If $u, v \in I$ satisfy* (3.31), *then*

$$|u - v| \le C \operatorname{dist}(u,F) \tag{3.33}$$

and

$$C^{-1}\operatorname{dist}(u,F) \le \operatorname{dist}(v,F) \le C \operatorname{dist}(u,F) \tag{3.34}$$

for a suitable constant C.

Let $u, v \in I$ be given, with u and v satisfying (3.31). Notice that (3.31) implies that

$$\frac{1}{2}\operatorname{dist}(\phi(u), g(F)) \le \operatorname{dist}(\phi(v), g(F)) \le \frac{3}{2}\operatorname{dist}(\phi(u), g(F))\,. \tag{3.35}$$

Suppose first that $u, v \in I_1$. In this case we can get (3.34) from (3.35) and (3.25). This implies (3.33) immediately, because $\operatorname{dist}(u, F) \geq b \operatorname{diam} F$.

Now suppose that exactly one of u and v lies in I_1, let us say v. From (3.25) we get that

$$\operatorname{dist}(\phi(v), g(F)) \geq b \operatorname{diam} g(F). \tag{3.36}$$

On the other hand we have that

$$\operatorname{dist}(\phi(u), g(F)) \leq C \operatorname{diam} g(F) \tag{3.37}$$

because of (3.22). Using (3.35) we conclude that

$$C^{-1} \operatorname{diam} g(F) \leq \operatorname{dist}(\phi(u), g(F)) \tag{3.38.a}$$

and

$$\operatorname{dist}(\phi(v), g(F)) \leq C \operatorname{diam} g(F). \tag{3.38.b}$$

Going back to (3.25) we get that

$$C^{-1} \operatorname{diam} F \leq \operatorname{dist}(v, F) \leq C \operatorname{diam} F\,. \tag{3.39}$$

Let us check that

$$C^{-1} \operatorname{diam} F \leq \operatorname{dist}(u, F) \leq C \operatorname{diam} F\,. \tag{3.40}$$

The upper bound is automatic, because $u \in I_0$, the lower bound is the interesting one. It follows from (3.38.a)), Lemma 3.21, and the quasisymmetry of g on F.

These last two estimates imply (3.34), and (3.33) follows since $\operatorname{dist}(u, F)$ is bounded from below by a constant times $\operatorname{diam} F$. (This would also be true if we switched the roles of u and v, and immediately so, since we were assuming that $v \in I_1$.)

We are left with the case where both u and v lie in I_0. Set $R(x) = |g(\pi(x)) - g(\rho(x))|$ for $x = u, v$. Then

$$C^{-1}\, R(x) \leq \operatorname{dist}(\phi(x), g(F)) \leq C\, R(x) \tag{3.41}$$

when $x = u, v$, because of Lemma 3.21. Thus

$$C^{-1}\, R(u) \leq R(v) \leq C\, R(u) \tag{3.42}$$

by (3.35). On the other hand $|\phi(x) - g(\pi(x))| \le C\,R(x)$ when $x = u, v$, because of (3.22), and this implies that $|\phi(x) - g(\rho(x))| \le C\,R(x)$ for $x = u, v$ too. Using our assumption (3.31) we get that all the points $g(\pi(u)), g(\pi(v)), g(\rho(u))$, and $g(\rho(v))$ have mutual distance bounded by $C\,R(u)$, and also by $C\,R(v)$. Quasisymmetry then applies to say that the points $\pi(u), \pi(v), \rho(u), \rho(v)$ all have mutual distances bounded by $C\,|\pi(u) - \rho(u)|$, and by $C\,|\pi(v) - \rho(v)|$. In particular

$$C^{-1}\,|\pi(v) - \rho(v)| \le |\pi(u) - \rho(u)| \le C\,|\pi(v) - \rho(v)|\,. \tag{3.43}$$

This implies (3.34), because of (3.20). We also get (3.33) from these bounds on the mutual distances and (3.20). This proves Claim 3.32.

Now let us derive Sublemma 3.30 from the claim. Fix $u \in I$, and let $I(u)$ denote the set of $v \in I$ for which (3.31) holds. Thus (3.33) and (3.34) hold for all $v \in I(u)$. Consider the collection of balls $B(v)$, $v \in I(u)$, where $B(v)$ is as in Lemma 2.6. These balls all have approximately the same radius as $B(u)$, because of (3.34), and they are all contained in the ball $k\,B(u)$, where k is a large constant, because of (3.33) and (3.34). They are also disjoint, because of Lemma 2.6. This implies a bound on their total number, and Sublemma 3.30 follows.

Let us return now to the proof of Lemma 3.26.

Sublemma 3.44. *If the constant $c > 0$ is chosen small enough, then for each $u \in I$ we can find a point $\psi(u) \in \mathbb{R}^n \backslash g(F)$ such that*

$$|\phi(u) - \psi(u)| \le 10^{-2} \operatorname{dist}(\phi(u), g(F)) \tag{3.45}$$

and property iii) *of Lemma* (3.26) *holds.*

To do this we arrange the points in I as a sequence $\{u_j\}_{j=1}^{\infty}$, in which each element of I appears exactly once, and we choose $\psi(u_j)$ for one j after another. More precisely we want to choose these points so that for each j we have that (3.45) holds for $u = u_i$, $i = 1, \dots, j$, and, if the balls $\beta(u)$ are as defined in Lemma 3.26.iii), then for each j we have that

$$2\,\beta(u_i) \cap 2\,\beta(u_k) = \varnothing \qquad \text{when } \ 1 \le i < k \le j\,. \tag{3.46}$$

If we can do this for each j then we shall be finished.

Set $\psi(u_1) = \phi(u_1)$. This satisfies all the requirements for $j = 1$ trivially.

Suppose that $\psi(u_i)$ has been chosen for $i < j$ in accordance with the requirements stated above, and let us try to choose $\psi(u_j)$. Of course the disjointness property (3.46) is the thing that we have to keep our eyes on, and it is only an issue for $i < j$, $k = j$.

Consider first an $i < j$ such that

$$|\phi(u_i) - \phi(u_j)| \geq \frac{1}{2}\,\mathrm{dist}\,(\phi(u_j), g(F))\,. \tag{3.47}$$

We are assuming that we chose $\psi(u_i)$ so that (3.45) holds. If we are also careful to choose $\psi(u_j)$ so that (3.45) holds, then (3.47) will ensure that the disjointness property (3.46) will hold (with $k = j$) as soon as c is small enough. This is not hard to check, using also a computation like (3.35).

The interesting issue is to deal with the i's such that (3.47) fails, so that

$$|\phi(u_i) - \phi(u_j)| < \frac{1}{2}\,\mathrm{dist}\,(\phi(u_j), g(F))\,. \tag{3.48}$$

The point is that Sublemma 3.30 ensures that there are at most a bounded number of such i's. If c is chosen small enough then we can choose $\psi(u_j)$ so that (3.45) holds and so that (3.46) holds for these dangerous i's. This is not hard to see, the point is that we have only to avoid a bounded number of points in a given ball, and we can then get a c which is bounded from below in a way that depends on our bound on the number of bad points. This is slightly vague, but the reader is probably happier filling in the details rather than reading them.

Thus one can choose $\psi(u_j)$ so as to have the required properties. We can repeat this indefinitely to do this for all the u_j's, and Sublemma 3.44 follows from this, as noted above.

Let us now finish the proof of Lemma 3.26. We take $\psi(u)$ to be as provided in Sublemma 3.44, so that we have property iii) of Lemma 3.26 already. There remains the problem of verifying properties i) and ii) of Lemma 3.26. We want to derive them from (3.45) and the corresponding properties of $\phi(u)$. Notice first that (3.45) implies that

$$\frac{1}{2}\,\mathrm{dist}\,(\phi(u), g(F)) \leq \mathrm{dist}\,(\psi(u), g(F)) \leq 2\,\mathrm{dist}\,(\phi(u), g(F))\,, \tag{3.49}$$

as one can easily check. From here we get (3.29) when $u \in I_1$, using also the equality (3.25) for $\phi(u)$. Similarly (3.28) holds when $u \in I_0$, because of (3.49) and (3.23), while (3.27) follows from (3.22), (3.45), and the fact that

$$\begin{aligned} C^{-1}\,|g(\pi(u)) - g(\rho(u))| &\leq \operatorname{dist}(\phi(u), g(F)) \\ &\leq C\,|g(\pi(u)) - g(\rho(u))| \end{aligned} \tag{3.50}$$

(which itself comes from combining (3.22) and (3.23)).

This completes the proof of Lemma 3.26.

Define $h : F \cup I \to \mathbb{R}^n$ by

$$h = \begin{cases} g & \text{on } F, \\ \psi & \text{on } I. \end{cases} \tag{3.51}$$

We want to show that this mapping is quasisymmetric. This is not difficult but neither is it pleasant. We begin with small observations.

Lemma 3.52. *If $p, q \in F \cup I$, $p \neq q$, then*

$$|p - q| \geq \frac{1}{6}\big(\operatorname{dist}(p, F) + \operatorname{dist}(q, F)\big). \tag{3.53}$$

This follows easily from (2.3).

At the moment $\pi(u)$ and $\rho(u)$ are defined only for $u \in I_0$. We extend them to $u \in F$ simply by taking $\pi(u) = \rho(u) = u$ when $u \in F$.

Lemma 3.54. *Let $x, z \in F \cup I_0$ be given, and suppose that u is either x, $\pi(x)$, or $\rho(x)$, and that w is either z, $\pi(z)$, or $\rho(z)$. Then*

$$|u - w| \leq C\,|x - z|, \tag{3.55}$$

where C depends only on the seriousness constant of F.

This is an easy consequence of Lemma 3.52 and (3.20).

Lemma 3.56. *If x and z are distinct elements of $F \cup I$, then*

$$|h(x) - h(z)| \geq C^{-1}\left(\operatorname{dist}(h(x), g(F)) + \operatorname{dist}(h(z), g(F))\right), \tag{3.57}$$

where C depends only on the dimension n, the seriousness constant of F, and the function that governs the quasisymmetry of g.

Indeed, if either x or z lies in F then this is a tautology. If both x and z lie in I, then this follows from Lemma 3.26.iii).

Lemma 3.58. *Let $x, z \in F \cup I_0$ be given, and suppose that u is either x, $\pi(x)$, or $\rho(x)$, and that w is either z, $\pi(z)$, or $\rho(z)$. Then*

$$|h(u) - h(w)| \leq C\,|h(x) - h(z)|\,. \tag{3.59}$$

This constant depends only on the dimension n, the seriousness constant of F, and the function that governs the quasisymmetry of g.

Let us check that

$$\operatorname{dist}(h(p), g(F)) \leq |h(p) - h(q)| \leq C \operatorname{dist}(h(p), g(F)) \tag{3.60}$$

when $p \in F \cup I_0$ and q is either $\pi(p)$ or $\rho(p)$. This is trivial when $p \in F$, all the relevant quantities vanish, and so we need only consider $p \in I_0$. The first inequality follows from the fact that $q \in F$ by definitions. The second inequality follows from (3.27) and (3.28). (Think first about $q = \pi(p)$ and then $q = \rho(p)$. Remember that $h(p) = \psi(p)$, by (3.51).) Thus (3.60) is true.

The bound (3.59) follows now from Lemma 3.56 and (3.60). This proves Lemma 3.58.

Lemma 3.61. *Let $x, z \in F \cup I_0$ be given. We can choose $x' \in \{\pi(x), \rho(x)\}$ and $z' \in \{\pi(z), \rho(z)\}$ so that*

$$|x' - z'| \geq \frac{|x - z|}{16}\,. \tag{3.62}$$

Note that the reverse inequality is provided by Lemma 3.54.

For the proof we follow a suggestion from the Unknown Finn. Let us check first that if $u \in F \cup I_0$ and $v \in \mathbb{R}^n$ is arbitrary, then

$$|u - v| \leq 2\,\left(|\pi(u) - v| + |\rho(u) - v|\right)\,. \tag{3.63}$$

We have that

$$|\pi(u) - v| + |\rho(u) - v| \geq |\pi(u) - \rho(u)| \geq \operatorname{dist}(u, F)\,, \tag{3.64}$$

by (3.20). If $|u-v| \le 2\,\mathrm{dist}\,(u,F)$ then we get (3.63) from (3.64). If $|u-v| > 2\,\mathrm{dist}\,(u,F)$ then

$$\begin{aligned} |u-v| &\le |u-\pi(u)| + |\pi(u)-v| \\ &= \mathrm{dist}\,(u,F) + |\pi(u)-v| \\ &< |u-v|/2 + |\pi(u)-v|\,, \end{aligned}$$

and so $|u-v| < 2\,|\pi(u)-v|$. Thus (3.63) holds in this case too.

From (3.63) (applied twice) we conclude that if $u,v \in F \cup I_0$, then

$$\begin{aligned} |u-v| \le 4\,(&|\pi(u)-\pi(v)| + |\rho(u)-\rho(v)| \\ &+ |\pi(u)-\rho(v)| + |\rho(u)-\pi(v)|)\,. \end{aligned} \tag{3.65}$$

Lemma 3.61 follows from this.

Lemma 3.66. *Let $x,y,z \in F \cup I_0$ and $t > 0$ be given, with $x \ne y$ and $|x-y| \le t\,|x-z|$. Then*

$$\mathrm{dist}\,(h(x), g(F)) \le C\,\eta(C\,t)\,|h(x)-h(z)|\,, \tag{3.67}$$

where η is the function that governs the quasisymmetry of g, and where C depends only on n, the seriousness constant of F, and η.

This lemma is trivial when $x \in F$, and so we assume that $x \in I_0$. Lemma 3.52 permits us to convert our hypothesis into

$$\mathrm{dist}\,(x,F) \le 6\,t\,|x-z|\,. \tag{3.68}$$

Let x', z' be as in Lemma 3.61. We can convert (3.68) into

$$|x'-q| \le C\,t\,|x'-z'| \tag{3.69}$$

for $q = \pi(x), \rho(x)$. This follows from (3.68), using (3.62) and (3.20). All these points x', z', q lie in F, on which h equals g, and so we can use the quasisymmetry of g to get

$$|h(x')-h(q)| \le \eta(C\,t)\,|h(x')-h(z')| \tag{3.70}$$

for $q = \pi(x), \rho(x)$, where η is the function that governs the quasisymmetry of g. Because x' is one of $\pi(x), \rho(x)$, we can take q to be the other one, and we get

$$|h(\pi(x)) - h(\rho(x))| \le \eta(C\,t)\,|h(x')-h(z')|\,. \tag{3.71}$$

Lemma 3.26.i) permits us to replace this with

$$\operatorname{dist}(h(x), g(F)) \le C\,\eta(C\,t)\,|h(x') - h(z')|\,. \tag{3.72}$$

Using Lemma 3.58 we get that

$$\operatorname{dist}(h(x), g(F)) \le C\,\eta(C\,t)\,|h(x) - h(z)|\,. \tag{3.73}$$

This proves Lemma 3.66.

Lemma 3.74. *The restriction of h to $F \cup I_0$ is quasisymmetric, with the quasisymmetry governed by the function $C\,\eta(C\,t)$, where η is the function that governs the quasisymmetry of g, and where C depends only on the dimension n, the seriousness constant of F, and η.*

Let $x, y, z \in F \cup I_0$, $t > 0$, be given, such that

$$|x - y| \le t\,|x - z|\,. \tag{3.75}$$

We want to show that

$$|h(x) - h(y)| \le C\,\eta(C\,t)\,|h(x) - h(z)|\,, \tag{3.76}$$

where C and η are as above. We may as well assume that $y \neq x$.

Let x', z' be associated to x, z as in Lemma 3.61. Then (3.75) implies that $|x - y| \le C\,t\,|x' - z'|$, by (3.62). Therefore

$$|x' - q| \le C\,t\,|x' - z'| \tag{3.77}$$

for each of $q = \pi(y), \rho(y)$, because of Lemma 3.54 (applied to x and y). Because x', z', $\pi(y)$, $\rho(y)$ all lie in F, and because h equals g on F, we conclude that

$$|h(x') - h(q)| \le \eta(C\,t)\,|h(x') - h(z')| \tag{3.78}$$

for each of $q = \pi(y), \rho(y)$, where η is the function that governs the quasisymmetry of g. Lemma 3.58 permits us to convert this into

$$|h(x') - h(q)| \le C\,\eta(C\,t)\,|h(x) - h(z)|\,, \tag{3.79}$$

for each of $q = \pi(y), \rho(y)$.

Lemma 3.66 implies that

$$\operatorname{dist}(h(x), g(F)) \leq C\,\eta(C\,t)\,|h(x) - h(z)|\,. \tag{3.80}$$

From (3.60) (with $p = x$, $q = x'$) we have that $|h(x) - h(x')| \leq C \operatorname{dist}(h(x), g(F))$. Combining these estimates with (3.79) we get that

$$|h(x) - h(q)| \leq C\,\eta(C\,t)\,|h(x) - h(z)|\,, \tag{3.81}$$

for each of $q = \pi(y), \rho(y)$.

In particular we have that

$$|h(\pi(y)) - h(\rho(y))| \leq C\,\eta(C\,t)\,|h(x) - h(z)|\,, \tag{3.82}$$

and hence

$$|h(y) - h(\pi(y))| \leq C\,\eta(C\,t)\,|h(x) - h(z)| \tag{3.83}$$

by (3.27). Combining this with (3.81) (with $q = \pi(y)$) we get that

$$|h(x) - h(y)| \leq C\,\eta(C\,t)\,|h(x) - h(z)|\,. \tag{3.84}$$

This proves the lemma.

Our next main goal is to prove the following.

Lemma 3.85. *$h : F \cup I \to \mathbb{R}^n$ is quasisymmetric, with bounds that depend only on the dimension n, the seriousness constant of F, and the function that governs the quasisymmetry of g.*

In order to prove this we may as well assume that

$$\operatorname{diam} F = \operatorname{diam} g(F) = 1\,, \tag{3.86}$$

because we can always make rescalings on the domain and image without altering our assumptions. This assumption will be in force throughout the proof of Lemma 3.85.

In the following the constants C are permitted to depend only on the dimension n, the seriousness constant of F, and the function that governs the quasisymmetry of g.

The reader might wish to review the definitions of I_0 and I_1, which are given shortly before (3.20). In particular they imply that

$$\operatorname{diam}(F \cup I_0) \leq 3\,. \tag{3.87}$$

Using this and Lemma 3.26.i) we get that

$$\operatorname{diam} h(F \cup I_0) \le C\,. \tag{3.88}$$

Sublemma 3.89. *If $p \in F \cup I$ and $q \in I_1$, then*

$$C^{-1}\,|p-q| \le |h(p)-h(q)| \le C\,|p-q|\,, \tag{3.90}$$

We may as well assume that $p \neq q$.

Let us prove the upper bound first. We can do it crudely, starting with

$$\begin{aligned} |h(p)-h(q)| &\le \operatorname{dist}(h(p), g(F)) + \operatorname{diam} g(F) \\ &\qquad + \operatorname{dist}(h(q), g(F)) \\ &= \operatorname{dist}(h(p), g(F)) + 1 + \operatorname{dist}(h(q), g(F))\,. \end{aligned} \tag{3.91}$$

On the other hand we have that

$$\begin{aligned} |p-q| &\ge \frac{1}{6}\left(\operatorname{dist}(p,F) + \operatorname{dist}(q,F)\right) \\ &\ge C^{-1}(\operatorname{dist}(p,F) + \operatorname{dist}(q,F) + 1)\,. \end{aligned} \tag{3.92}$$

The first inequality comes from Lemma 3.52, while the second follows from our assumption that $q \in I_1$. From (3.29) we get that

$$\operatorname{dist}(h(q), g(F)) \le C \operatorname{dist}(q,F)\,. \tag{3.93}$$

If $p \in I_1$ we have the analogous inequality for p instead of q, and then the upper bound in (3.90) follows from (3.91) and (3.92). If $p \in F \cup I_0$, then $\operatorname{dist}(h(p), g(F)) \le C$ by (3.88), and the upper bound in (3.90) again follows from (3.91) and (3.92). This proves the upper bound in (3.90).

Let us now prove the lower bound. Lemma 3.56 implies that

$$|h(p)-h(q)| \ge C^{-1}\left(\operatorname{dist}(h(p), g(F)) + \operatorname{dist}(h(q), g(F))\right), \tag{3.94}$$

Using (3.29) we get that

$$\operatorname{dist}(h(q), g(F)) \ge C^{-1}\operatorname{dist}(q,F) \ge C^{-1}\left(\operatorname{dist}(q,F)+1\right). \tag{3.95}$$

We are also employing the assumption that $q \in I_1$ to get the last inequality. If $p \in I_1$, then we get the analogue of (3.95) for p as well, and then

$$|h(p) - h(q)| \geq C^{-1} \left(\mathrm{dist}\,(q, F) + 1 + \mathrm{dist}\,(p, F)\right). \tag{3.96}$$

This implies the lower bound in (3.90), using also (3.86). If $p \in F \cup I_0$, then we have

$$\begin{aligned} |h(p) - h(q)| &\geq C^{-1} \left(\mathrm{dist}\,(q, F) + 1\right) \\ &\geq C^{-1} \left(\mathrm{dist}\,(q, F) + \mathrm{diam}\,(F \cup I_0)\right), \end{aligned} \tag{3.97}$$

by (3.87). This implies the lower bound in (3.90) in this case.

This proves Sublemma 3.89.

Let us come back now to the proof of Lemma 3.85. Let $x, y, z \in F \cup I$ and $t > 0$ be given, with

$$|x - y| \leq t\,|x - z|\,. \tag{3.98}$$

We want to show that

$$|h(x) - h(y)| \leq \theta(t)\,|h(x) - h(z)|\,, \tag{3.99}$$

where $\theta : [0, \infty) \to [0, \infty)$ vanishes at the origin, is continuous at the origin, and is bounded on bounded sets. (Lemma 2.42 is relevant here.)

If all three of x, y, z lie in $F \cup I_0$ then we can use Lemma 3.74 to get the required estimate.

If $x \in I_1$ then we have that

$$|h(x) - h(y)| \leq C\,t\,|h(x) - h(z)|\,, \tag{3.100}$$

because of Sublemma 3.89. Thus we may assume that

$$x \in F \cup I_0\,. \tag{3.101}$$

If both y and z lie in I_1, then we get (3.100) again from Sublemma 3.89. If they both lie in $F \cup I_0$ then all three points lie there and we are back to a case that we know. Thus we may require that

$$\text{exactly one of } y \text{ and } z \text{ lies in } F \cup I_0\,. \tag{3.102}$$

Let us pause for a small observation.

Sublemma 3.103. *If* $p \in F \cup I_0$ *and* $q \in I_1$, *then* $|p - q| \geq C^{-1}$.

Indeed, in this case $6\,|p - q| \geq \operatorname{dist}(q, F)$, by Lemma 3.52, and Sublemma 3.103 follows from the assumption that $q \in I_1$.

Let us come back now to the task of proving an estimate like (3.99) under the conditions (3.98), (3.101), and (3.102). Assume first that $y \in I_1$. In this case we have

$$|x - y| \geq C^{-1}, \tag{3.104}$$

by Sublemma 3.103. This implies that $\operatorname{diam} F \leq C\,t\,|x - z|$, and so the quasisymmetry of h on $F \cup I_0$ (Lemma 3.74) implies that

$$1 = \operatorname{diam} g(F) \leq C\,\eta(C\,t)\,|h(x) - h(z)|\,, \tag{3.105}$$

where η is the function that controls the quasisymmetry of g. On the other hand

$$|h(x) - h(y)| \leq C\,|x - y|\,, \tag{3.106}$$

by Sublemma 3.89, and so

$$\begin{aligned} |h(x) - h(y)| &\leq C\,t\,|x - z| \\ &\leq C\,t\operatorname{diam}(F \cup I_0) \\ &\leq C\,t \leq C\,t\,\eta(C\,t)\,|h(x) - h(z)| \end{aligned} \tag{3.107}$$

by (3.87) and (3.105). This is the kind of estimate that we want.

Assume now that $z \in I_1$, so that $x, y \in F \cup I_0$. Notice that

$$|x - z| \geq C^{-1} \tag{3.108}$$

and

$$C^{-1}\,|x - z| \leq |h(x) - h(z)| \leq C\,|x - z| \tag{3.109}$$

by Sublemmas 3.103 and 3.89. Our assumption (3.98) implies that either

$$|x - y| \leq \sqrt{t} \tag{3.110}$$

or

$$1 \leq \sqrt{t}\, |x - z| \,. \tag{3.111}$$

Assume first that (3.111) holds. Then we have that

$$\begin{aligned} |h(x) - h(y)| &\leq \operatorname{diam} h(F \cup I_0) \\ &\leq C \\ &\leq C \sqrt{t}\, |x - z| \\ &\leq C \sqrt{t}\, |h(x) - h(z)| \,, \end{aligned} \tag{3.112}$$

by (3.109). This estimate does the job for this case. So suppose now that (3.110) holds. In this case we have that

$$|h(x) - h(y)| \leq \omega(\sqrt{t}) \tag{3.113}$$

for a certain function ω on $[0, \infty)$ which vanishes at the origin, is continuous at the origin, and is bounded. Indeed, we have $x, y \in F \cup I_0$ in the present situation, and so (3.113) follows from the quasisymmetry of h on $F \cup I_0$ and (3.87), (3.88). Using (3.108) and (3.109) we get that

$$|h(x) - h(y)| \leq C\, \omega(\sqrt{t})\, |h(x) - h(z)| \,, \tag{3.114}$$

which does the job in this case.

This completes the proof of Lemma 3.85.

Note that we have not tried to give sharp estimates here, it was more interesting to just get it over with.

Let us now finish the proof of Proposition 1.14. Let S be as in (3.1), and let us define a mapping G on S. We set $G = g$ on F, and if $x \in I$ we set

$$G(p) = h(x) + \frac{c \operatorname{dist}(h(x), g(F))}{20^{-1} \operatorname{dist}(x, F)}\, (p - x) \qquad \text{for } p \in B(x) \,. \tag{3.115}$$

Here c is chosen as in Lemma 3.26.iii); the ratio in (3.115) is simply the ratio between the radius of the ball $\beta(x)$ defined in Lemma 3.26.iii) and the radius of the ball $B(x)$ which is used in (3.1). In fact G maps the center of $B(x)$ to the center $\psi(x) = h(x)$ of $\beta(x)$, by definitions, and so we get that $G(B(x)) = \beta(x)$ for all $x \in I$.

We want to say that $G : S \to \mathbb{R}^n$ is quasisymmetric with a suitable bound. We apply Lemma 2.29, with $A = F$, $H = G$, and with the balls B_i and β_i taken to be the $B(x)$'s and $\beta(x)$'s, with the obvious changes in notation. We have to check that the hypotheses of Lemma 2.29 hold in this case. The requirement that "the restriction of H to A' is quasisymmetric" is satisfied in this case because of Lemma 3.85. We just checked that H maps the B_i's onto the β_i's, and the restriction of H to each B_i is a similarity, and hence quasisymmetric with uniform bounds. We know from Lemma 2.6 that the doubles of the B_i's are disjoint, and they are disjoint from $A = F$ by their definition. Similarly the β_i's have disjoint doubles because of Lemma 3.26.iii), and the doubles are disjoint from $H(A) = g(F)$ by their definition. The bounds (2.30) also follow from the definitions of the B_i's and β_i's. Thus the hypotheses of Lemma 2.29 are satisfied in this case, and we conclude that $H : A^* \to \mathbb{R}^n$ is quasisymmetric, which is the same as saying that $G : S \to \mathbb{R}^n$ is quasisymmetric. Of course we also get the correct bounds.

This completes the proof of Proposition 1.14.

4. The proof of Proposition 1.15.

Let us address first a preliminary point.

Proposition 4.1. *If F is a serious subset of $\mathbb{R}^n$ and $g : F \to \mathbb{R}^n$ is quasisymmetric, then $g(F)$ is also serious, with a constant which depends only on the seriousness constant of F and the function that governs the quasisymmetry of g.*

This is less amusing than Proposition 1.15, because it is really a fact about (quasi-) metric spaces rather than subsets of Euclidean spaces.

Let $x \in F$ be given. For each $0 < t < \operatorname{diam} F$ choose a point $y(t) \in F$ so that

$$C_0^{-1}\, t \le |x - y(t)| \le t\,, \tag{4.2}$$

where C_0 is the seriousness constant of F.

Claim 4.3. *There is a constant $C > 0$ so that for each $0 < s < \operatorname{diam} g(F)$ we can find a $0 < t < \operatorname{diam} F$ such that*

$$C^{-1}\, s \le |g(x) - g(y(t))| \le C\, s\,. \tag{4.4}$$

To prove the claim we use a continuity argument. We have that

$$C^{-1}\operatorname{diam} g(F) \le |g(x) - g(y(t))| \le C \operatorname{diam} g(F) \tag{4.5}$$

when $t > \operatorname{diam}(F)/2$ and $\operatorname{diam} F < \infty$, because of (4.2) and quasisymmetry, and

$$\lim_{t\to\infty} |g(x) - g(y(t))| = \infty \tag{4.6}$$

when $\operatorname{diam} F = \infty$. The continuity of g implies that

$$\lim_{t\to 0} |g(x) - g(y(t))| = 0\,. \tag{4.7}$$

We also have that

$$C^{-1}\,|g(x) - g(y(t))| \le |g(x) - g(y(t/2))| \le C\,|g(x) - g(y(t))| \tag{4.8}$$

when $0 < t < \operatorname{diam} F$, by quasisymmetry. The claim follows from these three observations.

Proposition 4.1 follows easily from Claim 4.2.

Now let us prove Proposition 1.15.

Let S be a strong subset of $\mathbb{R}^n$ and let $G : S \to \mathbb{R}^n$ be quasisymmetric. We want to show that $G(S)$ is strong, with bounds. We know from Proposition 4.1 that $G(S)$ is serious. Of course S is unbounded, since it is strong, and so $G(S)$ is also unbounded. $G(S)$ is also closed, since S is.

Let $x \in \mathbb{R}^n \backslash G(S)$ be given. Choose $x_0 \in G(S)$ so that $\operatorname{dist}(x, G(S)) = |x - x_0|$, and choose $x_1 \in G(S)$ so that

$$C^{-1}\,|x_1 - x_0| \le |x - x_0| \le |x_1 - x_0|\,. \tag{4.9}$$

We can do this because $G(S)$ is serious and unbounded.

We can apply Lemma 3.17 (with $X = G(S)$, $f = G^{-1}$ (remember Lemma 1.4), $x = x_0$, $y = x_1$, and $z = x$) to get a point $w \in \mathbb{R}^n \backslash S$ such that

$$|w - G^{-1}(x_0)| \le C\,|G^{-1}(x_0) - G^{-1}(x_1)| \tag{4.10}$$

and

$$\operatorname{dist}(w, S) \ge C^{-1}|G^{-1}(x_0) - G^{-1}(x_1)|\,. \tag{4.11}$$

Our assumption that S is strong implies the existence of a point $v \in S$ such that

$$|w - v| \le C \operatorname{dist}(w, S) \tag{4.12}$$

and

$$\operatorname{dist}(v, \mathbb{R}^n \backslash S) \ge C^{-1} \operatorname{dist}(w, S)\,. \tag{4.13}$$

Let us rephrase (4.13) as

$$B(v, C^{-1} \operatorname{dist}(w, S)) \subseteq S\,. \tag{4.14}$$

Set $y = G(v)$. We want to show that

$$|x - y| \le C \operatorname{dist}(x, G(S)) \tag{4.15}$$

and

$$\operatorname{dist}(y, \mathbb{R}^n \backslash G(S)) \ge C^{-1} \operatorname{dist}(x, G(S))\,. \tag{4.16}$$

We shall derive these from (4.12) and (4.13) using the quasisymmetry of G.

From (4.10) and (4.11) we have that

$$\begin{aligned} C^{-1} \left|G^{-1}(x_0) - G^{-1}(x_1)\right| &\le \operatorname{dist}(w, S) \\ &\le C \left|G^{-1}(x_0) - G^{-1}(x_1)\right|. \end{aligned} \tag{4.17}$$

Combining (4.12) and (4.10) we get that

$$\left|v - G^{-1}(x_0)\right| \le C \left|G^{-1}(x_0) - G^{-1}(x_1)\right|. \tag{4.18}$$

Since G is quasisymmetric we conclude that

$$|y - x_0| \le C\,|x_0 - x_1|\,. \tag{4.19}$$

Using (4.9) we can convert this into

$$|y - x| \le C\,|x - x_0|\,. \tag{4.20}$$

This implies (4.15), because of our choice of x_0.

It remains to prove (4.16), which we can rewrite as

$$B(y, C^{-1}\mathrm{dist}\,(x, G(S))) \subseteq G(S)\,. \tag{4.21}$$

Of course the point is to use (4.13). Let B denote the ball on the left side of (4.14). Because of invariance of domain we have that $G(B)$ is an open subset of $\mathbb{R}^n$ which contains y.

Claim 4.22. $\mathrm{dist}\,(y, G(B\backslash(B/2))) \geq C^{-1}\mathrm{dist}\,(x, G(S))\,.$

To see this we want to show that

$$\mathrm{dist}\,(y, G(B\backslash(B/2))) \geq C^{-1}\,|x_0 - x_1|\,. \tag{4.23}$$

To prove this we use the quasisymmetry of G. Let $z \in B\backslash(B/2)$ be given. Then

$$C^{-1}\,|G^{-1}(x_0) - G^{-1}(x_1)| \leq |z - v| \leq C\,|G^{-1}(x_0) - G^{-1}(x_1)|\,, \tag{4.24}$$

because of (4.17). This implies that

$$|z - v| \geq C^{-1}\,|v - G^{-1}(x_i)|\,, \qquad i = 0, 1\,, \tag{4.25}$$

by (4.18). Using this and quasisymmetry it is not hard to show that

$$|G(z) - y| \geq C^{-1}\,|x_0 - x_1|\,. \tag{4.26}$$

(Remember that $y = G(v)$.) With (4.26) in hand we get (4.23) immediately, and Claim 4.22 follows from (4.9) and our choice of x_0.

Let us now use the claim to derive (4.21). Let $p \in \mathbb{R}^n\backslash G(B)$ be chosen so that $|p - y|$ is as small as possible. We can do this because $G(B)$ is an open subset of $\mathbb{R}^n$, and we also get that $|p-y| > 0$. Set $p_t = y + t\,(p-y)$ for $0 < t < 1$, so that each p_t lies in $G(B)$. For t sufficiently close to 1 we must have that $p_t \in G(B\backslash(B/2))$; for if this were not the case, then p would lie in $G(\overline{B}/2)$, in contradiction to our choice of p (lying outside $G(B)$). Thus $p_t \in G(B\backslash(B/2))$ for t sufficiently close to 1, and we conclude from Claim 4.22 that $|p - y| \geq C^{-1}\mathrm{dist}\,(x, G(S))$. This proves (4.16).

Thus we have proved that $G(S)$ is a strong set, and Proposition 1.15 follows.

5. The proof of Proposition 1.16.

Let S be a strong subset of $\mathbb{R}^n$, and let us try to prove (1.17). It suffices to show that there is a constant $k > 1$ so that

$$|S \cap B(p, k\,r)| \geq k^{-1}\,|B(p,r)\backslash S| \tag{5.1}$$

for all $p \in S$ and $r > 0$.

Let $p \in S$ and $r > 0$ be given, and let us apply Lemma 2.1 with $E = S$ and $H = B(p,r)\backslash S$. Lemma 2.1 produces a subset I of H with the properties listed there. From (2.2) we get that

$$|B(p,r)\backslash S| \leq C \sum_{x \in I} (\mathrm{dist}\,(x,S))^n. \tag{5.2}$$

Given $x \in I$ choose $\tau(x) \in S$ so that

$$|x - \tau(x)| \leq C\,\mathrm{dist}\,(x,S) \tag{5.3}$$

and

$$\mathrm{dist}\,(\tau(x), \mathbb{R}^n\backslash S) \geq C^{-1}\mathrm{dist}\,(x,S)\,. \tag{5.4}$$

We can do this because S is strong. Note that these inequalities imply that

$$C^{-1}\mathrm{dist}\,(x,S) \leq \mathrm{dist}\,(\tau(x), \mathbb{R}^n\backslash S) \leq C\,\mathrm{dist}\,(x,S)\,. \tag{5.5}$$

Given $x \in I$, set

$$\beta(x) = B(\tau(x), \mathrm{dist}\,(\tau(x), \mathbb{R}^n\backslash S)/2)\,. \tag{5.6}$$

From (5.2) and (5.5) we have that

$$|B(p,r)\backslash S| \leq C \sum_{x \in I} |\beta(x)|\,. \tag{5.7}$$

We want to use this to prove (5.1).

Lemma 5.8. *For each $x \in I$ there are at most a bounded number of $z \in I$ such that $\beta(x)$ intersects $\beta(z)$.*

Suppose that $x, z \in I$ satisfy $\beta(x) \cap \beta(z) \neq \varnothing$. Then

$$(5.9) \qquad |\tau(x) - \tau(z)| < \frac{1}{2}\left(\operatorname{dist}(\tau(x), \mathbb{R}^n \backslash S) + \operatorname{dist}(\tau(z), \mathbb{R}^n \backslash S)\right).$$

This implies that

$$(5.10) \quad \frac{1}{3}\operatorname{dist}(\tau(x), \mathbb{R}^n \backslash S) \leq \operatorname{dist}(\tau(z), \mathbb{R}^n \backslash S) \leq 3\operatorname{dist}(\tau(x), \mathbb{R}^n \backslash S).$$

Using (5.5) we conclude that

$$(5.11) \qquad C^{-1}\operatorname{dist}(x, S) \leq \operatorname{dist}(z, S) \leq C\operatorname{dist}(x, S).$$

We also get that

$$(5.12) \qquad |x - z| \leq C\operatorname{dist}(x, S),$$

because of (5.3), (5.9), (5.5), and (5.11).

Let $I(x)$ denote the set of $z \in I$ such that $\beta(x) \cap \beta(z) \neq \varnothing$. From (2.3) and (5.11) we obtain that

$$(5.13) \qquad |y - z| \geq C^{-1}\operatorname{dist}(x, S) \quad \text{when } y, z \in I(x),\ y \neq z.$$

It is easy to see that $I(x)$ can have only a bounded number of elements, using (5.12) and (5.13). This proves the lemma.

Lemma 5.8 permits us to convert (5.7) into

$$(5.14) \qquad |B(p, r) \backslash S| \leq C \left| \bigcup_{x \in I} \beta(x) \right|.$$

Let us check that

$$(5.15) \qquad \bigcup_{x \in I} \beta(x) \subseteq B(p, C\,r) \cap S.$$

We have $\beta(x) \subseteq S$ from the definition (5.6). We also know that its radius is bounded by $C\operatorname{dist}(x, S)$, and this is at most $C\,|x - p| \leq C\,r$ for $x \in I$. The inclusion (5.15) follows easily from these observations, and the fact that $I \subseteq B(p, r)$ by definitions.

Combining (5.15) with (5.14) we get (5.1). This completes the proof of Proposition 1.16.

6. The proof of Proposition 1.22.

The proof of Proposition 1.22 is a straightforward consequence of the previous results and the definitions, but let us be slightly careful. Let S be a strong subset of $\mathbb{R}^n$ and let $G : S \to \mathbb{R}^n$ be quasisymmetric, as in the proposition. Define the measure ν on $\mathbb{R}^n$ by $\nu(A) = |A \cap G(S)|$. Note that $G(S)$ is a strong subset of $\mathbb{R}^n$, because of Proposition 1.15. Thus

$$C^{-1}\, r^n \le \nu(B(x,r)) \le C\, r^n \tag{6.1}$$

for some constant C and all $x \in G(S)$, $r > 0$, by Proposition 1.16. Of course ν has support equal to $G(S)$.

Define the measure μ on $\mathbb{R}^n$ by $\mu(A) = |G(A \cap S)|$, as in the statement of Proposition 1.22. Thus μ is a measure with support equal to S which is obtained by pulling back ν using the homeomorphism G.

That μ is doubling on S, as in Definition 1.18 a), is easy to check, using (6.1) and the quasisymmetry of G. The point is that if we are given $x \in S$ and $r > 0$, then we can find a ball $B = B(G(x), t)$ such that $G(B(x,r) \cap S) \supseteq B \cap G(S)$ and $G(B(x, 2\,r) \cap S) \subseteq k\,B \cap G(S)$, where k is a constant that does not depend on x or r.

To see that μ is a metric doubling measure on S, as in Definition 1.18.b), it suffices to show that

$$C^{-1}\,|G(x) - G(y)| \le \delta(x,y) \le C\,|G(x) - G(y)| \tag{6.2}$$

for some C and all $x, y \in S$, where $\delta(x,y)$ is as in (1.20). This is sufficient because $d(x,y) = |G(x) - G(y)|$ is obviously a metric on S. To get these bounds the main point is that

$$\begin{aligned} &B(G(x), C^{-1}\,|G(x) - G(y)|) \cap G(S) \\ &\qquad \subseteq G(B(x, |x-y|) \cap S) \\ &\qquad \subseteq B(G(x), C\,|G(x) - G(y)|) \cap G(S)\,. \end{aligned} \tag{6.3}$$

These inclusions follow from the quasisymmetry of G. Once we have them (6.2) follows easily from the definition (1.20) of $\delta(x,y)$ and the estimate (6.1).

This completes the proof of Proposition 1.22.

7. The proof of Proposition 1.23.

Let S be a strong subset of $\mathbb{R}^n$, and let μ be a metric doubling measure on S. We want to find a metric doubling measure ν on all of $\mathbb{R}^n$ which equals μ on S.

Let $\{Q_i\}_{i\in I}$ be a Whitney decomposition of $\mathbb{R}^n\backslash S$. Thus the Q_i's are closed cubes with disjoint interiors whose union is all of $\mathbb{R}^n\backslash S$ and which satisfy

$$\operatorname{diam} Q_i \leq \operatorname{dist}(Q_i, S) \leq 4 \operatorname{diam} Q_i\,, \tag{7.1}$$

as in [St, Theorem 1, p. 167].

We shall use this Whitney decomposition to define ν, we shall define it in a simple way on each Q_i and then combine the pieces. In order to define ν on the Q_i's we need to look at μ inside S, and we need to use our assumption that S is a strong set.

For each $i \in I$ choose $q_i \in Q_i$ so that

$$\operatorname{dist}(q_i, S) = \operatorname{dist}(Q_i, S)\,. \tag{7.2}$$

Using the fact that S is a strong set we can find a cousin for each q_i inside S, namely a point p_i such that

$$|p_i - q_i| \leq C \operatorname{dist}(q_i, S) \tag{7.3}$$

and

$$\operatorname{dist}(p_i, \mathbb{R}^n\backslash S) \geq C^{-1} \operatorname{dist}(q_i, S)\,. \tag{7.4}$$

These inequalities imply easily that

$$C^{-1}\operatorname{dist}(q_i, S) \leq \operatorname{dist}(p_i, \mathbb{R}^n\backslash S) \leq C \operatorname{dist}(q_i, S)\,. \tag{7.5}$$

Given $i \in I$ set

$$\beta_i = B(p_i, \operatorname{dist}(p_i, \mathbb{R}^n\backslash S)/2)\,. \tag{7.6}$$

Define ν by

$$\nu(A) = \mu(A \cap S) + \sum_{i\in I} \frac{\mu(\beta_i)}{|Q_i|}\, |A \cap Q_i|\,. \tag{7.7}$$

We want to show that this is a metric doubling measure on $\mathbb{R}^n$.

(It is not hard to see that (7.7) is the right way to define ν. There are various ways to package this extension, but basically there is only one reasonable way to do it, and this is it.)

The proof that ν is a metric doubling measure is pretty straightforward, a matter of checking that certain things follow from certain other things. We begin with some small technical observations. The constants C that appear below are allowed to depend only on the dimension n, the metric doubling constants for μ, and the strongness constant for S.

Lemma 7.8. *If* Q_i *is a Whitney cube and* $\operatorname{dist}(x, Q_i) \le \operatorname{diam}(Q_i)/10$, *then*

$$\frac{9}{10}\operatorname{diam} Q_i \le \operatorname{dist}(x, S) \le 6 \operatorname{diam} Q_i \,. \tag{7.9}$$

This is an immediate consequence of (7.1).

Lemma 7.10. *If two Whitney cubes* Q_i *and* Q_j *satisfy* $\operatorname{dist}(Q_i, Q_j) \le \operatorname{diam}(Q_i)/10$, *then*

$$\frac{9}{60}\operatorname{diam} Q_i \le \operatorname{diam} Q_j \le 6 \operatorname{diam} Q_i \,. \tag{7.11}$$

This follows from (7.1) and Lemma 7.8.

Lemma 7.12. *If two Whitney cubes* Q_i *and* Q_j *satisfy* $\operatorname{dist}(Q_i, Q_j) \le \operatorname{diam}(Q_i)/10$, *then*

$$C^{-1}\,\mu(\beta_i) \le \mu(\beta_j) \le C\,\mu(\beta_i)\,. \tag{7.13}$$

If Q_i and Q_j are as above, then

$$|p_i - p_j| \le C \operatorname{diam} Q_i \,, \tag{7.14}$$

by (7.3), (7.2), and (7.11). Also the radii of both β_i and β_j are comparable to $\operatorname{diam} Q_i$, because of (7.5) and (7.2). Thus we conclude that β_i is contained in some bounded multiple of β_j, and vice-versa. The doubling condition then yields (7.13).

Lemmma 7.15. *For each $i \in I$ there is only a bounded number of $j \in I$ such that β_i intersects β_j.*

This is very similar to Lemma 5.8. If $\beta_i \cap \beta_j \neq \varnothing$, then one can show that the radii of β_i and β_j are the same to within a factor of 3, for the same reason as in (5.10). This implies that

$$C^{-1}\operatorname{diam} Q_i \leq \operatorname{diam} Q_j \leq C \operatorname{diam} Q_i \tag{7.16}$$

for some constant C. Next $\beta_i \cap \beta_j \neq \varnothing$ implies that $|p_i - p_j| \leq C \operatorname{diam} Q_i$, because of (7.5), (7.2), and (7.1). Using (7.3) we get that

$$\operatorname{dist}(Q_i, Q_j) \leq C \operatorname{diam} Q_i\,. \tag{7.17}$$

If we fix i, then there can be only a bounded number of j's for which (7.16) and (7.17) are valid, because the Q_j's have disjoint interiors. Lemma 7.15 follows from this.

Lemma 7.18. *$\mu(B(x,r)) \leq \nu(B(x,r)) \leq C\,\mu(B(x,r))$ whenever $x \in S$ and $r > 0$.*

Let $x \in S$ and $r > 0$ be given. The first inequality is trivial. For the second it suffices to show that

$$\nu(B(x,r)\backslash S) \leq k\,\mu(B(x,k\,r)) \tag{7.19}$$

for some constant k, since μ is doubling on S.

Set $J = \{i \in I:\ Q_i \cap B(x,r) \neq \varnothing\}$. Then

$$\nu(B(x,r)\backslash S) \leq \sum_{i\in J} \mu(\beta_i)\,, \tag{7.20}$$

by the definition (7.7) of ν. Lemma 7.15 permits us to convert this into

$$\nu(B(x,r)\backslash S) \leq C\,\mu(\bigcup_{i\in J} \beta_i)\,. \tag{7.21}$$

Thus we are reduced to proving that

$$\bigcup_{i\in J} \beta_i \subseteq S \cap B(x, C\,r)\,. \tag{7.22}$$

Of course the β_i's are all contained in S, by their definition, and so it is just a question of showing that $\beta_i \subseteq B(x, C\,r)$ for all $i \in J$. If $i \in J$, then $\operatorname{dist}(Q_i, S) \leq \operatorname{dist}(Q_i, x) \leq r$. Hence $\operatorname{diam} Q_i \leq r$, by (7.1), and so $|p_i - x| \leq C\,r$, by (7.3) and (7.2). We also get that the radius of β_i is bounded by $C\,r$, by (7.5) and (7.2). Therefore $\beta_i \subseteq B(x, C\,r)$, and (7.22) follows. Of course (7.19) follows from (7.21) and (7.22), and so the proof of Lemma 7.18 is complete.

Lemma 7.23. *ν is a doubling measure on $\mathbb{R}^n$.*

Let $x \in \mathbb{R}^n$ and $r > 0$ be given. We want to prove that

$$\nu(B(x, 2r)) \leq C\,\nu(B(x, r))\,. \tag{7.24}$$

If $x \in S$, then this follows from Lemma 7.18 and the doubling condition for μ. Thus we may assume that $x \in \mathbb{R}^n \backslash S$.

Suppose that $r \geq 2 \operatorname{dist}(x, S)$. Pick a point $z \in S$ such that $|x - z| = \operatorname{dist}(x, S)$. Then

$$\nu(B(x, r)) \geq \nu(B(z, r/2))\,. \tag{7.25}$$

Since $z \in S$ we can use the preceding case to conclude that

$$\nu(B(z, r/2)) \geq C^{-1} \nu(B(z, 3r))\,. \tag{7.26}$$

Clearly $B(z, 3r) \supseteq B(x, 2r)$, and so we get (7.24) in this case.

Now suppose that $r \leq 10^{-3} \operatorname{dist}(x, S)$. Fix a Whitney cube Q_i such that $x \in Q_i$. Then $\operatorname{dist}(x, S) \leq 6 \operatorname{diam} Q_i$, by Lemma 7.8, and therefore every element z of $B(x, 2r)$ satisfies $\operatorname{dist}(z, Q_i) \leq 2\,r \leq \operatorname{diam}(Q_i)/50$. This means that if $j \in I$ and Q_j intersects $B(x, 2r)$, then $C^{-1}\mu(\beta_i) \leq \mu(\beta_j) \leq C\mu(\beta_i)$, by Lemma 7.12. For this set of j's –let us call it J– we also have that $|Q_j|$ is comparable to $|Q_i|$, because of Lemma 7.10. Of course $B(x, 2r)$ does not intersect S in this case, and so we get

$$\begin{aligned}
\nu(B(x, 2r)) &= \sum_{j \in J} \frac{\mu(\beta_j)}{|Q_j|}\,|B(x, 2r) \cap Q_j| \\
&\leq C \sum_{j \in J} \frac{\mu(\beta_i)}{|Q_i|}\,|B(x, 2r) \cap Q_j| \\
&= C\,\frac{\mu(\beta_i)}{|Q_i|}\,|B(x, 2r)|
\end{aligned}$$

$$
\begin{aligned}
(7.27) \qquad & \leq C \, \frac{\mu(\beta_i)}{|Q_i|} \, |B(x,r)| \\
&= C \sum_{j \in J} \frac{\mu(\beta_i)}{|Q_i|} \, |B(x,r) \cap Q_j| \\
&\leq C \sum_{j \in J} \frac{\mu(\beta_j)}{|Q_j|} \, |B(x,r) \cap Q_j| \\
&= \nu(B(x,r)) \, .
\end{aligned}
$$

(In brief, ν is comparable in size to $\mu(\beta_i)/|Q_i|$ times Lebesgue measure on $B(x,2r)$. We shall use this again in the proof of Lemma 7.35 below.) Thus we have (7.24) under these circumstances as well.

We are left with the case where $10^{-3}\mathrm{dist}\,(x,S) < r < 2\,\mathrm{dist}\,(x,S)$. Again choose $i \in I$ so that $x \in Q_i$, and observe that

$$
\nu(B(x,r)) \geq \frac{\mu(\beta_i)}{|Q_i|} \, |B(x,r) \cap Q_i| \geq C^{-1} \, \mu(\beta_i) \, . \tag{7.28}
$$

This uses Lemma 7.8 too. In this case there is a constant $k > 1$ such that

$$
k \, \beta_i \supseteq B(x, 2r). \tag{7.29}
$$

Indeed, the radius of β_i is comparable to $\mathrm{diam}\, Q_i$, and hence to r, by Lemma 7.8, and the distance from x to β_i is bounded by $C \,\mathrm{diam}\, Q_i \leq C\, r$, by (7.3), (7.2), and (7.1). The inclusion (7.29) follows from these facts. The doubling condition for μ together with Lemma 7.18 implies that

$$
\mu(\beta_i) \geq C^{-1} \, \mu(k\,\beta_i) \geq C^{-1} \, \nu(k\,\beta_i) \geq C^{-1} \, \nu(B(x,2r)) \, . \tag{7.30}
$$

Thus we get (7.24) from combining (7.28) and (7.30).

This completes the proof of Lemma 7.23.

Remark 7.31. For the proof of Lemma 7.23 we did not need to know that μ is a *metric* doubling measure on S. In other words, if S is a strong set and μ is doubling on S, and if we define ν as above, then ν is a doubling measure on $\mathbb{R}^n$.

It remains to prove that ν is a metric doubling measure. Let $\delta_\mu(x,y)$ be defined for $x, y \in S$ as in (1.20), and let $\delta_\nu(x,y)$ be defined for all $x, y \in \mathbb{R}^n$ in the analogous manner. Our assumption that

μ be a metric doubling measure on $\mathbb{R}^n$ means that there is a metric $d_\mu(x,y)$ on S such that

$$C^{-1}\, d_\mu(x,y) \le \delta_\mu(x,y) \le C\, d_\mu(x,y)\,, \tag{7.32}$$

for all $x, y \in S$.

There is a simple compatibility property between $\delta_\mu(x,y)$ and $\delta_\nu(x,y)$, which is given by the following.

Lemma 7.33. *$\delta_\mu(x,y) \le \delta_\nu(x,y) \le C\,\delta_\mu(x,y)$ for all $x, y \in S$.*

This is an easy consequence of Lemma 7.18 and the definitions.

We shall prove that ν is a metric doubling measure using the following criterion.

Lemma 7.34. *In order to show that ν is a metric doubling measure on $\mathbb{R}^n$ it suffices to show that there is a constant C_0 so that*

$$\delta_\nu(x_1, x_k) \le C_0 \sum_{i=1}^{k-1} \delta_\nu(x_i, x_{i+1})\,, \tag{7.35}$$

for any finite sequence $\{x_i\}_{i=1}^k$ of points in $\mathbb{R}^n$, $k \ge 2$. (Of course C_0 is not permitted to depend on k.)

This is Lemma 3.1 in [S1]. It is proved by taking $d(x,y)$ to be the infimum of $\sum_{i=1}^{k-1} \delta_\nu(x_i, x_{i+1})$ over all finite sequence $\{x_i\}_{i=1}^k$ of points in $\mathbb{R}^n$, $k \ge 2$, which connect x to y. The inequality (1.21) follows then from (7.35), and $d(x,y)$ is a metric because the triangle inequality is built into its definition.

Note that the sufficient condition of the lemma is also necessary.

The proof that ν satisfies this criterion is not difficult but neither is it so lovely. We begin with some minor technical observations.

Lemma 7.36. *$\delta_\nu(x,y)$ is a quasimetric, i.e., there is a constant $C > 0$ so that*

$$\delta_\nu(x,z) \le C\,(\delta_\nu(x,y) + \delta_\nu(y,z))\,, \tag{7.37}$$

for all $x, y, z \in \mathbb{R}^n$.

This is a straightforward consequence of the doubling property for ν.

Lemma 7.38. *Define $B(p)$ for $p \in \mathbb{R}^n$ by $B(p) = \overline{B}(p, 10^{-4}\mathrm{dist}\,(p,S))$. (Thus $B(p) = \{p\}$ when $p \in S$.) Then there is a number $\lambda = \lambda(p)$ and a constant C such that*

$$C^{-1}\,\lambda\,|x-y| \le \delta_\nu(x,y) \le C\,\lambda\,|x-y|\,, \tag{7.39}$$

for all $x, y \in B(p)$.

Indeed, let p be given as above, and assume that $p \notin S$, since otherwise the lemma is trivial. Choose $i \in I$ so that $p \in Q_i$. The same sort of argument as used in the paragraph containing (7.27) yields

$$C^{-1}\,\frac{\mu(\beta_i)}{|Q_i|}\,|A| \le \nu(A) \le C\,\frac{\mu(\beta_i)}{|Q_i|}\,|A| \tag{7.40}$$

when $A \subseteq 10\,B(p)$. Once we have this we get (7.39) from the definition of $\delta_\nu(x,y)$, with $\lambda = (\mu(\beta_i)/|Q_i|)^{1/n}$. This proves Lemma 7.38.

In Lemma 7.38 we do not have any control over the number λ, but we do not care. Once we know that $\delta_\nu(x,y)$ is comparable to a multiple of the Euclidean metric on $B(p)$ we have the information that we need. (All we really need to know is that it is comparable to some metric there.)

Let us now start to prove that ν is a metric doubling measure. Let a finite sequence $\{x_i\}_{i=1}^k$ of points in $\mathbb{R}^n$ be given, as in Lemma 7.34, and let us try to prove (7.35).

Lemma 7.41. *We can find a subsequence $\{y_i\}_{i=1}^j$ of $\{x_i\}_{i=1}^k$ (i.e., the y_i's are taken from the x_i's, with no repetitions, and the ordering is preserved) with the following properties:*

$$y_1 = x_1\,, \qquad y_j = x_k\,, \tag{7.42}$$

$$\text{there exists } 1 \le l \le j \text{ such that } y_{i+1} \notin B(y_i) \text{ when } 1 \le i < l \text{ and } y_i \in B(y_l) \text{ when } i \ge l\,, \tag{7.43}$$

$$\sum_{i=1}^{j-1} \delta_\nu(y_i, y_{i+1}) \le C \sum_{i=1}^{k-1} \delta_\nu(x_i, x_{i+1})\,. \tag{7.44}$$

This is pretty easy to prove. Let us first choose some integers $\alpha(m)$ as follows. Set $\alpha(1) = 1$. If $x_i \in B(x_1)$ for all $i > 1$ then we stop, otherwise we choose $\alpha(2)$ to be the smallest $i > 1$ such that $x_i \notin B(x_1)$. If $\alpha(1), \ldots, \alpha(m)$ have been chosen already then we proceed as follows. If $x_i \in B(x_{\alpha(m)})$ for all $i > \alpha(m)$ then we stop. Otherwise we choose $\alpha(m+1)$ to be the smallest integer $i > \alpha(m)$ such that $x_i \notin B(x_{\alpha(m)})$. Of course we are always restricting ourselves to $i \leq k$ here.

Let l denote the largest value of m for which $\alpha(m)$ is defined. Set $y_i = x_{\alpha(i)}$ when $1 \leq i \leq l$ and set $y_{l+i} = x_{\alpha(l)+i}$ for as long as this make sense. More precisely, we do nothing for the second definition if $\alpha(l) = k$, and otherwise we use it for $1 \leq i \leq k - \alpha(l)$. This defines our subsequence $\{y_i\}_{i=1}^{j}$, with $j = l + k - \alpha(l)$.

It is not hard to check that (7.42) holds, by construction. We also have (7.43) automatically from our construction.

Let us check (7.44). It suffices to show that

$$\delta_\nu(y_i, y_{i+1}) \leq C \sum_{m=\alpha(i)}^{\alpha(i+1)-1} \delta_\nu(x_m, x_{m+1}) \tag{7.45}$$

when $i < l$. Keep in mind that $y_i = x_{\alpha(i)}$ and $y_{i+1} = x_{\alpha(i+1)}$ here. We may as well assume that $\alpha(i+1) > \alpha(i) + 1$, otherwise (7.45) is trivial. By construction we have that $x_m \in B(x_{\alpha(i)})$ when $\alpha(i) \leq m < \alpha(m+1)$. From Lemma 7.38 we conclude that

$$\delta_\nu(x_{\alpha(i)}, x_{\alpha(i+1)-1}) \leq C \sum_{m=\alpha(i)}^{\alpha(i+1)-2} \delta_\nu(x_m, x_{m+1})\,. \tag{7.46}$$

That is, Lemma 7.38 permits us to get back to the triangle inequality for the Euclidean metric in this case. On the other hand we have that

$$\begin{aligned}\delta_\nu(x_{\alpha(i)}, x_{\alpha(i+1)}) \leq C \big(&\delta_\nu(x_{\alpha(i)}, x_{\alpha(i+1)-1}) \\ &+ \delta_\nu(x_{\alpha(i+1)-1}, x_{\alpha(i+1)})\big)\end{aligned} \tag{7.47}$$

by Lemma 7.36. Combining this with (7.46) yields (7.45). The estimate (7.44) follows from (7.45).

Of course it is very important here that these constants C do not depend on k or l or the x_i's, etc.

This completes the proof of Lemma 7.41.

We let $\{y_i\}_{i=1}^{j}$ and l be as in Lemma 7.41 from now on. In order to prove (7.35) it suffices to show that

$$\delta_\nu(y_1, y_j) \le C \sum_{i=1}^{j-1} \delta_\nu(y_i, y_{i+1})\,. \tag{7.48}$$

This assertion follows from (7.42) and (7.44).

Lemma 7.49. *In order to prove* (7.48) *we may assume that* $l > 1$*, and it suffices to show that*

$$\delta_\nu(y_1, y_l) \le C \sum_{i=1}^{l-1} \delta_\nu(y_i, y_{i+1})\,. \tag{7.50}$$

The point here is that we have replaced the j in (7.48) with l. In particular we may as well assume that $l < j$, otherwise there is nothing to do.

To prove the lemma we observe that

$$\delta_\nu(y_l, y_j) \le C \sum_{i=l}^{j-1} \delta_\nu(y_i, y_{i+1})\,. \tag{7.51}$$

Indeed, we have that $y_i \in B(y_l)$ when $i \ge l$, because of (7.43), and so (7.51) follows from Lemma 7.38. Once we have (7.51) we see that (7.48) is automatic when $l = 1$, and that (7.48) would follow when $l > 1$ if we had (7.50), because of Lemma 7.36. This proves Lemma 7.49.

Thus we assume from now on that $l > 1$, and we want to prove (7.50).

Choose $z_i \in S$, $1 \le i \le l$, so that

$$|z_i - y_i| = \operatorname{dist}(y_i, S) \tag{7.52}$$

for each i. Thus $z_i = y_i$ when $y_i \in S$.

Lemma 7.53. $\delta_\nu(z_i, y_i) + \delta_\nu(z_{i+1}, y_{i+1}) \le C\, \delta_\nu(y_i, y_{i+1})$ *for* $1 \le i < l$.

Indeed, we know from (7.43) that $y_{i+1} \notin B(y_i)$, whence $|y_{i+1} - y_i| \ge 10^{-4}\operatorname{dist}(y_i, S)$. This implies that $|y_{i+1} - y_i| \ge 10^{-5}\operatorname{dist}(y_{i+1}, S)$, as one can check. (If $|y_{i+1} - y_i| < 10^{-5}\operatorname{dist}(y_{i+1}, S)$, then $\operatorname{dist}(y_i, S) \leq 2\operatorname{dist}(y_{i+1}, S)$, etc.) Once we have these inequalities it is not hard to derive Lemma 7.53 from the definition of $\delta_\nu(x,y)$ and the doubling property for ν.

Lemma 7.54. $\delta_\nu(z_1, z_l) \leq C \sum_{i=1}^{l-1} \delta_\nu(y_i, y_{i+1})$.

Indeed, Lemma 7.53 implies that

$$\sum_{i=1}^{l-1} \delta_\nu(z_i, z_{i+1}) \leq C \sum_{i=1}^{l-1} \delta_\nu(y_i, y_{i+1}) . \tag{7.55}$$

Here is where we use our hypothesis that μ is a metric doubling measure. Because the z_i's lie in S we have that $\delta_\mu(z_i, z_{i+1}) \leq \delta_\nu(z_i, z_{i+1})$ for each i, as in Lemma 7.33. Thus

$$\sum_{i=1}^{l-1} \delta_\mu(z_i, z_{i+1}) \leq \sum_{i=1}^{l-1} \delta_\nu(z_i, z_{i+1}) . \tag{7.56}$$

The metric doubling condition for μ (see (7.32)) implies that

$$\delta_\mu(z_1, z_l) \leq C \sum_{i=1}^{l-1} \delta_\mu(z_i, z_{i+1}) . \tag{7.57}$$

Combining these inequalities we get

$$\delta_\mu(z_1, z_l) \leq C \sum_{i=1}^{l-1} \delta_\nu(y_i, y_{i+1}) . \tag{7.58}$$

This implies Lemma 7.54, because of Lemma 7.33.

Lemma 7.59. $\delta_\nu(y_1, y_l) \leq C \sum_{i=1}^{l-1} \delta_\nu(y_i, y_{i+1})$.

This follows from Lemmas 7.54, 7.53, and 7.36.

Lemma 7.59 asserts the validity of (7.50). Lemma 7.49 implies that (7.48) holds, and we saw already that (7.48) implies that (7.35) is true.

Thus we have proved that ν satisfies the criterion for being a metric doubling measure in Lemma 7.34.

This completes the proof of Proposition 1.23.

Acknowledgements. I am grateful to Guy David, Juha Heinonen, Jussi Väisälä, and the referee for some helpful discussions about this matter.

References.

[BA] Beurling, A. and Ahlfors, L., The boundary correspondence under quasiconformal mappings. *Acta Math.* **96** (1956), 125-142.

[DS] David, G., Semmes, S., *Strong A_∞-weights, Sobolev inequalities, and quasiconformal mappings* in *Analysis and Partial Differential Equations.* Edited by C. Sadosky, Lecture Notes in Pure and Applied Mathematics **122**, Marcel Dekker 1990.

[G] Gehring, F. W., The L^p integrability of the partial derivatives of a quasiconformal mapping. *Acta Math.* **130** (1973), 265-277.

[S1] Semmes, S., Bilipschitz mappings and strong A_∞-weights. *Ann. Acad. Sci. Fenn. Series A I Math.* **18** (1993), 211-248.

[S2] Semmes, S., On the nonexistence of bilipschitz parameterizations and geometric problems about A_∞-weights. *Revista Mat. Iberoamericana* **12** (1996), 337–410.

[St] Stein, E. M., *Singular Integrals and Differentiability Properties of Functions.* Princeton Univ. Press, 1970.

[TV] Tukia, P., Väisälä, J., Quasisymmetric embeddings of metric spaces. *Ann. Acad. Sci. Fenn. Series A I Math.* **5** (1980), 97–114.

[V1] Väisälä, J., *Lectures on n-dimensional quasiconformal mappings.* Lecture Notes in Math. **229**, Springer-Verlag, 1971.

[V2] Väisälä, J., Quasi-symmetric embeddings in Euclidean spaces. *Trans. Amer. Math. Soc.* **264** (1981), 191-204.

[V3] Väisälä, J., Porous sets and quasisymmetric maps. *Trans. Amer. Math. Soc.* **299** (1987), 525–533.

[VVW] Väisälä, J., Vuorinen, M., Wallin, H., Thick sets and quasisymmetric maps. *Nagoya Math. J.* **135** (1994), 121-148.

Recibido: 6 de octubre de 1.995
Revisado: 10 de noviembre de 1.995

Stephen Semmes*
Department of Mathematics
Rice University
Houston, Texas 77251, U.S.A.
semmes@math.rice.edu

and

I.H.E.S.
35 Route de Chartres
91440 Bures-sur-Yvette, FRANCE
semmes@ihes.fr

* The author was supported in part by the U.S. National Science Foundation, and is grateful to IHES for its hospitality.

Revista Matemática Iberoamericana
Vol. 12, N.º 3, 1996

A theorem of Semmes and the boundary absolute continuity in all dimensions

Juha Heinonen

Abstract. We use a recent theorem of Semmes to resolve some questions about the boundary absolute continuity of quasiconformal maps in space.

In [S3], Semmes proves that the quasisymmetric image of any set in $\mathbb{R}^d$, $d \geq 2$, of d-measure zero is again of d-measure zero. More formally, if $F \subset \mathbb{R}^d$, $d \geq 2$, and if $h : F \to \mathbb{R}^d$ is a quasisymmetric embedding, then

$$\mathcal{H}_d(F) = 0 \qquad \text{if and only if} \qquad \mathcal{H}_d(h(F)) = 0 \, .$$

Here and hereafter, $\mathcal{H}_p$ denotes the p-dimensional Hausdorff measure for some positive integer p. Recall that an embedding $h : F \to \mathbb{R}^d$ is *quasisymmetric* if there is a homeomorphism $\eta : [0, \infty) \to [0, \infty)$ so that

$$|x - a| \leq t|x - b| \quad \text{implies} \quad |h(x) - h(a)| \leq \eta(t)|h(x) - h(b)| \, ,$$

for all $x, a, b \in F$. Semmes's theorem has important consequences for the problem of boundary absolute continuity of quasiconformal maps in space, as will be explained in this note. In particular, we shall give a positive answer to Problem 1 in [H3, Section 6], and thereby improve upon the main result in [H2].

First we require some definitions. Suppose that $A \subset X \subset \mathbb{R}^n$. We say that A is *linearly locally connected in* X if there is a constant $C \geq 1$ so that, for all $a \in A$ and $r > 0$,

1) points in $A \cap B(a,r)$ can be joined in $X \cap B(a, Cr)$, and
2) points in $A \setminus \overline{B}(a,r)$ can be joined in $X \setminus \overline{B}(a, r/C)$.

Here *joining* means joining by a continuum, $B(z,t)$ is an open n-ball with center z and radius t, and bar denotes the closure. The importance of the concept of linear local connectivity in the quasiconformal mapping theory was observed by Gehring in the sixties (in the nonrelative form, where $A = X$).

A metric space Y is said to be a *bi-Lipschitz p-ball* if there is a bi-Lipschitz homeomorphism φ of the open unit ball $\mathbb{B}^p$ of $\mathbb{R}^p$ onto Y. A metric space E is said to be *contained in a bi-Lipschitz p-ball Y* if there is an isometric embedding $i : E \to Y$.

Finally, recall (from geometric measure theory) that a subset of $\mathbb{R}^n$ is *p-rectifiable* if it is contained in a countable union of Lipschitz images of $\mathbb{R}^p$ plus a set of $\mathcal{H}_p$-measure zero; a set is *purely p-unrectifiable* if it contains no p-rectifiable subset of positive and finite $\mathcal{H}_p$-measure.

Theorem 1. *Suppose that f is a quasiconformal mapping of $\mathbb{B}^n$ onto a domain $D \subset \mathbb{R}^n$, $n \geq 3$, and that $A \subset D$ is bounded, pathwise connected, and linearly locally connected in D. If $E \subset \overline{A} \cap \partial D$ is such that $\mathcal{H}_{n-1}(E) = 0$ and $\mathcal{H}_{n-1}(f^{-1}(E)) > 0$, then there is no bi-Lipschitz $(n-1)$-ball containing E. If $E \subset \overline{A} \cap \partial D$ is such that $\mathcal{H}_{n-1}(f^{-1}(E)) = 0$, then E is purely $(n-1)$-unrectifiable.*

The meaning of $f^{-1}(E)$ will be explained in the proof below; note that *a priori* f^{-1} is not defined on $\overline{A} \cap \partial D$. Also note that the second assertion is non-vacuous only if E has positive, possibly infinite $(n-1)$-measure.

Theorem 1 says, in particular, that if $n \geq 3$, then a quasiconformal homeomorphism $f : \mathbb{B}^n \to D$ preserves the null sets of Hausdorff $(n-1)$-measure on the part of the boundary ∂D that both lies on a bi-Lipschitz $(n-1)$-ball and can be touched from inside of D by a nice subset A. Notice, however, that we do not require that A meet the boundary in rectilinear cones, or anything like it; in principle, A can be a fractal object wildly twisting and spiraling when approaching ∂D.

The first assertion in Theorem 1 is reminiscent of the Bishop-Jones theorem [BJ], which claims that if f is a conformal map of $\mathbb{B}^2$ into the complex plane that maps (via its radial extension) a subset $E \subset \partial\mathbb{B}^2$

of positive length onto a set $f(E)$ of zero length, then $f(E)$ cannot lie on a rectifiable curve. It is well-known that Theorem 1 is false for quasiconformal maps in dimension $n = 2$.

PROOF. The proof is simply a combination of Semmes's aforementioned theorem [S3] and the generalized subinvariance principle [H1, Theorem 6.6]. It follows from [H1, 6.6] that

$$f^{-1} : A \longrightarrow f^{-1}(A)$$

is a quasisymmetric map, and hence extends to a quasisymmetric map

$$f^{-1} : \overline{A} \longrightarrow \overline{f^{-1}(A)}\,.$$

We understand $f^{-1}(E) \subset \overline{f^{-1}(A)}$ precisely as the image of a set $E \subset \overline{A} \cap \partial D$ under this extension, which is uniquely determined by f. Note that the inverse of a quasisymmetric map is quasisymmetric as well, so we have a quasisymmetric map

$$(f^{-1})^{-1} : \overline{f^{-1}(A)} \longrightarrow \overline{A}\,. \tag{2}$$

A couple of remarks need to be made here. The domains D and D' in [H1, 6.6] are assumed to be bounded, but this is a redundant assumption which was unfortunately made in [H1]; if we only assume that A is bounded, the same proof works verbatim. Also, the assumptions on A in [H1, 6.6] are slightly differently phrased, but easily seen to be equivalent to the assumptions of Theorem 1 above.

Suppose now that $E \subset \overline{A} \cap \partial D$ satisfies $\mathcal{H}_{n-1}(E) = 0$. Suppose also that $\varphi : \mathbb{B}^{n-1} \to Y$ is a bi-Lipschitz homeomorphism such that Y contains E, and let $i : E \to Y$ be an isometric embedding. Then the quasisymmetric embedding

$$h = \varphi^{-1} \circ i \circ (f^{-1})^{-1} : f^{-1}(E) \longrightarrow \mathbb{R}^{n-1}$$

maps the set $f^{-1}(E) \subset \partial\mathbb{B}^n$ into $\mathbb{R}^{n-1}$. Here $(f^{-1})^{-1}$ is the map given in (2). Because $n - 1 \geq 2$, we can apply Semmes's result to the map h. (The fact that the set $f^{-1}(E)$ lies on a smooth $(n-1)$-dimensional surface $\partial\mathbb{B}^n$ instead of $\mathbb{R}^{n-1}$ makes of course no difference here.) The conclusion $\mathcal{H}_{n-1}(f^{-1}(E)) = 0$ then follows upon observing that bi-Lipschitz maps preserve the null sets of every Hausdorff measure $\mathcal{H}_p$. This proves the first assertion of the theorem.

The proof of the second assertion follows similarly. We need a theorem about Lipschitz maps (see [F, 3.2.2]): if $h : \mathbb{B}^p \to \mathbb{R}^n$, $2 \leq p \leq n$, is Lipschitz such that the image $h(\mathbb{B}^p)$ has positive p-measure, then there is a subset in $h(\mathbb{B}^p)$ of positive p-measure on which h has a Lipschitz inverse. The rest is definition and Semmes's theorem. This completes the proof of Theorem 1.

A point w lying on the boundary of a domain D in $\mathbb{R}^n$ is said to be an *interior cone point* of D if, for some ball $B \subset D$, the cone

$$wB = \{\lambda w + \mu x \ : \ \ x \in B\,, \ \ \lambda, \mu \geq 0\,, \ \ \text{with} \ \ \lambda + \mu = 1\}$$

lies in $D \cup \{w\}$. Note that the height and the opening of the cone is allowed to depend on w. Denote by $\mathcal{IC}_D$ the subset of ∂D consisting of all the interior cone points of D.

Theorem 3. *Suppose that f is a quasiconformal mapping of $\mathbb{B}^n$ onto a domain $D \subset \mathbb{R}^n$, $n \geq 3$. Then, for any set $E \subset \mathcal{IC}_D$, we have that $\mathcal{H}_{n-1}(E) = 0$ if and only if $\mathcal{H}_{n-1}(\dot{f}^{-1}(E)) = 0$, where $\dot{f}$ denotes the radial extension of f which exists outside a set of n-capacity zero, hence of Hausdorff dimension zero, on $\partial\mathbb{B}^n$.*

Remarks 4. a) The fact that $\dot{f}(z) = \lim_{r\to 1} f(rz)$ exists for $z \in \partial\mathbb{B}^n$ outside an exceptional set of zero n-capacity is well known. Moreover, an easy application of the quasiconformal Lindelöf's theorem shows that, for each $w \in \mathcal{IC}_D$, there is a point $z \in \partial\mathbb{B}^n$ such that $\dot{f}(z) = w$.

b) A weaker version of Theorem 3 where $\mathcal{IC}_D$ is replaced by the set of boundary points admitting both an exterior and interior cone, was proved in [H2] in dimensions $n \geq 3$, $n \neq 4$. Again, the theorem is false in dimension $n = 2$.

Proof. Suppose that $E \subset \mathcal{IC}_D$ with $\mathcal{H}_{n-1}(E) > 0$. It follows by standard arguments (*cf.* [H2, Proof of Theorem 4.3]) that there is a subset in E of positive $\mathcal{H}_{n-1}$ measure that lies on the boundary of a bi-Lipschitz n-ball A contained in D. (The set A is a union of cones of the form wB, $w \in E$, where B has rational radius and rational coordinates for its center.) Because A is a bi-Lipschitz ball, it is linearly locally connected and its boundary is a union of two bi-Lipschitz $(n-1)$-balls. It thus follows from Theorem 1 that $\mathcal{H}_{n-1}(\dot{f}^{-1}(E)) > 0$. A similar argument shows that $\mathcal{H}_{n-1}(\dot{f}^{-1}(E)) = 0$ if $\mathcal{H}_{n-1}(E) = 0$ for $E \subset \mathcal{IC}_D$. The theorem follows.

Theorem 1 is interesting, and new, already in the case when $A = D$, which is equivalent to f being quasisymmetric in all of $\mathbb{B}^n$. A sufficient (but not necessary) condition for this occurrence is that f be *quasiconformally flat*, *i.e.* f extends to a quasiconformal homeomorphism $\mathbb{R}^n \to \mathbb{R}^n$. What is more, Semmes's result brings some new light into the absolute continuity properties of quasisymmetric embeddings of lower dimensional sets into $\mathbb{R}^n$. For completeness, we record the following theorem.

Theorem 5. *Suppose that $E \subset \mathbb{R}^p$ and that $f : E \to \mathbb{R}^n$ is a quasisymmetric embedding, where $2 \le p < n$. If $\mathcal{H}_p(E) > 0$ and $\mathcal{H}_p(f(E)) = 0$, then there is no bi-Lipschitz p-ball containing $f(E)$. If $\mathcal{H}_p(E) = 0$, then $f(E)$ is purely p-unrectifiable.*

PROOF. The first assertion is an immediate corollary of Semmes's theorem; the second assertion likewise reduces to it upon invoking [F, 3.2.2] as in the proof of Theorem 1.

REMARKS 6. a) Gehring [G] (quasiconformally flat case) and Väisälä [V] (the general case) proved that if f is a quasisymmetric embedding of an open set $G \subset \mathbb{R}^p$ into $\mathbb{R}^n$, where $2 \le p < n$, and if the p-measure of $f(G)$ is finite, then f is absolutely continuous; that is, $\mathcal{H}_p(f(E)) = 0$ if $\mathcal{H}_p(E) = 0$ for $E \subset G$. It is not known, even if f is quasiconformally flat, whether f^{-1} is absolutely continuous in this case.

b) Stephen Semmes raised the interesting question whether it is always the case that the quasisymmetric image inside $\mathbb{R}^n$ of a set $E \subset \mathbb{R}^p$ of positive p-measure has positive p-measure. Here $2 \le p < n$. No counterexamples are known to me, and the only positive results that are known assume that the map is defined in a neighborhood of E whose image is an Ahlfors-David p-regular set, *cf.* [S1, 3.4], [H2, 2.7]. In contrast to the equidimensional case required in [S3], the case $p < n$ is not symmetric any more in that a set $E \subset \mathbb{R}^p$ of zero p-measure may easily transform to a set of positive p-measure, or to a set of lower dimension, under a quasisymmetric embedding $f : E \to \mathbb{R}^n$; in fact, it is well known that this can happen for a global quasiconformal self map f of $\mathbb{R}^n$. The point in Semmes's question is that E has positive measure in top dimension. One may also wonder what happens if the range space $\mathbb{R}^n$ is replaced by an arbitrary metric space.

c) A question similar to the one in b) arises in the study of the boundary behavior of a quasiconformal map $f : \mathbb{B}^n \to \mathbb{R}^n$. Thinking of

f being defined capacity almost everywhere on $\partial\mathbb{B}^n$ (*cf.* Remark 4.a)), one may ask how small can the set $f(E)$ be for $E \subset \partial\mathbb{B}^n$ of positive $(n-1)$-measure. To this effect, it was shown in [HK] that the Hausdorff dimension of $f(E)$ is at least $(2K)^{1/(1-n)}$, if f is K-quasiconformal. It is possible that, for $n \geq 3$, the Hausdorff dimension of $f(E)$ has a lower bound that does not depend on K, and it is possible that this bound be $n-1$. If the latter is true, we would have an analog of Makarov's theorem for quasiconformal maps in space.

One is tempted to approach this problem by trying to embed $f(E)$ into $\mathbb{R}^{n-1}$ by a bi-Lipschitz map, if the dimension of $f(E)$ is low. Then Semmes's result would give a contradiction. However, such attempts are futile, as there are countable sets in $\mathbb{R}^3$ that do not admit bi-Lipschitz embeddings into $\mathbb{R}^2$. For example, one can take the three-fold Cartesian product of the sequence $\{1, 1/2, 1/3, \dots\}$. (I learned this example from Jouni Luukkainen.)

d) One could replace throughout the text a bi-Lipschitz p-ball Y by a *quasisymmetric p-ball Y* that has the additional property that all quasisymmetric homeomorphisms $h : \mathbb{B}^p \to Y$ preserve null sets of Hausdorff p-measure. The family of such spaces Y is known to be strictly larger than the family of bi-Lipschitz balls (see [S1], [S2]), but whether or not a given space is in this family is quite difficult to check. This is of course true for bi-Lipschitz balls as well; I simply chose to formulate the theorems in this paper in terms of the latter.

Acknowledgements. I thank Stephen Semmes and the referee for some useful remarks.

References.

[BJ] Bishop, C. J., Jones, P. W., Harmonic measure and arclength. *Ann. Math.* **132** (1990), 511-547.

[F] Federer, H., *Geometric Measure Theory.* Springer, 1969.

[G] Gehring, F. W., Lower dimensional absolute continuity properties of quasiconformal mappings. *Math. Proc. Camb. Phil. Soc.* **78** (1975), 81-93.

[H1] Heinonen, J., Quasiconformal mappings onto John domains. *Revista Mat. Iberoamericana* **5** (1989), 97-123.

[H2] Heinonen, J., The boundary absolute continuity of quasiconformal mappings. *Amer. J. Math.* **116** (1994), 1545-1567.

[H3] Heinonen, J., The boundary absolute continuity of quasiconformal mappings II. *Revista Mat. Iberoamericana* **12** (1996), 697-725.

[HK] Heinonen, J. and Koskela, P., The boundary distortion of a quasiconformal mapping. *Pacific J. Math.* **165** (1994), 93-114.

[S1] Semmes, S., Bilipschitz mappings and strong A_∞-weights. *Ann. Acad. Sci. Fenn. Series A I Math.* **18** (1993), 211-248.

[S2] Semmes, S., On the nonexistence of bilipschitz parameterizations and geometric problems about A_∞-weights. *Revista Mat. Iberoamericana* **12** (1996), 337-410.

[S3] Semmes, S., Quasisymmetry, measure, and a question of Heinonen. *Revista Mat. Iberoamericana* **12** (1996), 727-781.

[V] Väisälä, J., Quasi-symmetric embeddings in Euclidean spaces. *Trans. Amer. Math. Soc.* **264** (1981), 191-204.

Recibido: 20 de noviembre de 1.995

Juha Heinonen*
University of Michigan
Department of Mathematics
Ann Arbor, MI 48109, U.S.A.
juha@math.lsa.umich.edu

* Supported in part by NSF. The author is a Sloan fellow.

Revista Matemática Iberoamericana
Vol. 12, N.º 3, 1996

Analysis on Lie groups

Nick Th. Varopoulos

Dedicated to the memory of my mother

0. Introduction.

0.1. Statement of the theorems.

In what follows G will denote a real connected Lie group and $\Delta = -\sum_{j=1}^{n} X_j^2 + X_0$ will denote some subelliptic left invariant Laplacian (*cf.* [1]). This, for us here, will mean that $X_0, X_1, \ldots, X_n$ are left invariant fields on G (*i.e.* $Xf_g = (Xf)_g$, $f_g(x) = f(gx)$) and that $X_1, \ldots, X_n$ are generators of the Lie algebra of G (*i.e.* together with all their successive brackets they span the Lie algebra of G (*cf.* [2])). I shall denote by $dg = d^\ell g$ the left Haar measure of G and by $d^r g = d^\ell(g^{-1}) = m(g)\, d^\ell g$ the right Haar measure and by $m(g) = m_G(g)$ the modular function.

We can then construct $T_t = e^{-t\Delta}$ $(t > 0)$ the Heat diffusion semigroup and $\phi_t(g)$ the corresponding Heat diffusion kernel that is defined by

$$T_t f(x) = \int_G f(y)\, \phi_t(y^{-1}x)\, dy\,, \qquad t > 0,\ x \in G,\ f \in C_0^\infty(G)\,.$$

When $X_0 = 0$ we say that $\Delta = \Delta_0$ is driftless. A driftless Laplacian Δ_0 is formally self adjoint with respect to $d^r g$. It follows that the modified Laplacian $\tilde{\Delta} = m^{1/2}\Delta_0 m^{-1/2}$ is formally self adjoint with respect to dg. It is then more convenient to consider the modified semigroup $\tilde{T}_t = m^{1/2} e^{-t\Delta_0} m^{-1/2}$ and it is very easy to see (*cf.* [3]) that the

$L^2(G; d^r g) \to L^2(G; d^r g)$ norm of the operator $T_t = e^{-t\Delta_0}$ (which is the same as the $L^2(G; dg) \to L^2(G; dg)$ norm of $\tilde{T}_t$) satisfies

$$\|T_t\|_{2\to 2} = e^{-\lambda t}\,,$$

where λ is the spectral gap of Δ_0 defined by:

$$\lambda = \inf\left\{\int_G |\nabla f|^2\, d^r g \;:\; \int_G f^2\, d^r g = 1\right\}\,,$$

where $|\nabla f|^2 = \sum_{j=1}^n |X_j f|^2$.

In Chapter 1 of this paper we shall give an algebraic classification of $\mathfrak{g}$, the Lie algebra of G, into two classes: the B-algebras and the NB-algebras. We say of course that G is a B- (respectively NB-) group if $\mathfrak{g}$ is a B- (respectively NB-) algebra. We shall refer the reader to Chapter 1 for the precise definition that is algebraically very natural but fairly long to explain. In general terms one considers the minimal parabolic subgroups P (*cf.* [26] for the definition of these subgroups when G is semisimple. Here we extend the notion to general Lie groups by considering "maximal amenable subgroups" or rather a special class of such subgroups). One then considers the corresponding dynamical system $\mathrm{Ad}(P)$ and the classification amounts to the "hyperbolicity" or not of that system. If we denote, here and throughout, by $e \in G$ the neutral element of G we have

Theorem *A. Let G be a Lie group as above and let Δ_0 be a driftless Laplacian, let further $\phi_t \in C^\infty(G)$, $\lambda \geq 0$ be the corresponding heat diffusion kernel and spectral gap respectively. Then*

A_1) *If we assume that G is a* B*-group then there exists C, c_1, $c_2 > 0$ such that*

$$C^{-1}\exp\left(-\lambda\, t - c_2\, t^{1/3}\right) \leq \phi_t(e) \leq C\, \exp\left(-\lambda\, t - c_1\, t^{1/3}\right), \qquad t \geq 1\,. \tag{0.1}$$

A_2) *If we assume that G is an* NB*-group then there exists $C > 0$, $\nu \geq 0$ such that*

$$C^{-1} t^{-\nu} e^{-\lambda t} \leq \phi_t(e) \leq C\, t^{-\nu} e^{-\lambda t}\,, \qquad t \geq 1\,. \tag{0.2}$$

By the standard local Harnack estimate (*cf.* [1], [4]) we can of course replace $\phi_t(e)$ by $\phi_t(g)$ $(g \in G)$ but then the constant $C = C(g) > 0$ depends on g.

Observe that the upper estimate (0.2) with $\nu = 0$ is very easy (*cf.* [3]). The proof that the same index $\nu \geq 0$ can be used for both the upper and the lower estimate in A_2) is very technical. This will be done elsewhere. Here we shall show that some $\nu \geq 0$ exists for which the lower estimate in (0.2) holds.

Another way to write the Heat diffusion semigroup is $T_t f = f * \mu_t$ where $d\mu_t(g) = \phi_t(g)\, d^r g$ is a probability measure on G that in addition has a number of properties that qualifies it to be a "Gaussian" (Gs in short!) measure on G in a sense that we shall make precise in Chapter 3. For any bounded measure μ on G, I shall denote by $\|\mu\|_{2\to 2}$ the $L^2(G; d^r g) \to L^2(G; d^r g)$ norm of the operator $f \to f * \mu$. We have then

Theorem B**.** *Let G be a* B*-group as above and let $\mu \in \mathbb{P}(G)$ be a* Gs*-probability measure on G. Then there exists $c > 0$ such that for all $\varphi \in C_0^\infty(G)$ we have*

$$\langle \varphi, \mu^{*n} \rangle = O\left(\|\mu^n\|_{2\to 2}\, e^{-cn^{1/3}} \right). \tag{0.3}$$

The above clearly (*cf.* [1]) contains the upper estimates of (0.1). It is easy to see that in the estimate (0.3) we can replace $\|\mu^n\|_{2\to 2}$ by $\sigma(\mu)^n$ where $\sigma(\mu) \geq 0$, the spectral radious of μ (*i.e.* the spectral radious of the operator $f \to f * \mu$). The above theorem also holds if we replace the Gs-measure $\mu \in \mathbb{P}(G)$ by some measure that is compactly supported and has continuous (or even just L^2) density. The proof is but an easy modification of the one given in this paper and is if anything easier. The details will however not be given here. Observe finally that for symmetric measures we can easily adapt our methods to give lower estimates in Theorem B that are in the same spirit as Theorem A.

0.2. Guide to the paper.

Chapter 1 is pure algebra and it presents some independent interest. Chapter 2 analyzes the geometry of Lie groups and it shows how the spectral gap can be "isolated" from the rest of the decay of the

Heat kernel. Both the above sections are basic and are likely to play an important role in further developments. Chapter 3 is technical and is only one among many possible approachs to carry out the details of the proof of the upper estimates. The proof of the upper estimate is completed in chapter 4 where a fair amount of global structure theory of Lie group is needed. Observe however that for these upper estimates one needs chapter 2 only up to Section 2.4 and one needs very little algebra (essentially only the definitions of B-groups). A good way for the reader to start with this paper seems to me therefore to go streight for that upper estimate in chapter 4 and refer back to chapter 1, chapter 2 as needed. For the upper estimate one also needs Section B in the appendix.

Almost all of the algebra and the more intricate parts of chapter 2 are only needed for the lower estimates. In the proof of the lower estimate of (0.2) one more (rather unexpected) difficulty arises. The proof as I give it here is considerably easier if Δ is elliptic. The complications that arise when Δ is only subelliptic are quite formidable. This distinction disappears in the alternative, much more sophisticate (at the potential theoretic level), approach that will be used to show that the same ν can be used for the upper and lower estimate at (0.2). This approach will be presented elsewhere. My advise to the reader is therefore to ignore that difficulty and pretend, at least in a first reading, that Δ is elliptic.

The role of the appendix is crucial since it contains all the probability and potential theory that is needed in the rest of the paper. The appendix can (and should) be studied independently, and it has its own independent “guide to the reader” where I explain in particular what is needed for what. Whether it was a good idea to separate the material in this way is of course debatable. One thing is certain, this paper is very long and putting the appendix apart made my life a little easier.

Chapter OV. An overview.

The aim of this chapter (which properly speaking is not part of the paper since it was written after the rest of the writting was completed) is to give to the reader an overview of the subject that is developed in this paper as well as in some of my previous work in the area.

The material is presented here in general terms and with an emphasis on ideas and on the “intuitive picture”. The price that one pays for this is in the precision or even in the accuracy in the presentation. I warn the reader that many assertions made in this chapter are, as such, *incorrect*. The "deviation" from what is actually correct however can be controled and one can say that the aim of this work is to make these ideas into real mathematics.

The only thing that a non expert needs for the reading of this chapter is the definition of a semidirect product $A \leftthreetimes B$ of two groups (*cf.* [6]) and to have some idea of what a random walk and brownian motion is (*cf.* [36]). In reading the first two sections of this chapter the reader could also profit from [37], [38] (in [37] some explanation is offered for the missing page in [38]).

OV.1. The $ax + b$ group.

Let G be the (only non abelian) two dimensional Lie group of affine transformations on $\mathbb{R}$, $\sigma : x \mapsto ax + b$ ($x \in \mathbb{R}$) with $0 < a = e^\alpha \in \mathbb{R}_+^*$ and α, $b \in \mathbb{R}$. This group is the semidirect product $\mathbb{R} \leftthreetimes \mathbb{R}_+^*$ since

$$\sigma_1\sigma_2 : x \longmapsto a_1a_2\, x + b_2\, a_1 + b_1 \ ,$$

where the action of $\mathbb{R}_+^*$ on $\mathbb{R}$ is $b \mapsto ba$.

Let us now consider two probability distributions $\mu^* \in \mathbb{P}(\mathbb{R}_+^*)$, $\mu \in \mathbb{P}(\mathbb{R})$ and let $\nu = \mu \times \mu^*$ be the "product" measure that we obtain on G by putting μ on $\mathbb{R}$ and μ^* on $\mathbb{R}_+^*$.

The beginning of the present work was when several years ago I observed that one could represent the random walk on G generated by ν (alternatively the convolution powers ν^{*n}) in a very simple and managable way. This idea I shall explain in this section.

Let $[g_n = (x_n, s_n^*) \in G \ : \ n \geq 1]$ be the paths of that random walks which formally is defined by $\mathbb{P}\,[g_{n+1} \in dx : \ g_n = y] = d\nu(y^{-1}x)$. By projecting $G \to \mathbb{R}_+^*$ we see that $s_n^* = x_1^* \cdots x_n^*$ performs a (multiplicative) random walk on $\mathbb{R}_+^*$ ($\cong \mathbb{R}$) with transition probability μ^*. The motion of $x_n \in \mathbb{R}$ does not, on the other hand, obey a simple stochastic law and there lies the difficulty of the problem.

The key observation is that once we "fix" (*i.e. condition* in formal probabilistic terms) the path $\omega = (s_1^*, s_2^*, \dots)$ of the random walk on $\mathbb{R}_+^*$ then the motion $x_1, x_2, \dots \in \mathbb{R}$ *also* becomes Markovian. The Markov process that we thus obtain is time inhomogeneous and we have

$$\mathbb{P}\,[x_{n+1} \in dx \ // \ x_n = y\,;\, \omega] = d\mu_n\,(x - y) \ ,$$

where μ_n is the measure on $\mathbb{R}$ that is obtained from μ after the dilatation $x \mapsto s_n^* x$ $(x \in \mathbb{R})$.

This idea, simple though it is, goes a long way. Let us for simplicity make the assumption that $\mu \in \mathcal{N}(0,1)$ is a normal (*i.e.* Gaussian) variable (mean zero and covariance 1). Then $\mu_n \in \mathcal{N}(0, s_n^*)$. From this we can easily estimate the return probability of our random walk

$$p(n) = \mathbb{P}\left[g_n \in [-b_0, b_0] \leftthreetimes [a_0^{-1}, a_0] \subset G\right]$$

for some fixed $0 < b_0 \in \mathbb{R}$, $a_0 \in \mathbb{R}_+^*$. The first step is to estimate the return probability of the conditioned random walk (*i.e.* fixed ω) and this is clearly

$$\text{(OV.1)} \qquad (s_1^* + s_2^* + \cdots + s_n^*)^{-1/2} \sim \left(\int_0^n e^{b(s)}\, ds\right)^{-1/2},$$

where $(b(s) \in \mathbb{R};\ s > 0)$ denotes standard Brownian motion. The reason why we take $e^{b(s)}$ is that $\alpha \mapsto e^\alpha$ is the standard homomorphism between the additive $\mathbb{R}$ and $\mathbb{R}_+^*$. We then clearly have to take the expectation of the expression (OV.1) demanding in addition that $b(n)$ returns to 0 *i.e.* we have the estimate

$$\text{(OV.2)} \qquad p(t) \sim \mathbb{E}\left[\left(\int_0^t e^{b(s)}\, ds\right)^{-1/2} ;\ b(t) \in [-1,1]\right].$$

Brownian functionals as this, have become a big industry these days and are being considered by several authors under the glamorous and sexy label of “Financial Mathematics”. This is as good a name as any for the flavour of the month, I am sure, but all we need is

$$p(t) \sim t^{-3/2}, \qquad t \to \infty.$$

Instead of considering a random walk on G we can take an analytic point of view. We should consider then X, X_* two invariant unit fields along the one parameters subgroups $\mathbb{R}$ and $\mathbb{R}_+^*$ of G and $\Delta = -X^2 - X_*^2$ the corresponding invariant Laplacian. If $\phi_t(g)$ denotes the kernel on G of the heat semigroup $e^{-t\Delta}$ generated by Δ we have

$$\phi_t(g) \sim p(t) \sim t^{-3/2}, \qquad t \to \infty.$$

The task that lies ahead is twofold:

1) to show that the above geometric construction generalizes to an arbitrary Lie group,

2) to estimate the corresponding Brownian functionals thus obtained.

Both the above tasks are quite formidable at the technical level. In the rest of this chapter, I shall try as much as possible to clarify the general picture.

OV.2. The barrier problem for random walks.

In this section I shall swich back to standard random walk

$$S_n = X_1 + \cdots + X_n \in \mathbb{R},$$

where X_j are *i.i.d.* Bernoulli (*i.e.* $\mathbb{P}[X_j = \pm 1] = 1/2$) variables. The issue in (OV.2) is of course to estimate

$$p^*(n) = \mathbb{E}\Big[\Big(\sum_{j=1}^{n} e^{S_j}\Big)^{-A} \; ; \; S_n = 0\Big],$$

So that

$$p^*(n) \leq \mathbb{E}\left(e^{-AM_n} \; , \; S_n = 0\right) \sim n^{-3/2},$$

where $M_n = \sup_{1 \leq j \leq n} S_j$ (recall that $S_j \cong -S_j$). The above asymptotic is obtained because the expectation can be explicitely computed. Indeed there exist standard formulas for the probabilities (*cf.* [36])

$$\begin{aligned}\mathbb{P}\big[M_n \leq \lambda \; ; \; S_n = 0\big]& \\ &= \mathbb{P}\big[S_j, \; 1 \leq j \leq n \text{ lies below the barrier } \lambda \; ; \; S_n = 0\big].\end{aligned}$$

We can now use the "maximal oscilation"

$$\operatorname{osc}(t) = \sup_{0 < t_1, t_2 \leq t} |b(t_1) - b(t_2)| \; ; \qquad |t_1 - t_2| \leq 1,$$

which is a very "small" variable (and in particular $\|\exp(\operatorname{osc}(t))\|_{L^1} = O(t)$), and Hölder inequality and we obtain at once that the actual brownian functional (OV.2) (and not just the random walk functional) satisfies

$$p(n) = O\big(n^{-3/2+\varepsilon}\big)$$

for all $\varepsilon > 0$.

We see, in particular, that it is easy to estimate our (OV.2) functional "up to an ε". This phenomenon recurs all the time during the whole theory, *i.e.* the functionals that occur are straight forward to estimate "up to an ε". On the other hand to obtain the exact asymptotics becomes rather involved.

OV.3. A generalization.

The next obvious generalization of the $ax + b$ group is the group $\mathbb{R}^n \ltimes \mathbb{R}$ where the action of $\mathbb{R}$ on $\mathbb{R}^n$ is given by $x \to e^{\alpha} x$ ($\alpha \in \mathbb{R}$, $x \in \mathbb{R}^n$). The analysis that we made for $ax + b$ extends "verbatum" to this group. What we have to estimate is the functional

$$\mathbb{E}\left[\left(\int_0^t e^{b(s)}\,ds\right)^{-n/2} ;\ b(t) \in [-1,1]\right],$$

which, as we already saw is also $\sim t^{-3/2}$ (or at least $O(t^{-3/2})$). We can push this generalization a step further and consider the group

$$G = \mathbb{R}^n \ltimes \mathbb{R}^a = V \ltimes A\,, \tag{OV.3}$$

where for simplicity (and since the essential aspects of the problem do not change by this assumption) we shall assume that the action of A on V is semisimple with real roots, *i.e.* that it is given by $\theta : A \to GL(V)$ where there exists $L_1, \dots, L_n \in A^*$ (the dual space) linear functions, that are normally referred to as "roots", such that

$$\theta(\xi) = \begin{pmatrix} e^{L_1(\xi)} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & e^{L_n(\xi)} \end{pmatrix}, \qquad \xi \in A = \mathbb{R}^a\,.$$

By the same analysis it is then very easy to see that the return probability of the corresponding random walk can be estimated by

$$\begin{aligned} p(t) = \mathbb{E}\Bigg[&\left(\int_0^t e^{L_1[b(s)]}\,ds\right)^{-1/2} \\ &\cdots \left(\int_0^t e^{L_n[b(s)]}\,ds\right)^{-1/2} ;\ |b(s)| \le 1\Bigg], \end{aligned} \tag{OV.4}$$

where now $(b(s) = b_a(s) \in A = \mathbb{R}^a\ ;\ s > 0)$ is the a-dimensional standard brownian motion. I do not know whether the functional (OV.4) will help any one make a lot of money with Asian options at the Chicago Exchange. What I do know is that estimates of functionals like this are not easy to get. What we obtain is that (as $t \to \infty$)

$$p(t) \sim t^{-\alpha} \qquad \text{or} \qquad p(t) \sim e^{-ct^{1/3}}$$

and that it all depends on the geometry of the roots $L_1, \dots, L_n \in A^*$. There are two types of geometries that we need to consider in this context. The first is the linear geometry *i.e.* the invariants under $GL(A)$. This allows us to make the following basic classification. We distinguish first the case when the origin $(0 \in A^*)$ lies in the convex combination of the roots $L_1, \dots, L_n$. We say then that the roots satisfy the C-condition. And then the NC-case (non-C) which is the opposite situation when all the roots lie strictly on one side of a hyperplane in A^*. In the C-case we have $p(t) \sim e^{-ct^{1/3}}$.

To give a glims of what is happening, let us consider the functional (OV.4) under the C-condition for Brownian motion and for the Bernoulli random walk of Section OV.2 (*i.e.* $n = 2$, $a = 1$; and the roots are: $L_1 = +1$, $L_2 = -1$). We have then

$$\begin{aligned}
\mathbb{E}\Big[\Big(\int_0^t e^{b(s)}\,ds\Big)^{-1/2}\Big(\int_0^t e^{-b(s)}\,ds\Big)^{-1/2}\Big]&\\
&\sim \mathbb{E}\Big[\Big(\sum_{j=1}^n e^{S_j}\Big)^{-1/2}\Big(\sum_{j=1}^n e^{-S_j}\Big)^{-1/2}\Big]\\
&\le \mathbb{E}\Big(\sum_{j=1}^n e^{|S_j|}\Big)^{-1/2}\\
&\le \mathbb{E}\Big[\exp\Big(-\frac{1}{2}\sup_{1\le j\le n}|S_j|\Big)\Big]\\
&= \mathbb{E}\big(e^{-m_n}\big)\,,
\end{aligned}$$

and it is well known and easy to see that with $m(t) = (\sup_{0<s<t}|b(s)|)/2$ we have

$$\mathbb{E}\big(e^{-m(t)}\big) \sim e^{-c\,t^{1/3}}\,.$$

In the NC-situation the *Euclidean* geometry (*i.e.* the $O(a;\mathbb{R}) = O(A)$ invariants) of A^* becomes relevant and we have $p(t) \sim t^{-\alpha}$. The exponent α depends on the "geometry" of the cone

$$\Omega = \big[x \in \mathbb{R}^a\ ;\ L_j(x) > 0,\ j = 1, \dots, k\big] \subset A \cong \mathbb{R}^a\,,$$

where $L_1, \dots, L_k$ are the *non* zero roots. Observe incidentally that the NC-condition is equivalent to the fact that $\Omega \neq \varnothing$. When *all* the roots are 0 one should set $\Omega = A$. This geometry, of course, is considered with respect to the Euclidean structure on A. At this point we should go back to the definitions of (OV.4) and observe that a Euclidean structure has to be given on $A = \mathbb{R}^a$, if brownian motion is to be defined. The question arises what that Euclidean structure is and how is it determined from the group G in (OV.3). The answer to this question is simple: we project the measure $\mu \in \mathbb{P}(G)$ that controls our random walk on $A = \mathbb{R}^a$ and take its covariance matrix. This determines the Euclidean structure.

This Euclidean structure is, with hindsight, very natural. It came to me however as a big surprise. First of all this Euclidean structure depends on the random walk on G and not just on G. Therefore $t^{-\alpha}$ depends on the measure μ and, in general, the α varies continously with μ and can be any large enough real value (*e.g.* $\alpha = 10^{\sqrt{3}}\pi + \sqrt{2}$). This contradicts the intuition that we all had in the subject (*cf.* [1], [41]) that lead us to believe that α had to be a 1/2-integer. It is worth noting that in Ph. Bougerol's work [41], a natural scalar product does exist in A, where $G = NAK$ is the Iwasawa decomposition of some semisimple group G. It is given by the Killing form and it gives rise to corresponding α's that are 1/2-integer. Contrary to what was said, that scalar product and the corresponding α is then independent of the particular measure. This contradiction with what was said above is, however, only apparent. Indeed, for a semisimple group G we have $\mu \in \mathbb{P}(G)$ and *not* $\mu \in \mathbb{P}(NA)$. The role of the Killing scalar product is important in our theory also. The cone Ω in (OV.5) should be thought as a generalization of the Weyl chamber of the semisimple theory and the $\alpha \in 1/2\,\mathbb{Z}$ is related with the symmetries of the Weyl group. This aspect of the theory will not be examined in this paper. The other case when $\alpha \in 1/2\,\mathbb{Z}$ is, of course, when G is a unimodular NC-group. For these groups we have $(L_1, L_2, \dots, L_k) = \varnothing$ since unimodularity amounts to $\sum_{j=1}^{k} L_j = 0$. It follows that $\Omega = A$ and therefore α is independent of μ. This unimodular theory was developed with different methods in [1].

For the same reasons as in Section OV.2, the functional (OV.4) is intimately connected with the following "conical barrier" problem. Let $x \in \Omega$ (*cf.* (OV.5)) be fixed, the problem is then to obtain the asymptotics as $t \to \infty$ of

$$p_\Omega(t) = \mathbb{P}_x\left[b_a(s) \in \Omega,\ 0 < s < t\right].$$

The answer is that there exists some $\lambda = \lambda(\Omega) > 0$ such that $p_\Omega(t) \sim t^{-\lambda}$, $(t > 1)$. The proof of this already takes some doing (*cf.* [13]).

Remark. We can bring out the qualitative difference between the C- and NC-geometry at the probabilistic level in the following manner. Consider the region

$$\Omega^* = [x \in \mathbb{R}^a \ ; \ L_i(x) \geq -1 \ ; \ i = 1, \ldots] ,$$

which is a "polyhedron". This polyhedron is bounded (respectively unbounded) under the C- (respectively NC-) condition and what is relevant in both cases is

$$p^*(t) = \mathbb{P}_0[b(s) \in \Omega^* ; \ 0 < s < t] .$$

The point is that while $p^*(t)$ behaves polynomially when Ω^* is unbounded, in the case when Ω^* is bounded we have

$$p^*(t) \sim \mathbb{P}\,[|b(s)| \leq 1 ; \ 0 < s < t] ,$$

which, by the scaling properties of brownian motion $b(s) \in \mathbb{R}^a$, is easily seen to have an exponential behaviour (as $t \to \infty$). This is the underlying reason for the difference in behaviour of the Heat kernel under the two geometries.

OV.4. The amenability of the group.

The analysis that we gave in the previous section extends to all amenable groups. Indeed the model for such a group is a soluble group $P = N \leftthreetimes A = N \leftthreetimes \mathbb{R}^a$ where now N is a general nilpotent group and not just a vector space $V = \mathbb{R}^n$. The root analysis of the action of A on N can be carried out as before and the corresponding brownian functionals can be estimated. The details give rise to considerable technical difficulties (*cf.* [13], [40]) but not fundamentally new ideas are involved.

New ideas are needed to deal with non amenable (*e.g.* semisimple) groups. It is these ideas that are developed in this paper. The first hint of how to go about this is supplied by what was already done. The point is that it is not quite exact that we can model a general amenable group by $P = N \leftthreetimes A$ as above. The correct model is more like $G = P \leftthreetimes K$ where P is the soluble radical and K is a compact semisimple

Levi factor. These groups can thus be thought as a P-principal bundle with a compact base space $G/P = K$. This model generalizes to any connected Lie group G: we can find $P \subset G$ a soluble subgroup such that the homogeneous space G/P is compact. If G is semisimple and $G = NAK$ is the Isawawa decomposition we take $P = NA$. In general we take for P any "Borel" subgroup (here I deviate slightly from the standard terminology). It thus turns out that the correct setting for our theory is to view G as the total space of a P-principal bundle with P soluble and G/P compact. There exists then $N \subset P$ a nilpotent normal subgroup such that $P/N \cong \mathbb{R}^a = A$, and if we quotient G by the action of N we obtain X an $\mathbb{R}^a$-principal bundle. Such a bundle is, topologicaly, trivial, *i.e.* $X \cong \mathbb{R}^a \times K$ (*cf.* [42]). Observe also that X is a genuine fiber bundle and that it does not admit, in general, a natural group structure.

In this fiber bundle representation of G the Laplacian Δ on G is identified with a P-invariant differential operator on the total space of the bundle. It is in this identification that the factor $e^{-\lambda t}$, where λ is the spectral gap of Δ, appears explicitely in the heat diffusion semigroup $e^{-t\Delta}$. It is futile to try to give an intuitive and yet convincing description of how this comes about. But "grosso modo" what happens is that on the fiber bundle $G = P \times K$ (the product is a Borelian trivialization of the bundle) we have to consider both the measures $d^l r \otimes dk$ and $d^r r \otimes dk$ for left and right Haar measure d^l, d^r. This brings out the modular function $m(x) = d^r x / d^l x$ and then, somehow, the action of Δ on m brings out the spectral gap. A similar phenomenom occurs in the construction of the principal series in the representation theory of the semisimple group $G = NAK$. The fact that P is amenable also plays a role here. In the present formalism one should think of $G = NX$ where $X = \mathbb{R}^a \times K$ is the generalization of $A = \mathbb{R}^a$. As for the brownian motion on A it is replaced by the $\mathbb{R}^a$-invariant diffusion on X that is generated on X by D, the image of Δ by $G \to G/N = X$.

The root analysis of the "action of X on N" can be carried out and the region $\Omega = [L_i > 0,\ i = 1, \dots] \subset X$ can be defined as in (OV.5). One can introduce the analogous brownian functionals and use these functionals to estimate the heat diffusion kernel as before.

The final step that remains is to analyze the second order differential operator D on X and the corresponding "brownian motion" that it generates.

OV.5. The Laplacian of an $\mathbb{R}^a$-principal bundle.

Let $X = \mathbb{R}^a \times K$ or $X = \mathbb{Z}^a \times K$ some (trivial) principal bundle with compact base K, and let T be some Markovian operator on X that is invariant by the action of $\mathbb{R}^a$ or $\mathbb{Z}^a$. For instance, we could be looking at the markovian semigroup $T_t = e^{-tD}$ on $X = \mathbb{R}^a \times K$ as in Section OV.4. Observe that D can be expressed in local coordinates as follows

$$\text{(OV.6)} \qquad \begin{aligned} D = \sum a_{ij}(k) \frac{\partial^2}{\partial x_i \partial x_j} + \sum b_i(k) \frac{\partial}{\partial x_i} \\ + \sum c_{i\alpha}(k) \frac{\partial^2}{\partial x_i \partial k_\alpha} + D\Big(k, \frac{\partial}{\partial k}\Big), \end{aligned}$$

where $D(k, \partial/\partial k)$ only involves the local coordinates $(k_1, \dots, k_j, \dots)$ of $k \in K$ and $(x_1, \dots, x_a)$ are the (global) coordinates of $\mathbb{R}^a$.

I shall denote by

$$\text{(OV.7)} \qquad z(n) = (z_R(n), z_K(n)) \in \mathbb{R}^a \times K = X, \qquad n = 1, 2, \dots,$$

the paths of the Markov process generated by T.

The simplest non trivial example of the above set up is clearly $X = \mathbb{Z} \times \{0,1\}$ (*i.e.* when $K = \{0,1\}$) is the two point space). T is the determined by $L = (L(i,j);\ i,\ j = 0,1)$ some 2×2 markovian matrix and by four probability measures $\mu_{i,j} \in \mathbb{P}(\mathbb{Z})$ $(i,\ j = 1,2)$. The Markov chain (OV.7) can then be determined as follows. First the motion of the K-coordinate $z_K(n)$ is a time homogeneous Markov chain with transition matrix L. As for the fiber coordinate $z_R(n)$ it moves accordingly to the law

$$\mathbb{P}\big[z_R(n+1) = \zeta' \,//\, z(n) = (\zeta, i),\, z_K(n+1) = j\big] = \mu_{i,j}(\zeta - \zeta'),$$

for $\zeta,\ \zeta' \in \mathbb{Z}$, $i, j = 0, 1$. In other words, if we condition the base point at time n to be i and at time $n+1$ to be j, then the n^{th} step on $\mathbb{Z}$ is the same as for a random walk with measure $\mu_{i,j}$. Just as in Section OV.1, therefore, if we condition on the path $z_K(0), z_K(1), \dots \in K$, the motion $z_R(n)$ becomes a time inhomogeneous random walk on $\mathbb{Z}$. It is clear, of course, that the above description of the process generalizes when $K = \{0, 1, \dots, n-1\}$ has n-points or when K is an arbitrary compact space.

A typical problem that we shall consider for the above process is the following barrier problem: Find the asymptotic behaviour (as $n \to \infty$) of

$$\mathbb{P}_{z(0)=(0,0)}\left[z_R(j) \geq -1\,,\ 1 \leq j \leq n\right],$$

or more generally when $X = \mathbb{R}^a \times K$ and when $\Omega \subset X$ is a connical domain as in (OV.5) and $x \in \Omega$, find the correct asymptotics of

$$p_\Omega(n) = \mathbb{P}_x\left[z_R(j) \in \Omega\,;\ 1 \leq j \leq n\right]. \tag{OV.8}$$

The above connical (and "twisted" in the bundle) barrier problem is difficult. Not surprisingly the first step consists in finding $\overline{T}$ the "limit operator" on $\mathbb{R}^a$. That operator determines a Markov chain on $\mathbb{R}^a$ ($\overline{z}(n) \in \mathbb{R}^a,\ n = 1, 2, \dots$) that suitably approximates the motion of $z_R(n)$ of our process for large times.

The construction of $\overline{T}$ is not trivial. For instance, when T is given by $T_t = e^{-tD}$ with D as in (OV.7) then the approximating semigroup is $\overline{T}_t = e^{-t\overline{D}}$ with

$$\overline{D} = \sum \overline{a}_{ij}\frac{\partial^2}{\partial x_i \partial x_j} + \sum \overline{b}_i \frac{\partial}{\partial x_i}\,,$$

but it is *not*, in general, true that the coefficients $\overline{a}_{ij} = \int_K a_{ij}(k)\,d\mu(k)$ are the average of the corresponding coefficients of D with respect to, say, the equilibrium measure $\mu \in \mathbb{P}(K)$ of the $z_K(n)$. Finding the above limit operator is a problem from Homogenization Theory (*cf.* [39]). Once we have determined the limit operator we proceed to show that, when we are in the NC-case (and $\Omega \neq \varnothing$), the correct assymptotics in (OV.8) are

$$p_\Omega(t) \sim t^{-\overline{\alpha}}\,,$$

where $\overline{\alpha} = \overline{\alpha}_\Omega$ is the index that corresponds to the cone Ω and the Euclidean structure determined by $\overline{D}$ as in Section OV.3. If we are in the C-case (*i.e.* if $\Omega = \varnothing$) we obtain, as expected, that $p_\Omega(t) \sim e^{-ct^{1/3}}$.

The details of the above procedure will not be given in this paper. Only a crude first approximation is given in the Appendix. The full solution will be given in a second instalment of this work.

Solving the above problems is interesting and rewarding because, among other things, they throw new light to classical homogenization theory.

1. Algebraic considerations.

1.1. Complex soluble algebras and their roots.

In this section we shall denote by $\mathfrak{q}$ a finite dimensional complex soluble Lie algebra (*cf.* [5], [6], [7]) and by $\mathfrak{n} \subset \mathfrak{q}$ its nilradical. We shall denote by $\mathfrak{n}^p = [\dots[\mathfrak{n},\mathfrak{n}]\mathfrak{n}]\dots\mathfrak{n}]$ the p^{th} commutator, $p = 1, 2, \dots,$ and by $\mathfrak{k}_p = \mathfrak{n}^p/\mathfrak{n}^{p+1}$ the corresponding factors. I shall further denote by $W = \mathfrak{k}_1$, $V = \mathfrak{q}/\mathfrak{n}$ and by $\mathfrak{k} = \mathfrak{k}_1 \oplus \mathfrak{k}_2 \oplus \cdots$, $\mathfrak{k}$ is the corresponding graded Lie algebra where, for the canonical multiplication, we have of course $[\mathfrak{k}_j, \mathfrak{k}_i] \subset \mathfrak{k}_{i+j}$.

The adjoint action of $\mathfrak{q}$

$$\operatorname{ad} x : \mathfrak{q} \to \mathfrak{q}\,, \qquad \operatorname{ad}(x)y = [x,y]\,, \qquad x, y \in \mathfrak{q}\,,$$

induces canonically the following actions

$$\operatorname{ad} x : \mathfrak{n}^p \to \mathfrak{n}^p\,, \qquad p \geq 1\,,\, x \in \mathfrak{q}\,, \tag{1.1.1}$$

$$\operatorname{ad} x : \mathfrak{k}_p \to \mathfrak{k}_p\,, \quad p \geq 1\,; \qquad \operatorname{ad} x \,:\, \mathfrak{k} \to \mathfrak{k}\,, \quad x \in \mathfrak{q}\,. \tag{1.1.2}$$

It is also clear that the action (1.1.2) vanishes if $x \in \mathfrak{n}$. It follows therefore that we also have the following natural actions

$$\operatorname{ad} v : \mathfrak{k}_p \to \mathfrak{k}_p\,, \qquad \operatorname{ad} v \,:\, \mathfrak{k} \to \mathfrak{k}\,, \qquad v \in V = \mathfrak{q}/\mathfrak{n}\,, \tag{1.1.3}$$

and in particular

$$\operatorname{ad} v : W \to W\,, \qquad v \in V\,. \tag{1.1.4}$$

V is an abelian Lie algebra, therefore the action (1.1.4) admits the standard root space decomposition

$$W = W_1 \oplus \cdots \oplus W_s\,, \tag{1.1.5}$$
$$W_j = \{x \in W;\ (\operatorname{ad} v - \lambda_j(v))^N x = 0\,;\ v \in V\}\,, \qquad j = 1, 2, \dots, s\,,$$

where $\lambda_j \in V_{\mathbb{C}}^* = \operatorname{Hom}_{\mathbb{C}}[V;\mathbb{C}]$ are the distinct roots of the action ($\lambda_i \neq \lambda_j$, $i \neq j$, *cf.* [5]) and the integer N in (1.1.5) is large enough, say $N = \dim\ W + 10$.

The actual roots $\lambda_1, \dots \lambda_s$ can of course also be defined by the property that for every $j = 1, \dots, s$ there exists $0 \neq \omega \in W$ such that

$$(\lambda_j(v) - \operatorname{ad} v)\omega = 0 \,, \qquad v \in V \,.$$

Analogous root space decompositions exist for the action (1.1.3) so that we have for instance

$$\mathfrak{k}_j = W_1^{(j)} \oplus \cdots \oplus W_{s_j}^{(j)} \,, \qquad j = 1, 2, \dots \,,$$

where the root space $W_r^{(j)}$ has the root $\lambda_r^{(j)}$ and

$$W_r^{(j)} = \sum [W_{i_1}, W_{i_2}, \dots, W_{i_j}] \,,$$

where the summation extends over all indices for which $\lambda_r^{(j)} = \lambda_{i_1} + \cdots + \lambda_{i_j}$ and where in this paper I shall adopt once and for all the following notation

$$[X, Y, \dots, Z] = [\dots[[X, Y]\dots], Z]$$

for a higher commutator.

Because of the above situation, as is customary, we shall sometimes say that

$$\lambda_1, \lambda_2, \dots, \lambda_s \tag{1.1.6}$$

are the simple roots of the adjoint action and

$$\lambda_r^{(j)} = \lambda_{i_1} + \cdots + \lambda_{i_j} \,, \qquad j = 1, \dots, \; r = 1, \dots, s_j \,, \tag{1.1.7}$$

are the multiple roots. It is important in what follows to examine more closely the above roots and to give what amounts to alternative definitions of the above notions.

Since $\mathfrak{q}$ is soluble, the action (1.1.1) (for $p = 1$) can be simultaneously triangulated (*cf.* [5], [6]). In other words we can choose a basis of $\mathfrak{n}$ with respect to which the action (1.1.1) takes the form

$$\operatorname{ad} x = \begin{pmatrix} \nu_1(x) & & * \\ & \ddots & \\ \mathbf{0} & & \nu_t(x) \end{pmatrix} . \tag{1.1.8}$$

In (1.1.8) $\nu_j \in \mathrm{Hom}_{\mathbb{C}}[\mathfrak{q};\mathbb{C}] = \mathfrak{q}^*_{\mathbb{C}}$ are complex linear functionals on $\mathfrak{q}$ that vanish identically on $\mathfrak{n}$ and can thus be identified with elements of $V^*_{\mathbb{C}}$.

By the standard Jordan-Hölder theorem on composition series [8] we see then that up to a different order the $\nu_1, \dots, \nu_t \in V^*_{\mathbb{C}}$ are exactly the roots $\lambda_r^{(j)}$ in (1.1.7).

The third definition of the roots is less elementary. Let $\mathfrak{h} \subset \mathfrak{q}$ be some Cartan subalgebra of $\mathfrak{q}$ (*cf.* [5], [7]) or more generally just some nilpotent subalgebra of $\mathfrak{q}$ that has the additional property (Cartan subalgebras have that property (*cf.* [7])),

$$\mathfrak{n} + \mathfrak{h} = \mathfrak{q}\,. \tag{1.1.9}$$

We can then consider the root space decomposition

$$\mathfrak{n} = \mathfrak{n}_1 \oplus \mathfrak{n}_2 \oplus \cdots \oplus \mathfrak{n}_\ell$$

of the ad-action of $\mathfrak{h}$ on $\mathfrak{n}$ where as before

$$\mathfrak{n}_j = \{y \in \mathfrak{n} \,:\, (\mathrm{ad}\, x - \mu_j(x))^N y = 0, \text{ for all } x \in \mathfrak{h}\} \tag{1.1.10}$$

(*cf.* [5]) with $\mu_j \in \mathfrak{h}^*_{\mathbb{C}} = \mathrm{Hom}_{\mathbb{C}}[\mathfrak{h}, \mathbb{C}]$ as before. The important thing here is that

$$[\mathfrak{n}_i, \mathfrak{n}_j] \subset \mathfrak{n}_k\,,$$

where $\mu_i + \mu_j = \mu_k$ (*cf.* [5]) and that, since the μ_j's vanish identically on $\mathfrak{h} \cap \mathfrak{n}$, we can identify these μ's to elements of $(\mathfrak{h}/\mathfrak{h} \cap \mathfrak{n})^*_{\mathbb{C}} = V^*_{\mathbb{C}}$ because of (1.1.9).

Therefore $\mu_1, \dots, \mu_\ell$ can be identified with elements of $V^*_{\mathbb{C}}$, and by the same composition series arguments (applied to the action of $\mathfrak{h}$ on $\mathfrak{n}$) we can identify the $\mu_1, \dots, \mu_\ell$ (up to a new order) with the roots (1.1.7).

1.2. Real soluble algebras and their roots.

In this section I shall denote by $\mathfrak{q}$ a finite dimensional real soluble Lie algebra and by $\mathfrak{n} \subset \mathfrak{q}$ its nilradical. I shall fix $\mathfrak{h}$ some Cartan subalgebra (or more generally some nilpotent subalgebra that satisfies (1.1.9)) and I shall denote by $\mathfrak{q}_c = \mathfrak{q} \otimes \mathbb{C}$, $\mathfrak{n}_c = \mathfrak{n} \otimes \mathbb{C}$ and $\mathfrak{h}_c = \mathfrak{h} \otimes \mathbb{C}$ the corresponding complexified algebras. It is then well known that $\mathfrak{n}_c$

is the nilradical of $\mathfrak{q}_c$ (it is also true, but irrelevant for us here, that $\mathfrak{h}_c$ is a Cartan subalgebra of $\mathfrak{q}_c$ if $\mathfrak{h}$ is a Cartan subalgebra of $\mathfrak{q}$). Let us also follow the same notations as in Section 1.1 and denote by

$$V = \mathfrak{q}/\mathfrak{n}\,, \qquad W = \mathfrak{n}/[\mathfrak{n},\mathfrak{n}]\,,$$

which are real vector spaces. We then have canonical identifications

$$V_c = V \otimes \mathbb{C} = \mathfrak{q}_c/\mathfrak{n}_c\,, \qquad W \otimes \mathbb{C} = \mathfrak{n}_c/[\mathfrak{n}_c,\mathfrak{n}_c]\,.$$

Relative to the complex algebra $\mathfrak{q}_c$ the roots (1.1.6) can then be identified with (I denote by t and not by s the number of these roots here)

$$\lambda_1,\ldots,\lambda_t \in \mathrm{Hom}_{\mathbb{R}}[V;\mathbb{C}] \tag{1.2.1}$$

that are defined by the property that there exists $0 \neq \omega \in W \otimes \mathbb{C}$ such that

$$(\lambda_j(x) - \mathrm{ad}\,x)\,\omega = 0\,, \qquad \text{for all } x \in V\,.$$

At this stage it is important to introduce a notation that we shall adopt throughout. The real algebra $\mathfrak{q}$ induces a "real structure" (*i.e.* a "complex conjugation", *cf.* [9], [10]) in the complex algebra $\mathfrak{q}_c = \mathfrak{q}\otimes\mathbb{C}$. I shall consider the complex subalgebras (or even complex subspaces) of $\mathfrak{q}_c$ that respect to the above real structure (*i.e.* are stable by the above complex conjugation). I shall reserve the suffix c to indicate by $\mathfrak{a}_c \subset \mathfrak{q}_c$ these subalgebras. This means that $\mathfrak{a}_c = \mathfrak{a}\otimes\mathbb{C}$ for some subalgebra $\mathfrak{a} \subset \mathfrak{q}$. Such subspaces will be called real.

The considerations of Section 1.1 apply to $\mathfrak{q}_c$ and the subalgebra $\mathfrak{h}_c$. The important thing is to "build" in the definition of the roots the above real structure. The key definition needed to do that is that of the real simple roots or simply the real roots when confusion does not arise.

These are $L_1,\ldots,L_s \in V^*$(= the real dual of V) which are the distinct non zero real parts of the roots $\lambda_1,\ldots,\lambda_t$ of (1.2.1) *i.e.*

$$L(v) = \mathrm{Re}\,\lambda(v)\,, \quad v \in V\,.$$

We can of course consider the graded algebra

$$\mathfrak{k}_c = \bigoplus_p (\mathfrak{n}_c)^p/(\mathfrak{n}_c)^{p+1}$$

and the corresponding action of V_c on $\mathfrak{k}_c$. The "multiple" real roots can thus be defined in the obvious way and these are finite linear combinations with positive integer coefficients of the L_j's

$$L_r^{(j)} = L_{i_1} + L_{i_2} + \cdots + L_{i_\alpha} \neq 0 \tag{1.2.2}$$

(with the i's not necessarily all distinct).

Let us denote by

$$\mathcal{L} = \left\{ \sum_{j=1}^{s} \alpha_j L_j \ : \ \alpha_j \geq 0,\ \sum_{j=1}^{s} \alpha_j = 1 \right\} \subset V^*$$

the convex hull in V^* of the real roots with the understanding that $\mathcal{L} = \varnothing$ if

$$\{L_1, L_2, \ldots, L_s\} = \varnothing\,. \tag{1.2.3}$$

Definition. *We shall say that the algebra $\mathfrak{q}$ is a* C-*algebra if* $0 \in \mathcal{L}$, *otherwise* (*if* $0 \notin \mathcal{L}$) *we shall say that $\mathfrak{q}$ is an* NC-*algebra.*

Algebras for which (1.2.3) holds are called R-algebras (*cf.* [11]). R-algebras are in particular NC-algebras. It should finally be observed that in the above definition nothing changes (*i.e.* we obtain the same classification of C-, NC-algebras) if we replace $\mathcal{L}$ by $\overline{\mathcal{L}}$ the corresponding convex hull in V^* of the "multiple" real roots (1.2.2) which are just the non zero real parts $\operatorname{Re}\lambda_r^{(j)}$ of the (multiple) roots $\lambda_r^{(j)}$ in (1.1.7).

Let us recall that quite generally we say that the Lie algebra $\mathfrak{g}$ is unimodular if

$$\operatorname{trace}(\operatorname{ad} x) = 0\,, \qquad x \in \mathfrak{g}\,.$$

It follows at once that if $\mathfrak{q}$ as above is unimodular and satisfies the NC-condition then (1.2.3) holds and $\mathfrak{q}$ is an R-algebra.

1.3. The structure of soluble NC-algebras.

In this section $\mathfrak{n}$, $\mathfrak{h} \subset \mathfrak{q}$ will be as in Section 1.2. All the notations of Section 1.2 will be preserved and we shall consider the root space decomposition

$$\mathfrak{n}_c = \mathfrak{n}_0^{(c)} \oplus \cdots \oplus \mathfrak{n}_\ell^{(c)} \tag{1.3.1}$$

of $\mathfrak{n}_c$ under the ad-action of $\mathfrak{h}_c$. In this decomposition (1.3.1) if $\mu = 0$ is a root as in (1.1.10) we denote by $\mathfrak{n}_0^{(c)}$ the root space corresponding to that root. When $\mu = 0$ is a root then the space $\mathfrak{n}_0^{(c)}$ is a real space and we have

$$\mathfrak{n}_0^{(c)} = (\tilde{\mathfrak{n}}_0)_c \ ,$$

where $\tilde{\mathfrak{n}}_0 \subset \mathfrak{n}$. Observe also that $\mathfrak{h}_c \cap \mathfrak{n}_c \subset \mathfrak{n}_0^{(c)}$ and therefore that

$$\mathfrak{n} \cap \mathfrak{h} \subset \tilde{\mathfrak{n}}_0 \tag{1.3.2}$$

when $\mu = 0$ is not a root we shall abusively set $\tilde{\mathfrak{n}}_0 = 0$.

The other root spaces $\mathfrak{n}_j^{(c)}$ are of course not necessarily real. We shall therefore partition all the roots $\mu_1, \ldots, \mu_\ell$ into disjoint subsets by the equivalence relation

$$\mu_i \sim \mu_j \quad \text{if and only if} \quad \operatorname{Re}\mu_i = \operatorname{Re}\mu_j \tag{1.3.3}$$

and block together the corresponding subspaces. We obtain thus a direct decomposition

$$\mathfrak{n} = \mathfrak{n}_0 \oplus \mathfrak{n}_1 \oplus \cdots \oplus \mathfrak{n}_k \ , \tag{1.3.4}$$

where $(\mathfrak{n}_i)_c = \mathfrak{n}_{i_1}^{(c)} \oplus \cdots \oplus \mathfrak{n}_{i_\alpha}^{(c)}$ with $\mu_{i_1}, \ldots, \mu_{i_\alpha}$ the roots in the equivalence class $\operatorname{Re}\mu_{i_1} = \operatorname{Re}\mu_{i_2} = \cdots = \operatorname{Re}\mu_{i_\alpha} = L_i$.

In the notations of (1.3.4) we shall (abusively) assume that $\mathfrak{n}_0$ may be $= \{0\}$ and will always correspond to the equivalence class $\operatorname{Re}\mu = 0$. We have of course

$$\tilde{\mathfrak{n}}_0 \subset \mathfrak{n}_0 \,, \qquad [\mathfrak{n}_i, \mathfrak{h}] \subset \mathfrak{n}_i \ , \qquad i = 0, 1, \ldots, k \,. \tag{1.3.5}$$

Finally for any two $i, j = 0, 1, \ldots, k$ we have

$$[\mathfrak{n}_i, \mathfrak{n}_j] \subset \mathfrak{n}_p \ , \tag{1.3.6}$$

where in the equivalence class of the roots of $\mathfrak{n}_p$ the real part is the sum of the two corresponding real parts (*cf.* [5]). The following important proposition immediately follows from (1.3.2), (1.3.5), (1.3.6) and (1.1.9).

Proposition. *If we assume that* $\mathfrak{q}$ *is an* NC-*algebra then* $\mathfrak{n}_R = \mathfrak{n}_1 \oplus \cdots \oplus \mathfrak{n}_k$ *is an ideal in* $\mathfrak{q}$ *and* $\mathfrak{q}_R = \mathfrak{n}_0 + \mathfrak{h}$ *is a subalgebra such that*

$\mathfrak{n}_R \cap \mathfrak{q}_R = \{0\}$ *and* $\mathfrak{q} = \mathfrak{n}_R + \mathfrak{q}_R$. *In other words we have a semidirect product decomposition*

$$\mathfrak{q} = \mathfrak{n}_R \leftthreetimes \mathfrak{q}_R \,. \tag{1.3.7}$$

Furthermore, we have

$$[\mathfrak{n}_j, \mathfrak{q}_R] \subset \mathfrak{n}_j \,, \qquad j = 1, \dots, k \,. \tag{1.3.8}$$

It is clear that $\mathfrak{q}_R$ is a soluble R-algebra (*i.e.* it satisfies (1.2.3)). Observe that quite generally if two ideals $\mathfrak{j}_1$, $\mathfrak{j}_2 \subset \mathfrak{q}$ have the property

$$\mathfrak{q}/\mathfrak{j}_i \quad \text{is an R-algebra} \tag{1.3.9}$$

then the ideal $\mathfrak{j}_1 \cap \mathfrak{j}_2$ has the same property. Indeed $\mathfrak{q}/\mathfrak{j}_1 \cap \mathfrak{j}_2$ can be identified to a subalgebra of $\mathfrak{q}/\mathfrak{j}_1 \times \mathfrak{q}/\mathfrak{j}_2$ which is an R-algebra. It follows in particular that the ideal $\mathfrak{n}_R \subset \mathfrak{q}$ can be given an intrinsic characterization and is the smallest ideal $\mathfrak{j}$ that has the property (1.3.9). It is in particular independent of the choice of $\mathfrak{h}$.

We shall finally need to examine more closely the action (1.3.7). The algebra $\mathfrak{q}_R$ is soluble. For every fixed j we can therefore chose a basis over $\mathbb{C}$ on $(\mathfrak{n}_j)_c = \mathfrak{n}_j \otimes_{\mathbb{R}} \mathbb{C}$ in such a way that with respect to that basis we have

$$\operatorname{ad} x = \begin{pmatrix} \rho_1(x) & & * \\ & \ddots & \\ \mathbf{0} & & \rho_{t_j}(x) \end{pmatrix}, \qquad x \in \mathfrak{q}_R \,,$$

where the ρ_k's vanish identically on $\mathfrak{n} \cap \mathfrak{h}$ since $\operatorname{ad} x$ is a nilpotent tranformation for $x \in \mathfrak{n} \cap \mathfrak{h}$. The ρ_k's can thus be identified with elements of $(\mathfrak{h}/\mathfrak{h} \cap \mathfrak{n})^*_{\mathbb{C}} = \operatorname{Hom}_{\mathbb{R}}[\mathfrak{h}/\mathfrak{h} \cap \mathfrak{n}; \mathbb{C}] = (\mathfrak{q}/\mathfrak{n})^*_{\mathbb{C}}$ and can thus be identified with the elements of the equivalence class of the roots μ of (1.3.3) that have a fixed non zero real part.

1.4. A general Lie algebra and the Levi decomposition.

In this section I shall consider a general finite dimensional real Lie algebra $\mathfrak{g}$ and I shall denote by $\mathfrak{n} \subset \mathfrak{q} \subset \mathfrak{g}$ its radical and its nilradical (*cf.* [5], [6]). We can then find $\mathfrak{s} \subset \mathfrak{g}$ a semisimple subalgebra (with the convention that $\mathfrak{s}$ could be $= \{0\}$) such that

$$\mathfrak{g} = \mathfrak{q} \leftthreetimes \mathfrak{s} . \tag{1.4.1}$$

This is of course the standard Levi decomposition and $\mathfrak{s}$ is called a Levi subalgebra of $\mathfrak{g}$ (*cf.* [5], [6], [7]). The following lemma was first proved and successfully exploited by G. Alexopoulos [12]. The proof I give below is different

Lemma (Alexopoulos [12]). *We can find $\mathfrak{h}_0 \subset \mathfrak{q}$ a nilpotent subalgebra such that*

$$\mathfrak{q} = \mathfrak{n} + \mathfrak{h}_0 , \qquad [\mathfrak{s}, \mathfrak{h}_0] = \{0\} . \tag{1.4.2}$$

PROOF. By H. Weyl's theorem (*cf.* [5], [6]) on the semisimplicity of a representation of any semisimple algebra, we can find $\mathfrak{l} \subset \mathfrak{q}$ a subspace such that $\mathfrak{q} = \mathfrak{n} \oplus \mathfrak{l}$ and such that $[\mathfrak{s}, \mathfrak{l}] \subset \mathfrak{l}$. But since $[\mathfrak{g}, \mathfrak{q}] \subset \mathfrak{n}$ (*cf.* [6]) we have $[\mathfrak{s}, \mathfrak{l}] = 0$. This means that

$$\mathfrak{l} \subset \mathfrak{q}_0 = \{x \in \mathfrak{q} \,:\, [\mathfrak{s}, x] = 0\} ,$$

where $\mathfrak{q}_0$ is a subalgebra of $\mathfrak{q}$. If follows in particular that $\mathfrak{t}_0 = \mathfrak{q}_0/\mathfrak{q}_0 \cap \mathfrak{n} = \mathfrak{q}/\mathfrak{q} \cap \mathfrak{n}$. If we set $\mathfrak{h}_0 \subset \mathfrak{q}_0$ to be some Cartan subalgebra of $\mathfrak{q}_0$ we see therefore that all the conditions of the lemma are verified because the canonical image of $\mathfrak{h}_0$ in $\mathfrak{t}_0$ is $\mathfrak{t}_0$ (*cf.* [7]).

The subalgebra $\mathfrak{h}_0$ is not in general a Cartan subalgebra of $\mathfrak{q}$ but what the lemma says is that it satisfies the condition (1.1.9). It follows therefore that we can make all the constructions of Section 1.3 starting from the algebra $\mathfrak{h}_0$. Using this we shall extend our previous definition to general algebras.

Definition. *Let $\mathfrak{q} \subset \mathfrak{g}$ be as above. We shall say that $\mathfrak{g}$ is a* C- (*respectively* NC-) *algebra if $\mathfrak{q}$ is.*

It follows that if with the above definition $\mathfrak{g}$ is an NC-algebra then we can define the ideal $\mathfrak{n}_R \subset \mathfrak{q}$ and decompose $\mathfrak{q} = \mathfrak{n}_R \leftthreetimes \mathfrak{q}_R$ where $\mathfrak{q}_R$ is defined as in proposition of Section 1.3 and depends on the choice of $\mathfrak{h}_0$ (as we already pointed out $\mathfrak{n}_R$ does not depend on that choice). The fact that $[\mathfrak{h}_0, \mathfrak{s}] = \{0\}$ implies that in the decomposition (1.3.1) all the subspaces $\mathfrak{n}_j^{(c)}$ are stable by the ad-action of $\mathfrak{s}$. Therefore it follows (with the notations of Section 1.3) that $\tilde{\mathfrak{n}}_0$ and all the subspaces $\mathfrak{n}_j$ in (1.3.4) are stable by the ad-action of $\mathfrak{s}$

$$[\tilde{\mathfrak{n}}_0, \mathfrak{s}] \subset \tilde{\mathfrak{n}}_0\,, \quad [\mathfrak{n}_j, \mathfrak{s}] \subset \mathfrak{n}_j\,, \qquad j = 0, 1, \ldots, s\,. \tag{1.4.3}$$

We obtain thus the semidirect product decomposition

$$\mathfrak{g} = \mathfrak{n}_R \leftthreetimes (\mathfrak{q}_R \leftthreetimes \mathfrak{s})\,.$$

Observe finally that when $\mathfrak{s}$ is of compact type and therefore $\mathfrak{q}_R \leftthreetimes \mathfrak{s}$ is an R-algebra then $\mathfrak{n}_R$ can be characterized as before as the smallest ideal $\mathfrak{j} \subset \mathfrak{g}$ for which $\mathfrak{g}/\mathfrak{j}$ is an R-algebra. A final observation is in order. We have

$$\mathfrak{q}_R \leftthreetimes \mathfrak{s} = (\mathfrak{n}_0 + \mathfrak{h}_0) \leftthreetimes \mathfrak{s} \tag{1.4.4}$$

and $\mathfrak{n}_0 \subset \mathfrak{q}_R \leftthreetimes \mathfrak{s}$ is an ideal by (1.4.3) and we can consider the projection

$$\pi\,:\, \mathfrak{q}_R \leftthreetimes \mathfrak{s} \to (\mathfrak{q}_R \leftthreetimes \mathfrak{s})/\mathfrak{n}_0 = (\mathfrak{h}_0/\mathfrak{h}_0 \cap \mathfrak{n}_0) \oplus \mathfrak{s} = \pi(\mathfrak{q}_R \leftthreetimes \mathfrak{s}) \tag{1.4.5}$$

i.e. $\mathfrak{s}$ and $\mathfrak{h}_0/\mathfrak{h}_0 \cap \mathfrak{n}_0$ commute in $\pi(\mathfrak{q}_R \leftthreetimes \mathfrak{s})$. This is because $[\mathfrak{g}, \mathfrak{q}] \subset \mathfrak{n}$ (*cf.* [6]) and (1.4.2), (1.4.4).

1.5. A lemma from linear algebra.

In this section I shall consider

$$M_j = D_j + T_j \in M_{n\times n}(\mathbb{C})\,, \qquad j = 1, \ldots, s\,.$$

A finite number of complex invertible matrices where $D_j = \mathrm{Diag}\,(d_1^{(j)}, \ldots, d_n^{(j)})$ is assumed diagonal with diagonal entrees $d_i^{(j)} \neq 0$ and $T_j = (t_{\alpha,\beta}^{(j)})_{\alpha,\beta=1}^n$ is assumed upper triangular *i.e.* we assume that $t_{\alpha,\beta}^{(j)} = 0$ $(j = 1, \ldots, s,\ \alpha \geq \beta)$. I shall set

$$\max_{1\leq j\leq s} \{\|M_j\|, \|M_j^{-1}\|\} = e^u\,, \qquad u \geq 0\,,$$

where $\|\cdot\|$ indicates the operator norm of the matrix (with respect to the canonical hermitian scalar product $\sum z_i\overline{u}_i$ on $\mathbb{C}^n$). I shall also set

$$\delta_j = \max_{1\le i\le n} |d_i^{(j)}|\,, \qquad \delta_1\delta_2\cdots\delta_s = e^\rho\,, \quad \rho\in\mathbb{R}\,.$$

(For our applications (*cf.* (1.6.5), (1.6.7)) we are in fact going to have $\delta_j = |d_i^{(j)}|$, $i=1,2,\dots$)

What will be proved in this section is that there exists C some numerical constant such that

$$\|M_1\cdots M_s\| \le C^n s^n \exp\,(Cnu+\rho)\,. \tag{1.5.1}$$

First of all we shall reduce the proof of (1.5.1) to the special case $\delta_j=1$, $j=1,\dots,s$ where (1.5.1) reduces to

$$\|M_1\cdots M_s\| \le C^n s^n e^{Cnu}\,. \tag{1.5.2}$$

Indeed we clearly have $\delta_j^{-n} \le |d_1^{(j)}\cdots d_n^{(j)}|^{-1} = \det\,(M_j^{-1}) \le e^{nu}$. Therefore $\delta_j^{-1}\le e^u$ and since trivially $\delta_j\le e^u$ the new matrix $\tilde{M}_j = \delta_j^{-1}M_j$ satisfy $\|\tilde{M}_j\|$, $\|\tilde{M}_j^{-1}\| \le e^{2u}$. The $\tilde{\delta}_j$ that correspond to these new matrices clearly satisfy $\tilde{\delta}_j=1$ and we are in the special case. The estimate (1.5.2) for these new matrices immediately implies the general result (1.5.1). It remains to give a proof of (1.5.2). Let us develop the product

$$\prod_{j=1}^{s}(D_j+T_j) = \sum_{\substack{\varepsilon_k=\pm1\\ k=1,\dots,s}} A_1^{(\varepsilon_1)}\cdots A_s^{(\varepsilon_s)}\,, \tag{1.5.3}$$

where $A_j^{(+1)} = D_j$, $A_j^{(-1)} = T_j$, $j=1,2,\dots,s$.

It is clear furthermore that every term of the form $A_1^{(\varepsilon_1)}\cdots A_s^{(\varepsilon_s)}$ is 0 if among the ε_j's we can find at least $n+1$ (-1)'s. It follows that in the summation of the right hand side of (1.5.3) there are at most s^n non zero terms and since we clearly have $\|A_j^{(\varepsilon)}\| \le 2\,e^{Cu}$ our estimate (1.5.2) follows.

In words what the estimate (1.5.1) says is the following: the norm of $\|M_1\cdots M_s\|$ which has the obvious exponential bound e^{su} can in fact be estimated by $\delta_1\cdots\delta_s$ (this in general does grow exponentially in s but it does so in a special way!) multiplied by a polynomial in s.

1.6. The geometric interpretation of the lemma for soluble Lie groups.

In this section we shall consider Q a real soluble connected Lie group (that is not assumed to be simply connected) and let

$$\pi : Q \to GL_n(\mathbb{R}) \tag{1.6.1}$$

a n-dimensional real representation of Q. I shall denote by $\mathfrak{q}$ the Lie algebra of Q and by

$$d\pi : \mathfrak{q} \to \mathfrak{g}\ell_n(\mathbb{R}) = \mathrm{End}_{\mathbb{R}}(\mathbb{R}^n) \tag{1.6.2}$$

the corresponding representation. The above representations can then be complexified and a basis over $\mathbb{C}$ can be chosen on $\mathbb{C}^n$ in such a way that $d\pi(x)$ ($x \in \mathfrak{q}$) is upper triangular

$$d\pi(x) = \begin{pmatrix} \xi_1(x) & & * \\ & \ddots & \\ \mathbf{0} & & \xi_n(x) \end{pmatrix} = m \in M_{n\times n}(\mathbb{C})\,, \ x \in \mathfrak{q}\,. \tag{1.6.3}$$

The ξ_i's are of course elements of $\mathrm{Hom}_{\mathbb{R}}(\mathfrak{q};\mathbb{C})$. If $g = \mathrm{Exp}\,(x) \in Q$ where

$$\mathrm{Exp} : \mathfrak{q} \to Q$$

is the standard exponential mapping from the Lie algebra $\mathfrak{q}$ in the group Q (this exponential mapping is not in general "onto") we have

$$\begin{aligned} \pi(g) &= \exp m = M \\ &= \begin{pmatrix} \Xi_1(g) & & * \\ & \ddots & \\ \mathbf{0} & & \Xi_n(g) \end{pmatrix} \in GL_n(\mathbb{C})\,, \qquad g \in \mathrm{Exp}(\mathfrak{q})\,, \end{aligned}$$

where $\Xi_j(g) = e^{\xi_j(x)}$ and

$$|\Xi_j(g)| = e^{L_j(g)} = e^{\mathrm{Re}\,\xi_j(x)}, \qquad g = \mathrm{Exp}\,(x),\ x \in \mathfrak{q}\,.$$

It follows (since $\mathrm{Exp}(\mathfrak{q})$ generates Q) that $\pi(g)$ can be simultaneously triangulated for all $g \in Q$ and that $g \to \Xi_j(g)$ is a global homomorphism

$Q \to \mathbb{C}^* = \mathbb{C}\backslash\{0\}$ (with the multiplicative structure). This also defines a group homomorphism

$$Q \to \mathbb{R}\,, \qquad g \mapsto L_j(g)\,. \tag{1.6.4}$$

What the estimate (1.5.1) says in this context is that if $g_1, \ldots, g_s \in Q$ are such that $|g_j| \leq u$, $j = 1, \ldots, s$ (*cf.* [1] and Section 3.1 for the definition of $|g| = |g^{-1}|$) and if $L_1 = L_2 = \cdots = L$ and

$$L(g_1 \cdots g_s) \leq \rho\,, \qquad \rho \in \mathbb{R}\,,$$

then

$$\|\pi(g_1 \cdots g_s)\| \leq C^n s^n \exp\left(Cnu + \rho\right). \tag{1.6.5}$$

Let us illustrate the above considerations further in terms of NC-algebras. Let us assume that $\mathfrak{q}$ is a real NC-algebra and let

$$\mathfrak{q} = \mathfrak{n}_R \leftthreetimes \mathfrak{q}_R$$

be the decomposition (1.3.7) that corresponds to some choice of $\mathfrak{h} \subset \mathfrak{q}$. I shall denote by Q the simply connected real Lie group that corresponds to the algebra $\mathfrak{q}$. The analytic subgroup $N_R \subset Q$ that corresponds to the ideal $\mathfrak{n}_R$ is clearly closed and simply connected (*cf.* [6]). We can also construct "ad hoc" Q_R the simply connected Lie group whose algebra is $\mathfrak{q}_R$. The group Q_R acts canonically (as a group of automorphisms) on N_R. Indeed for $\Omega \subset Q_R$ a small enough neighbourhood of the identity we define that action by the obvious inner automorphism. The simple connectedness of Q_R does the rest. We can define thus the semidirect product $N_R \leftthreetimes Q_R$ and the simple connectedness of G implies that we can identify

$$Q = N_R \leftthreetimes Q_R$$

and that Q_R can be identified to the analytic subgroup of Q that corresponds to the subalgebra $\mathfrak{q}_R$. Q_R is thus a closed subgroup.

We can apply our previous considerations to the representations (1.6.1), (1.6.2)

$$\begin{cases} \pi = \mathrm{Ad} : Q_R \to GL(\mathfrak{n}_R)\,, \\ d\pi = \mathrm{ad} : \mathfrak{q}_R \to \mathfrak{g}\ell(\mathfrak{n}_R)\,, \end{cases} \tag{1.6.6}$$

(recall that $\mathrm{Ad}\, g = dI_g|_e$, $I_g x = g^{-1}xg$, $g, x \in G$, *cf.* [6]).

Let us now consider $N_0 \subset Q_R$ the closed subgroup that corresponds to the ideal $\mathfrak{n}_0 \subset \mathfrak{q}_R = \mathfrak{n}_0 + \mathfrak{h}$. For the above example (1.6.6) the elements $d\pi(x)$ ($x \in \mathfrak{n}_0$) are nilpotent transformations and therefore all the ξ_j's of (1.6.3) vanish identically on $\mathfrak{n}_0$. It follows that the L_j's define in (1.6.4) factor through $\kappa : Q_R \to Q_R/N_0$ and can be considered as group homomorphisms

$$L_j : Q/N \cong Q_R/N_0 \to \mathbb{R},$$

where now $N \subset Q$ is the nilradical of Q. The results in Section 1.5 give then here the following estimate:

Let $g_1, \ldots, g_s \in Q_R \subset Q$ and let assume that $|g_j| \leq u$ ($j = 1, \ldots, s$) (observe that $|\ |_Q$ and $|\ |_{Q_R}$ are equivalent *cf.* Chapter 3) and that $L_k[\kappa(g_1 \cdots g_s)] \leq \rho$, $k = 1, \ldots, n$. Then

$$\|\mathrm{Ad}(g_1 \cdots g_s)|_{\mathfrak{n}_R}\| \leq C^n s^n \exp(Cnu + \rho)\,. \tag{1.6.7}$$

The condition $L_1 = L_2 = \cdots$, that was needed for the validity of (1.6.5), is here guaranteed by (1.3.8). Indeed it is on each $\mathfrak{n}_j$, ($j = 1, \ldots, k$) separately, that we apply our Lemma.

Let now $q_1, \ldots, q_s \in Q_R$ be as before and let us assume that: $|q_j| \leq u$, $j = 1, \ldots, s$; $L_k(\kappa(q_1 \cdots q_i)) \leq \rho$, $i = 1, \ldots, s-1$, $k = 1, \ldots, n$. Let further

$$B(r) = \{n \in N : \ |n|_N \leq r\}$$

denote the r-ball in N. We then clearly have

$$\begin{aligned} \mathcal{B} &= B(r)q_1B(r)q_2 \cdots B(r)q_s \\ &= \left(B(r)B(r)^{q_1}B(r)^{q_1q_2} \cdots B(r)^{q_1\cdots q_{s-1}}\right) q_1 \cdots q_s\ , \end{aligned} \tag{1.6.8}$$

where as usual for any group G we set $g^h = hgh^{-1}$ (g, $h \in G$). It follows therefore from (1.6.7) and (1.6.8) (*cf.* [13]) that

$$\mathcal{B} \subseteq B(R)q_1 \cdots q_s\ , \tag{1.6.9}$$

where $R = r\,C^n s^{n+1} \exp(Cnu + \rho)$. The estimate (1.6.9) implies in particular that

$$|n|_N \leq C \exp(C|n|_G), \qquad n \in N \subset Q\,.$$

A fact that as we shall point out in Section 3.1 holds in general.

In the spirit of Section 1.4 the above considerations extend to a general Lie group G that it is not necessarily soluble. No use of this will be made in this paper but since this construction is important in an other related problem (*cf.* [13]) I shall briefly outline this generalization. If we denote by $\mathfrak{g}$ the Lie algebra of the simply connected group G and if $\mathfrak{g}$ is assumed to be an NC-algebra then we can decompose as in Section 1.4 $\mathfrak{g} = \mathfrak{n}_R \leftthreetimes (\mathfrak{q}_R \leftthreetimes \mathfrak{s})$ and this gives the obvious semidirect product decomposition of the simply connected group G associated to $\mathfrak{g}$

$$G = N_R \leftthreetimes (Q_R \leftthreetimes S)\,,$$

where S is semisimple and simply connected. The lemma of Section 1.5 gives then the following:

Let $g_1, \dots, g_s \in Q_R \leftthreetimes S$ and let us assume that S is compact (in other words we are assuming that G is amenable which was the hypothesis in [13]) let further

$$|g_j| \leq u\,, \quad j = 1, \dots, s\,,$$
$$L_k \circ \kappa(g_1 \cdots g_i) \leq \rho\,, \qquad k = 1, \dots\,, \qquad i = 1, 2, \dots, s-1\,,$$

where now κ is the composition (*cf.* [6])

$$Q_R \leftthreetimes S \to (Q_R \leftthreetimes S)/N_0 \cong (Q_R/N_0) \times S \to Q_R/N_0 \cong Q/N \cong \mathbb{R}^d\,.$$

The conclusion of the above hypothesis is then that the estimate (1.6.7) holds. The details will be left to the reader.

Remark. Implicit in the considerations of this section is the definition of the "roots" for a general (not necessarily simply connected) soluble Lie group G. Indeed we have as above

$$\mathrm{Ad} : Q \to GL(\mathfrak{n}_c)\,,$$

where $\mathfrak{n} \subset \mathfrak{q}$ is the nilradical of the Lie algebra $\mathfrak{q}$ of Q. From the above we see that we can simultaneously triangulate Ad so that

$$\mathrm{Ad}(q) = \begin{pmatrix} \Xi_1(q) & & * \\ & \ddots & \\ \mathbf{0} & & \Xi_n(q) \end{pmatrix}, \qquad q \in Q\,,$$

where $\Xi_j : Q \to \mathbb{C}_* = \mathbb{C}\backslash\{0\}$ is a group homomorphism ($\mathbb{C}_*$ has of course the multiplicative group structure). The above defines uniquely

$\rho_j : Q \to \mathbb{R}$ and $\theta_j : Q \to \mathbb{T} = \mathbb{R}\,(\mathrm{mod}\, 2\pi)$, $(j = 1, \ldots, n)$ two group homomorphisms such that $\Xi_j(q) = \exp(\rho_j(q) + i\,\theta_j(q))$ where clearly

$$\rho_j(\mathrm{Exp}\,(x)) = L_j(x)\,, \qquad x \in \mathfrak{q}\,.$$

Using easy standard considerations (involving determinants) we can express the modular function

$$m(q) = \frac{d^r q}{d^\ell q} = \exp\left(\sum_{j=1}^{n} \rho_j(q)\right). \tag{1.6.10}$$

One sees in particular that G is unimodular if and only if $\mathrm{tr}(\mathrm{ad}_{\mathfrak{q}}(x)) = 0$ $(x \in \mathfrak{q})$.

Finally just as before if $Q/N \cong V \times T$ where N is the nilradical of Q with $V \cong \mathbb{R}^m$ and $T \cong \mathbb{T}^k$ then the "roots" ρ_j are defined on V (*i.e.* $\rho_j|_N \equiv 1$).

1.7. Non amenable Lie algebras.

In this section I shall consider $\mathfrak{g}$ a finite dimensional real Lie algebra and I shall denote by $\mathfrak{n} \subset \mathfrak{q} \subset \mathfrak{g}$ its radical and nilradical. The algebra $\mathfrak{g}/\mathfrak{q}$ is then semisimple or zero. Let us recall the following standard

Definition. *We say that $\mathfrak{g}$ is amenable if $\mathfrak{g}/\mathfrak{q}$ is of compact type or zero. Otherwise we say that $\mathfrak{g}$ is non amenable.*

Quite generally the Lie algebra $\mathfrak{g}$ can be written

$$\mathfrak{g} = \mathfrak{q} \leftthreetimes \mathfrak{s}\,,$$

where $\mathfrak{s}$ is some Levi subalgebra (*cf.* [6]) and where of course $\mathfrak{s} \cong \mathfrak{g}/\mathfrak{q}$. When $\mathfrak{s} \neq 0$ we shall apply the Iwasawa decomposition on $\mathfrak{s}$ (*cf.* [9], [10], [14])

$$\mathfrak{s} = \mathfrak{n}_S + \mathfrak{a} + \mathfrak{k}\,,$$

where $\mathfrak{n}_S$ is nilpotent and $\mathfrak{a}$ is abelian and normalizes $\mathfrak{n}_S$ so that $\mathfrak{n}_S + \mathfrak{a}$ is a soluble algebra. As for $\mathfrak{k}$ it is never 0 and it is the Lie algebra of some compact group. If $\mathfrak{g}$ is amenable we have $\mathfrak{n}_S = \mathfrak{a} = 0$. Since $\mathfrak{s}$ normalizes $\mathfrak{q}$ it is clear that

$$\mathfrak{r} = \mathfrak{q} + \mathfrak{n}_S + \mathfrak{a} \subset \mathfrak{g}$$

is a soluble subalgebra of $\mathfrak{g}$ which I shall call an Iwasawa radical of $\mathfrak{g}$. The definition of $\mathfrak{r}$ is not "unique". In general several Iwasawa radicals exist in $\mathfrak{g}$. When $\mathfrak{g}$ is amenable $\mathfrak{q}$ is the only Iwasawa radical. Finally when $\mathfrak{s} = 0$ and $\mathfrak{g} = \mathfrak{q}$ is soluble we shall agree to say that the Iwasawa radical of $\mathfrak{g}$ is $\mathfrak{r} = \mathfrak{q} = \mathfrak{g}$.

We can give now the following basic

Definition. *We shall say that* $\mathfrak{g}$ *as above is a* B*-algebra* (*respectively* NB*-algebra*) *if some Iwasawa radical of* $\mathfrak{g}$ *is a* C*-algebra* (*respectively* NC*-algebra*).

It is not obvious that a non amenable algebra cannot be both a B and an NB-algebra at the same time. But of course as we shall see this cannot be the case and the above definition gives a genuine classification on Lie algebras.

What is well known (but anything but trivial) is that if $\mathfrak{g}$ is semi-simple of non compact type then it is an NB-algebra. This follows from the classification theorems that give the complete description of the reduced roots (*i.e.* the roots of the action of $\mathfrak{a}$ on $\mathfrak{n}_S$).

Let now $\mathfrak{q} = \mathfrak{q}_1 \times \mathfrak{q}_2$ be the direct product two soluble algebras, and let $\mathfrak{n} = \mathfrak{n}_1 \times \mathfrak{n}_2$ be the nilradical. It is clear that the set of real roots L of $\mathfrak{q}$ can be identified with the set $(L_1 \times \{0\}) \cup (\{0\} \times L_2) \subset V^*$ where $L_i \subset V_i^* = (\mathfrak{q}_i/\mathfrak{n}_i)^*$, $i = 1, 2$, are the real roots of $\mathfrak{q}_i$ $(i = 1, 2)$ and $V = V_1 \times V_2 = \mathfrak{q}/\mathfrak{n}$. From this it follows that $\mathfrak{q}$ is an NC-algebra if and only if both $\mathfrak{q}_1$ and $\mathfrak{q}_2$ are.

Let now $\mathfrak{g}_i$, $i = 1, 2$, be two general Lie algebras and let $\mathfrak{g} = \mathfrak{g}_1 \times \mathfrak{g}_2$. It is then clear that $\mathfrak{r} \subset \mathfrak{g}$ is an Iwasawa radical of $\mathfrak{g}$ if and only if $\mathfrak{r} = \mathfrak{r}_1 \times \mathfrak{r}_2$ where $\mathfrak{r}_i$ is an Iwasawa radical of $\mathfrak{g}_i$ $(i = 1, 2)$. From this it follows that (even without knowing that the above definition gives a classification) that $\mathfrak{g}$ is an NB-algebra if and only if $\mathfrak{g}_1$ and $\mathfrak{g}_2$ are both NB-algebras.

By the above definition, if $\mathfrak{g}$ is amenable then $\mathfrak{g}$ is a B-algebra (respectively NB-algebra) if and only if its radical $\mathfrak{q} \subset \mathfrak{g}$ is a C-algebra (respectively NC-algebra). We also have

Proposition. *Let* $\mathfrak{g}$ *be an arbitrary real Lie algebra, let* $\mathfrak{q} \subset \mathfrak{g}$ *be its radical. Let us assume that* $\mathfrak{q}$ *is a* C*-algebra. Then* $\mathfrak{g}$ *is a* B*-algebra.*

Let us also state formally the classifying property of our definition.

Classification. *Let* $\mathfrak{g}$ *be an arbitrary real Lie algebra, then* $\mathfrak{g}$ *cannot be simultaneously a* B- *and an* NB-*algebra.*

The above classification is indirectly an automatic consequence of the main theorem of this paper. A direct algebraic proof can also be given. That algebraic proof does not seem to be very relevant for the rest of this paper and will therefore be deferred until the end of this chapter. The rest of this section will be devoted to the proof of the proposition.

Before I give the proof of the proposition, I shall have to examine more closely the Iwasawa radicals of the Lie algebra $\mathfrak{g}$. Let

$$\mathfrak{r} = \mathfrak{q} + \mathfrak{n}_S + \mathfrak{a} \subset \mathfrak{g}$$

be such an Iwasawa radical where I shall assume throughout in this section that $\mathfrak{g}/\mathfrak{q} \neq 0$ and let us denote by $\overline{\mathfrak{n}} = \mathfrak{n} + \mathfrak{n}_S$. We have then

Lemma. $\overline{\mathfrak{n}}$ *is the nilradical of* $\mathfrak{r}$.

PROOF. I shall denote by $\mathfrak{n}_r \subset \mathfrak{r}$ the nilradical of $\mathfrak{r}$ and I shall prove first that

$$\overline{\mathfrak{n}} \subset \mathfrak{n}_r\,. \tag{1.7.1}$$

To prove (1.7.1) observe first that $\mathfrak{n}$ is the nilradical of $\mathfrak{g}$ and therefore clearly $\mathfrak{n} \subset \mathfrak{n}_r$. We have on the other hand

$$\mathfrak{n}_S = [\mathfrak{n}_S + \mathfrak{a}, \mathfrak{n}_S + \mathfrak{a}]\,. \tag{1.7.2}$$

This holds by the structure theory of semisimple algebras and the construction of the Iwasawa decompositions (*cf.* [9], [14, Proposition 5.10]). The conclusion is that

$$\mathfrak{n}_S = [\mathfrak{n}_S + \mathfrak{a}, \mathfrak{n}_S + \mathfrak{a}] \subset [\mathfrak{r}, \mathfrak{r}] \subset \mathfrak{n}_r\,, \tag{1.7.3}$$

where the last inclusion holds because $\mathfrak{r}$ is soluble (*cf.* [6]). (1.7.1) follows.

Now $\mathfrak{n}_r \cap \mathfrak{q}$ is a nil-ideal of $\mathfrak{q}$ therefore $\mathfrak{n}_r \cap \mathfrak{q} \subset \mathfrak{n}$ (= the nilradical of $\mathfrak{q}$). It follows therefore from (1.7.1) that

$$\mathfrak{n}_r \cap \mathfrak{q} = \mathfrak{n}\,. \tag{1.7.4}$$

Let us now consider the projection

$$\pi : \mathfrak{r} \to \mathfrak{r}/\mathfrak{q} = \mathfrak{n}_S + \mathfrak{a}\,.$$

From general considerations it follows that $\pi(\mathfrak{n}_r)$ is a nilpotent ideal of $\mathfrak{r}/\mathfrak{q}$ and that therefore it lies inside the nilradical of $\mathfrak{r}/\mathfrak{q} = \mathfrak{n}_S + \mathfrak{a}$. That nilradical is exactly $\mathfrak{n}_S$ by (1.7.2) and the fact that, for all $0 \neq x \in (\mathfrak{n}_S + \mathfrak{a}) \backslash \mathfrak{n}_S$, $[x, [x, [\dots [x, \mathfrak{n}_S] \dots] \neq 0$. The conclusion is that

$$\pi(\mathfrak{n}_r) \subset \mathfrak{n}_S$$

and if we combine this with (1.7.4) we deduce that

$$\mathfrak{n}_r \subset \mathfrak{n} + \mathfrak{n}_S = \overline{\mathfrak{n}}$$

our lemma follows.

From the above lemma we see that we have the identification

$$\mathfrak{r}/\mathfrak{n}_r = (\mathfrak{q}/\mathfrak{n}) + \mathfrak{a} = V \tag{1.7.5}$$

Let us now complexify $\mathfrak{n}_c = \mathfrak{n} \otimes \mathbb{C}$, $\overline{\mathfrak{n}}_c = \overline{\mathfrak{n}} \otimes \mathbb{C}$ and consider

$$W = \mathfrak{n}_c/[\mathfrak{n}_c, \mathfrak{n}_c] \subseteq \overline{W} = \overline{\mathfrak{n}}_c/[\overline{\mathfrak{n}}_c, \overline{\mathfrak{n}}_c]\,.$$

The natural (induced by ad-) action of V on $\overline{W}$ that was considered in Section 1.2 stabilizes W. Let us consider the root space decomposition

$$W = W_1 \oplus \cdots \oplus W_m$$

with respect to the above action of $\mathfrak{q}/\mathfrak{n}$ $(\subset V)$ on W. $\lambda_1, \dots, \lambda_m \in \mathrm{Hom}_{\mathbb{R}}(\mathfrak{q}/\mathfrak{n}; \mathbb{C})$ are the corresponding roots. The important thing to observe is that (since $\mathfrak{q}/\mathfrak{n}$ and $\mathfrak{a}$ commute in V!) each root space W_j is stable by the action of $\mathfrak{a}$ $(\subset V)$ and admits thus its proper root space decomposition

$$W_j = W_1^{(j)} \oplus \cdots \oplus W_{m_j}^{(j)} \tag{1.7.6}$$

under that action. The corresponding roots are $\rho_1, \dots, \rho_{m_j} \in \mathrm{Hom}_{\mathbb{R}}$ $[\mathfrak{a}; \mathbb{C}]$, (strictly speaking we need also a "j" index and we should denote these roots by $\rho_i^{(j)} = \rho_i$, $i = 1, 2, \dots, m_j$), and we have

$$\sum_{i=1}^{m_j} \rho_i = 0, \qquad j = 1, \dots, m\,. \tag{1.7.7}$$

To see (1.7.7) observe that the ad-action of $\mathfrak{q}/\mathfrak{n}$ on W and the ad-action of $\mathfrak{s}$ on W commute (since $[\mathfrak{s},\mathfrak{q}] \subset [\mathfrak{g},\mathfrak{q}] \subset \mathfrak{n}$, *cf.* [6]). It follows that the natural ad-action of $\mathfrak{a}$ on W_j extends to a representation of the semisimple Lie algebra $\mathfrak{s}$ on W_j. The trace of such a representation is zero and (1.7.7) follows.

The very definition of (1.1.5) implies that

$$(\operatorname{ad} x - \lambda_j(x))^N \omega = 0\,, \qquad (\operatorname{ad} y - \rho_s(y))^N \omega = 0\,,$$

where $x \in \mathfrak{q}/\mathfrak{n}$, $y \in \mathfrak{a}$, $\omega \in W_s^{(j)}$ (for N large enough). Since the action of $\mathfrak{q}/\mathfrak{n}$ and of $\mathfrak{a}$ commute it follows that

$$W = \sum_{j=1}^{m} \sum_{s=1}^{m_j} W_s^{(j)}$$

is a root space decomposition of W under the action of $V = (\mathfrak{q}/\mathfrak{n}) + \mathfrak{a}$, and that the corresponding roots are

$$\pi_{j,s} : V \ni x + y \longmapsto \lambda_j(x) + \rho_s(y)\,, \qquad x \in \mathfrak{q}/\mathfrak{n},\ y \in \mathfrak{a}\,.$$

(1.7.7) implies therefore that

$$(1.7.8)\quad \sum_{s=1}^{m_j} \pi_{j,s}(x+y) = m_j \lambda_j(x)\,, \qquad j = 1, \ldots, m,\ x \in \mathfrak{q}/\mathfrak{n},\ y \in \mathfrak{a}\,.$$

Let us now, as in our proposition, make the assumption that $\mathfrak{q}$ is a C-algebra and that there exists a non trivial representation of zero

$$(1.7.9)\qquad 0 = \sum_{j=1}^{m} \alpha_j \operatorname{Re} \lambda_j(x)\,, \qquad x \in \mathfrak{q}/\mathfrak{n}\,.$$

But (1.7.8) and (1.7.9) give then a non trivial representation of zero

$$(1.7.10)\qquad 0 = \sum_{j=1}^{m} \frac{\alpha_j}{m_j} \sum_{s=1}^{m_j} \operatorname{Re} \pi_{j,s}(v)\,, \qquad v \in V\,.$$

The final step that is needed to complete proof of our proposition is that the $\pi_{j,s}$'s can be identified to a subset of the roots of $\mathfrak{r}$ (in the sense of Section 1.1) *i.e.* referring to the action of $\mathfrak{r}/\mathfrak{n}_r$ on $\overline{W}$. This is of course easy by the obvious composition series argument and our proposition follows.

Observe that the converse of the above proposition does not hold. Indeed if $\mathfrak{s}$ is semisimple if $\mathfrak{q}$ is an R-algebra and if the semidirect product $\mathfrak{g} = \mathfrak{q} \leftthreetimes \mathfrak{s}$ is not direct then $\mathfrak{g}$ is in general a B-algebra. Indeed the trace of the action of $\mathfrak{s}$ on $\mathfrak{q}$ is zero.

1.8. Unimodular Lie algebras.

Let us recall that a finite dimensional Lie algebra $\mathfrak{g}$ is called unimodular if

$$\operatorname{tr}(\operatorname{ad}_{\mathfrak{g}} x) = 0, \qquad x \in \mathfrak{g}\,.$$

It follows at once that a unimodular Lie algebra that is in addition amenable is an NB-algebra if and only if it is an R-algebra. In this section we shall prove the following

Proposition. *Let $\mathfrak{g}$ be a unimodular Lie algebra. Then either $\mathfrak{g}$ is a* B*-algebra or $\mathfrak{g}$ is the direct product $\mathfrak{g} = \mathfrak{g}_1 \times \mathfrak{s}$ where $\mathfrak{g}_1$ is an* R*-algebra and $\mathfrak{s}$ is either* 0 *or semisimple.*

The proof will be done in several steps. We shall assume that $\mathfrak{g}$ is not soluble and fix once and for all $\mathfrak{g} = \mathfrak{q} \leftthreetimes \mathfrak{s}$ and $\mathfrak{s} = \mathfrak{n}_S + \mathfrak{a} + \mathfrak{k}$ a Levi decomposition of $\mathfrak{g}$ and an Iwasawa decomposition of $\mathfrak{s}$. We shall assume as we may that $\mathfrak{s}$ is not compact. We have then

Lemma. *Let $\mathfrak{g}$, $\mathfrak{q}$, $\mathfrak{n}_S$, $\mathfrak{a}$ be as above and let us assume that $[\mathfrak{q}, \mathfrak{n}_S + \mathfrak{a}] = \{0\}$. Then $\mathfrak{g}$ can be written as a direct product $\mathfrak{g} = \mathfrak{g}_1 \times \mathfrak{s}_1$ where $\mathfrak{g}_1$ is an* R*-algebra and $\mathfrak{s}_1$ is either semisimple or $\{0\}$.*

Proof. Indeed

$$I = \{x \in \mathfrak{s} \,:\, [\mathfrak{q}, x] = 0\} \subset \mathfrak{s}$$

is an ideal in $\mathfrak{s}$ and since $\mathfrak{n}_S + \mathfrak{a} \subset I$ it follows that $\mathfrak{s} = I \times \tilde{\mathfrak{s}}$ where $\tilde{\mathfrak{s}}$ is either $\{0\}$ or a compact semisimple algebra. It suffices therefore to set $\mathfrak{g}_1 = \mathfrak{q} \leftthreetimes \tilde{\mathfrak{s}}$ and $\mathfrak{s}_1 = I$ and our lemma follows.

Lemma. *Let $\mathfrak{g}$, $\mathfrak{q}$, $\mathfrak{n}_S$, $\mathfrak{a}$ be as above and let us assume that $[\mathfrak{n}, \mathfrak{n}_S + \mathfrak{a}] = \{0\}$ where $\mathfrak{n} \subset \mathfrak{q}$ is as before the nilradical of $\mathfrak{g}$. We have then $[\mathfrak{q}, \mathfrak{n}_S + \mathfrak{a}] = \{0\}$.*

PROOF. The semisimple algebra $\mathfrak{s}$ acts by ad on $\mathfrak{q}$ and stabilizes the subspace $\mathfrak{n}$. By H. Weyl's theorem therefore (*cf.* [5], [6]) we can find a direct complement $\mathfrak{q} = \mathfrak{n} \oplus \mathfrak{l}$ such that $[\mathfrak{s}, \mathfrak{l}] \subset \mathfrak{l}$. Since on the other hand we also have $[\mathfrak{s}, \mathfrak{q}] \subset \mathfrak{n}$ it follows that $[\mathfrak{s}, \mathfrak{l}] = \{0\}$ and therefore for all $x \in \mathfrak{s}$, $[x, \mathfrak{q}] \subset [x, \mathfrak{n}]$ and our lemma follows.

Let us now consider the ad-action of $\mathfrak{r}$ on $\overline{\mathfrak{n}} = \mathfrak{n} + \mathfrak{n}_S = \mathfrak{n}_r$ the nilradical of $\mathfrak{r}$ (*cf.* Section 1.7). It clearly stabilizes $\mathfrak{n}$ and, $\mathfrak{r}$ being soluble, a basis can be chosen on $\mathfrak{n}_c$ for which the adjoint action takes the form

$$\mathrm{ad}_{\mathfrak{n}_c}(x) = \begin{pmatrix} \lambda_1(x) & & * \\ & \ddots & \\ \mathbf{0} & & \lambda_p(x) \end{pmatrix}, \qquad x \in \mathfrak{r}\,.$$

since $\mathrm{ad}_{\mathfrak{n}}(x)$ is nilpotent for every $x \in \overline{\mathfrak{n}}$ it follows that we can identify each $\lambda_j \in \mathrm{Hom}_{\mathbb{R}}[V; \mathbb{C}]$ where as in (1.7.5) $V = \mathfrak{r}/\overline{\mathfrak{n}} = (\mathfrak{q}/\mathfrak{n}) + \mathfrak{a}$. We have then

Lemma.

i) *All the λ_j above are real valued on $\mathfrak{a}$, i.e.*

$$\lambda_j(x) \in \mathbb{R}, \qquad j = 1, \dots, p,\ x \in \mathfrak{a}\,.$$

ii) *The trace is zero on $\mathfrak{a}$, i.e.*

$$\mathrm{tr}\,(\mathrm{ad}_{\mathfrak{n}} x) = \sum_{j=1}^{p} \lambda_j(x) = 0\,, \qquad x \in \mathfrak{a}\,.$$

iii) *If we assume that $[\mathfrak{n}, \mathfrak{n}_S + \mathfrak{a}] \neq \{0\}$, then there exists a j $(1 \le j \le p)$, say $j = 1$, for which $\lambda_1(x) \neq 0$ for some $x \in \mathfrak{a}$.*

We shall defer the proof of the lemma until later and complete the proof of the proposition, assuming as we may because of our first two lemmas, that $[\mathfrak{n}, \mathfrak{n}_S + \mathfrak{a}] \neq \{0\}$.

The unimodularity of $\mathfrak{g}$ implies the unimodularity of the algebra $\mathfrak{q}$ which says that

$$\mathrm{tr}\big(\mathrm{ad}_{\mathfrak{n}}(x)\big) = \sum_{j=1}^{p} \lambda_j(x) = 0\,, \qquad x \in \mathfrak{q}/\mathfrak{n}\,.$$

Part ii) of our previous lemma implies therefore that

$$\operatorname{tr}\big(\operatorname{ad}_{\mathfrak{n}}(x)\big) = \sum_{j=1}^{p} \lambda_j(x) = 0\,, \qquad x \in V = \mathfrak{q}/\mathfrak{n} + \mathfrak{a}\,. \tag{1.8.1}$$

Since on the other hand by i) and iii) of our previous lemma and our hypothesis we have $\operatorname{Re}\lambda_1 = L_1 \neq 0$, (1.8.1) says that $\mathfrak{r}$ is a C-algebra. By definition therefore $\mathfrak{g}$ is a B-algebra and the proof of our proposition is complete. (As in the end of Section 1.7 we have to use a standard composition series argument to verify that the non zero among the $\operatorname{Re}\lambda_j$'s can be identified to real roots of $\mathfrak{r}$).

It remains to give the proof of the last lemma.

PROOF OF ii). This uses the same argument as in Section 1.7. Indeed the action of $\mathfrak{a}$ on $\mathfrak{n}$ extends to an action of $\mathfrak{s}$ on $\mathfrak{n}$ *i.e.* to a representation of a semisimple algebra and therefore has trace equal to zero.

To see parts i) and iii) of the lemma we start from the following construction:

Let $\mathfrak{g}$ be some real semisimple algebra and let $\mathfrak{u} \subset \mathfrak{g}$ be some real subalgebra that is a semisimple algebra of compact type. Let further

$$\theta = \mathfrak{g} \longmapsto \mathfrak{gl}_n(\mathbb{C})$$

a real algebra homomorphism. Then there exists $\langle\cdot,\cdot\rangle$ some Hermitian product on $\mathbb{C}^n$ that is invariant under $\mathfrak{u}$, *i.e.*

$$\langle\theta(x)z_1, z_2\rangle + \langle z_1, \theta(x)z_2\rangle = 0\,, \qquad x \in \mathfrak{u},\ z_1, z_2 \in \mathbb{C}^n\,. \tag{1.8.2}$$

The proof of this is of course very easy. Indeed let G be the simply connected semisimple group that corresponds to $\mathfrak{g}$ and let $U \subset G$ be the (compact) subgroup that corresponds to $\mathfrak{u}$. θ induces then

$$\theta : U \longrightarrow GL_n(\mathbb{C})$$

and, since for any non singular matrix $M \in M_{n\times n}(\mathbb{C})$ $\langle z, u\rangle_M = \langle Mz, Mu\rangle$ is a new Hermitian product on $\mathbb{C}^n$, the Hermitian product

$$\langle\langle z, u\rangle\rangle = \int_U \langle\theta(x)z, \theta(x)u\rangle\, dx$$

is invariant under the action of U. Taking the differential we obtain (1.8.2).

The above observation has to be combined with the fact that for the Iwasawa decomposition $\mathfrak{s} = \mathfrak{n}_S + \mathfrak{a} + \mathfrak{k}$ if we complexify $\mathfrak{s}_c$ we can write $\mathfrak{s}_c = \mathfrak{k}_0 \oplus \mathfrak{p}_0$ in such a way that $\mathfrak{u} = \mathfrak{k}_0 + i\,\mathfrak{p}_0$ is a compact real semisimple subalgebra of the (underlying) real algebra $\mathfrak{s}_c$, and we can do so in such a way that $\mathfrak{a} \subset \mathfrak{p}_0$. This fact is of course anything but obvious but in some sense it is the very basis of the construction of the Iwasawa decomposition (*cf.* [9], [14]).

This being said, we see, that if we let $\mathfrak{g}$ in (1.8.2) to be the underlying real algebra of $\mathfrak{s}_c$, there exists $\langle\cdot,\cdot\rangle$ some Hermitian product on $\mathbb{C}^n$ for which all the matrices $\theta(x)$ $(x \in i\,\mathfrak{p}_0)$ are skew-Hermitian. All the matrices $\theta(x)$ $(x \in \mathfrak{a} \subset \mathfrak{p}_0)$ are therefore Hermitian. It follows that all the matrices $\theta(x)$ $(x \in \mathfrak{a})$ have real eigenvalues and if all the eigenvalues of $\theta(x)$ are zero then $\theta(x) = 0$.

If we apply this last observation to the representation of $\mathfrak{s} \otimes \mathbb{C}$ on $\mathfrak{n} \otimes \mathbb{C}$ induced by the adjoint action of $\mathfrak{s}$ on $\mathfrak{n}$ the assertions i) and iii) of the lemma follows.

I shall finish this section with an example that shows that unimodularity is essential for the above proposition to hold. I shall consider the 2-dimensional group of "affine motions" which is the Lie group

(1.8.3) $$G = \mathbb{R}^2 \leftthreetimes (\mathbb{R} \times SL_2(\mathbb{R})) = (\mathbb{R}^2 \leftthreetimes \mathbb{R}) \leftthreetimes SL_2(\mathbb{R})\,,$$

where $\mathbb{R}$ acts on $\mathbb{R}^2$ by dilatation (*i.e.* $x \mapsto e^\rho x$, $x \in \mathbb{R}^2$, $\rho \in \mathbb{R}$) and $SL_2(\mathbb{R})$ acts on $\mathbb{R}^2$ by the natural action (of course $\mathbb{R}$ and $SL_2(\mathbb{R})$ commute in (1.8.3)). The third term in (1.8.3) is of course a Levi decomposition of G and G is not unimodular since the radical $Q = \mathbb{R}^2 \leftthreetimes \mathbb{R}$ is not unimodular. It is clear of course that $\mathfrak{g}$ the Lie algebra of G is not of the form $\mathfrak{g}_1 \times \mathfrak{s}$ as in our proposition. The above algebra $\mathfrak{g}$ however is an NB-algebra and therefore, by the classification in Section 1.7, it is not a B-algebra.

Indeed with the "standard" Iwasawa decomposition of $SL_2(\mathbb{R})$ and the corresponding Iwasawa radical $\mathfrak{r}$ obtained by the Levi decomposition (1.8.3) we have, with our previous notations,

$$\mathfrak{r}/\overline{\mathfrak{n}} = V = R + \mathfrak{a} \cong \mathbb{R}^2\,.$$

Here $\mathfrak{a} \cong \mathbb{R}$ is the $\mathfrak{a}$ component of the Iwasawa decomposition of $SL_2(\mathbb{R})$ and $R \cong \mathbb{R} \cong \mathfrak{q}/\mathfrak{n}$. The action of R on $\mathbb{R}^2$ is of course given by dilatation. This means that the two roots λ_1, λ_2 of $\mathfrak{q}$ (in the sense of sections 1.1 and 1.2) are real and $\lambda_1(x) = \lambda_2(x) \neq 0$ if $0 \neq x \in R$. The algebra $\mathfrak{r}$ is therefore an NC-algebra and our assertion follows.

1.9. The uniqueness of the Iwasawa radical and an intrinsic definition.

In this section I shall prove the following

Proposition. *Let $\mathfrak{g}$ be a real Lie algebra and let $\mathfrak{r}_1$, $\mathfrak{r}_2$ two Iwasawa radicals of $\mathfrak{g}$. Then there exists $\alpha \in \operatorname{Int}(\mathfrak{g})$ such that $\alpha(\mathfrak{r}_1) = \mathfrak{r}_2$.*

This proposition is not essential for the rest of this paper but it does help to give an intrinsic status to the notions introduced in the previous sections. The proof is an inmediate consequence of the following sequence of well known, but highly non trivial, facts:

1) Let $\mathfrak{g} = \mathfrak{q} \leftthreetimes \mathfrak{s}_1 = \mathfrak{q} \leftthreetimes \mathfrak{s}_2$ be two Levi decompositions of $\mathfrak{g}$ then there exists $\alpha : \mathfrak{g} \to \mathfrak{g}$ an inner automorphism of $\mathfrak{g}$ such that $\alpha(\mathfrak{s}_1) = \mathfrak{s}_2$ (*cf.* [6, Theorem 3.14.2]).

Let now $\mathfrak{s}_i = \mathfrak{k}_i + \mathfrak{p}_i$, $i = 1, 2$ be Cartan decomposition of the above two semisimple algebras. By composing, if necessary, the automorphism $\alpha \in \operatorname{Int}(\mathfrak{g})$ by an appropiate element of $\operatorname{Int}(\mathfrak{s}_i)$ we can then assume in addition the following fact (*cf.* [9, Theorem 7.2] of the first edition):

2) The inner automorphism α is such that

$$\alpha(\mathfrak{k}_1) = \mathfrak{k}_2 \,, \qquad \alpha(\mathfrak{p}_1) = \mathfrak{p}_2 \,.$$

Let now $\mathfrak{a}_i \subset \mathfrak{p}_i$, $i = 1, 2$ be a maximal abelian subalgebra and let

$$\mathfrak{s}_i = \mathfrak{k}_i + \mathfrak{a}_i + \mathfrak{n}_i, \qquad i = 1, 2\,,$$

the Iwasawa decompositions that correspond to these choices of $\mathfrak{a}_i$ and to some choice of $\Sigma_+^i \subset \mathfrak{a}_i^*$ $(= \operatorname{Hom}_{\mathbb{R}}(\mathfrak{a}_i; \mathbb{R}))$ the positive restricted roots on $\mathfrak{a}_i$ (*i.e.* the finitely many choices of the corresponding Weyl chambers). By the standard facts concerning the Iwasawa decomposition we see therefore that we can further compose the $\alpha \in \operatorname{Int}(\mathfrak{g})$ by an appropiate element in $\operatorname{Int}(\mathfrak{k}_i)$ and guarantee the following additional fact (*cf.* [14, Section 5.13 and Corollary 5.18]):

3) The inner automorphism $\alpha \in \mathrm{Int}(\mathfrak{g})$ is such that

$$\alpha(\mathfrak{k}_1) = \mathfrak{k}_2\ , \quad \alpha(\mathfrak{a}_1) = \mathfrak{a}_2\ , \quad \alpha(\mathfrak{n}_1) = \mathfrak{n}_2\ .$$

The final conclusion clearly is that $\alpha(\mathfrak{r}_1) = \mathfrak{r}_2$ and this proves our proposition.

I shall finish up this section by giving, without proofs, what amounts to an alternative, more intrinsic, but less manageable, definition of the Iwasawa radical. Let $\mathfrak{g}$ be an arbitrary real Lie algebra and let $\mathfrak{l} \subset \mathfrak{g}$ be a amenable subalgebra, such that for some Lie group G that corresponds to $\mathfrak{g}$, $\mathfrak{l}$ corresponds to a closed subgroup L such that $G = L \cdot K$ where K is a compact subgroup. We can then show that $\mathfrak{l}$ is a C- (respectively NC-) algebra if and only if $\mathfrak{g}$ is a B- (respectively NB-) algebra. This, among other things is a consequence of the analytic theory developed in this paper. The Iwasawa radical clearly has the above property. Other examples of such subalgebras are $\mathfrak{l} = \pi^{-1}$ (a minimal boundary subalgebra of $\mathfrak{g}/\mathfrak{q}$) with $\pi : \mathfrak{g} \to \mathfrak{g}/\mathfrak{q}$ the canonical map and the standard terminology of semi-simple groups (*cf.* [26]). Such subalgebras will be called minimal boundary subalgebras of $\mathfrak{g}$.

We have in particular (the proof will be left as an exercise for the reader) the following

Proposition. *All the minimal boundary subalgebras of* $\mathfrak{g}$ *are conjugate* (*under:* $\mathrm{int}(\mathfrak{g})$) *in* $\mathfrak{g}$. *The algebra* $\mathfrak{g}$ *is a* B-*algebra if the minimal boundary subalgebras are* C-*algebras* (*i.e. if their radicals are* C-*algebras, cf.* [13]).

Presented like this the B-NB classification becomes "subordinate" to the C-NC classification of the amenable algebras. "Philosophically" what the theorems of this paper say is that for non amenable groups the principal term $e^{-\lambda t}$ of the heat kernel ϕ_t comes from the "spectral gap" and that the error term $e^{\lambda t}\phi_t(g)$ is controlled by the geometry of the minimal boundary subgroups.

2. Basic geometric structure.

2.1. The Haar measure.

Let G be a locally compact group and let $Z \subset K \subset G$, $R \subset G$ be closed subgroups such that $G = RK$ and K/Z is compact and $Z \subset G$ central. Quite generally we shall denote throughout by $dh = d^\ell h$ and by $d^r h$ the left and the right Haar measures of the locally compact group H. Among the above groups K is unimodular and $dk = d^r k = d^\ell k$.

The example to keep in mind is the Iwasawa decomposition of a connected real semisimple group $S = NAK$ where $R = NA$ and Z is the discrete center of S. More generally when G is a simply connected real Lie group then we can write $G = Q \leftthreetimes S$ where Q is the radical and $S = NAK$ is a semisimple simply connected Levi subgroup. We can set $G = RK$ with $R = Q \leftthreetimes NA$. Indeed if $Z(K) \subset K$ is the discrete center of S then it is well known and easy to prove (*cf.* [15] for a proof) that there exists $Z \subset Z(K)$ of finite index (*i.e.* $[Z(K) : Z] < +\infty$) such that Z is central in G (when $\mathfrak{q}$ is the Lie algebra of Q this here amounts to saying that $\mathrm{Ad}(Z)|_{\mathfrak{q}} =$ identity). We have

Lemma. *Let G, R, K, Z be as above and assume that $R \cap K = \{e\}$. Then*
$$\int_G f(g)\, d^\ell g = \int_{R\times K} f(rk)\, d^\ell r\, dk$$
for an appropiate normalization of the Haar measures.

Remark. In the above lemma we can relax the conditions that K/Z is compact and Z is central and impose instead the unique condition that the modular function $m(g)$ on G satisfies $m(k) \equiv 1$ for all $k \in K$. One can also refer to [35, 1.5.1] or to [9], [25, Chapter 7, Section 2, number 9] for analogous and more general results.

Proof. There exists a unique $\Phi(r, k) > 0$ ($r \in R, k \in K$) such that
$$\int_G f(g)\, d^\ell g = \int_{R\times K} \Phi(r,k) f(rk)\, d^\ell r\, dk\,,$$
Φ is just the Jacobian of the mapping $R \times K \to G$ ($(r,k) \to rk$).

The uniqueness of the above Φ and the left invariance of $d^\ell g$ implies that
$$\Phi(r_1 r, k) = \Phi(r, k) = \Phi_0(k)\,, \qquad r, r_1 \in R,\ k \in K$$

with a new function $\Phi_0 > 0$ on K. If we use the right action $g \mapsto gk$ $(k \in K)$ on $d^\ell g$ we see that it sends $d^\ell g$ to $m(k)d^\ell g$ where $m(\cdot)$ is the modular function on G. By our hypothesis $m(k) = 1$ and therefore $\Phi_0(kk_1) = \Phi_0(k)$ $(k, k_1 \in K)$ and Φ_0 is a constant.

We shall introduce now a basic notation that will be used throughout this paper.

i) R will denote some locally compact group assigned with its $dr = d^\ell r$ and $d^r r$ measure. In practice R will always be a Lie group and more often than not a soluble Lie group.

ii) K will be some C^∞ manifold asigned with some C^∞ non vanishing measure dk. More generally K could be an abstract measure space.

iii) For any measure $\tilde{d}r$ on R we shall consider the measure $\tilde{d}x = \tilde{d}r \otimes dk$ on $X = R \times K$ (the product space rather than group product). We shall denote, in particular

$$dx = d^\ell x = d^\ell r \otimes dk\,, \quad d^r x = d^r r \otimes dk\,.$$

iv) More often than not I shall assume that there exists Z some discrete group acting (discretely) on K stabilizing dk and such that K/Z is a C^∞ manifold.

v) For our applications K as in iv) will be a locally compact (more often than not a Lie) group and dk will be the Haar measure, $Z \subset K$ will then be some discrete central subgroup.

vi) We shall say that we are in the "group case" $X = G$ if G is a locally compact group (more often than not a connected Lie group) and if R, K are closed subgroups such that $R \cap K = \{e\}$ and such that the conditions of the above lemma are verified. We set then $X = R \times K$ which we identify as a measure space, or even as a C^∞-manifold, with G.

The above construction admits a number of useful generalizations which although not essential for us are worth noting. For instance, in practice we can often write a connected Lie group in the form $G = RK$ where R is closed (but not necessarily connected) and K is an analytic subgroup (but not neccesarily closed) and such that $Z = R \cap K$ is a (closed) discrete central subgroup of G. The general Levi decomposition of G is of the above form. We can then identify Z to a closed central subgroup of the Lie group K (for its intrinsic analytic structure) and consider the projection $\pi : K \to K/Z$ if $\Sigma \subset K$ is some Borel section of π (*i.e.* π is (1-1) from Σ onto K/Z), we shall denote by $(\sigma) = \pi^{-1}\,\sigma \in \Sigma \subset G$, $\sigma \in K/Z$. We can then identify G with $R \times K/Z$ by the mapping

$$(r,\sigma) \longrightarrow r(\sigma) = g\,. \tag{2.1.1}$$

For the above section it is clear that

$$(\sigma_1\sigma_2) = r(\sigma_1)(\sigma_2)\,, \qquad (\sigma_1^{-1}) = \tilde{r}(\sigma_1)^{-1}\,,$$

where r, $\tilde{r} \in Z$. From these relations it immediately follows that the proof of the above lemma generalizes and that the above identification identifies $d^\ell r \otimes d_{K/Z}\sigma$ with $d^\ell g$ provided of course that $m_G(k) = 1$ $(k \in K)$.

2.2. The left invariant operators.

X will be here as in Section 2.1 and we shall examine positive R-left invariant operators on X

$$\begin{cases} T : C_0^\infty(X) \longrightarrow C^\infty(X)\,; \quad Tf \geq 0\,,\ \text{for all } f \geq 0\,, \\ T(f_r) = (Tf)_r\,;\ f_r(r_1,k) = f(rr_1,k)\,,\ r,r_1 \in R,\ k \in K\,. \end{cases} \tag{2.2.1}$$

Let $\mu_{h,k} \in M(R)$ $(h,k \in K)$ be a family of positive measure (more often than not I shall assume that they are bounded measures) and let $L(h,dk)$ be some positive "kernel" on K (*e.g.* $L(h,dk) = L(h,k)\,dk$ where $L(h,k) \geq 0$ but of course more general kernels could be considered). An invariant operator as in (2.2.1) can then be defined by the formula

$$\begin{aligned} Tf(r,h) &= \int_K L(h,dk)(f(\cdot,k) * \mu_{h,k})(r) \\ &= \int_K L(h,dk) \int_R f(rr_1^{-1},k)\,d\mu_{h,k}(r_1)\,, \qquad f \geq 0\,, \end{aligned}$$

provided that L and the μ's satisfy the appropiate smoothness conditions. If in particular $f = \varphi \otimes \psi$, $\varphi \in C_0(R)$, $\psi \in C_0(K)$ we have

$$Tf(r,h) = \int L(h,dk)(\varphi * \mu_{h,k}(r))\psi(k)\,.$$

Motivated by this we shall introduce the notation

$$(2.2.2) \qquad Lf(h) = \int_K L(h, dk) f(k)\,, \qquad f \in C_0(K)\,,$$

$$(2.2.3) \qquad T = L \otimes \{*\mu\} = L(h, dk) \otimes \{*\mu_{h,k}\}\,.$$

The representation (2.2.3) of T is clearly not unique (*e.g.* replace $L \to \alpha(h)L$; $\mu_{h,k} \to \alpha^{-1}(h)\mu_{h,k}$). We shall say that the representation (2.2.3) is normal if $\mu_{h,k} \in \mathbb{P}(R)$ is a probability measure for each $h, k \in K$. It is clear that under obvious (and reasonable) conditions a positive R-left invariant operator admits a unique normal representation as in (2.2.3). To see the uniqueness observe that for a normal representation we have

$$(2.2.4) \qquad Lf = g,\ f \in C_0^\infty(k) \quad \text{if and only if} \quad T(f \otimes \mathbf{1}) = g \otimes \mathbf{1}$$

with $\mathbf{1}(r) = 1$ $(r \in R)$. For normal representations it follows in particular from (2.2.4) that T is markovian (respectively sub-markovian) *i.e.* that $T1 = 1$ (respectively $T1 \leq 1$) if and only if L is markovian (respectively sub-markovian).

Let finally (T_j $j = 1, 2, \dots$) be a sequence of positive R-left invariant operators on X as above. We can then define the R-left invariant (time inhomogeneous in general) Markov chain ($x_n \in X$ $n = 1, 2, \dots$) by the condition that T_j $j = 2, 3, \dots$ are the transition operators

$$T_j f(x) = \int \mathbb{P}[x_j \in dy \;//\; x_{j-1} = x]\, f(y)\,.$$

2.3. The group case and the convolution operators.

We shall consider here G a locally compact group and $d\mu(g) = \varphi(g)\,dg = \psi(g)m(g)\,dg$ $(g \in G)$ some positive measure where m is the modular function. Let T be the corresponding convolution operator

$$(2.3.1) \quad Tf(g) = f * \mu(g) = \int_G f(gg_1^{-1})\, d\mu(g_1) = \int_G f(gg_1^{-1})\varphi(g_1)\, dg_1\,.$$

In this section we shall also assume that we are in the group case $X = G = RK$ as in vi) of Section 2.1, and we shall analyze the R-left invariant operator T on X.

We shall adopt the notation

$$g = rk\,,\quad g_i = r_i k_i\,,\qquad i = 1,2,\quad r, r_i \in R\,,\quad k, k_i \in K\,,$$

and by the Lemma in Section 2.1 we have

$$Tf(rk) = \iint_{R\times K} f(rkk_1^{-1}r_1^{-1})\,\varphi(r_1k_1)\,dr_1dk_1\;.$$

Let us fix $k \in K$ and consider the (1-1) correspondence $(r_1, k_1) \leftrightarrow (r_2, k_2)$ given by

(2.3.2) $$g = r_1k_1 = k_2^{-1}r_2k\,,\quad dg = dr_1dk_1 = J(r_2; k_2, k)\,dr_2dk_2\;,$$

where $J(\,\cdot\,;\,\cdot\,, k)$ is of course the Jacobian. We have thus

(2.3.3) $$Tf(rk) = \iint_{R\times K} f(rr_2^{-1}k_2)\,\varphi(k_2^{-1}r_2k)\,J(r_2; k_2, k)\,dr_2\,dk_2\;.$$

We have

Lemma. *The Jacobian $J(r; k', k) = J(r)$ is independent of $k, k' \in K$ and*

(2.3.4) $$J(r) = \frac{m_R(r)}{m_G(r)}\,,\qquad r \in R\,,$$

where $m_R(\,\cdot\,)$ is the modular function of R amd $m_G(\,\cdot\,)$ is the modular function of G.

PROOF. By the unimodularity of K we have (with obvious notations)

$$dk = dk^{-1} = d(k_0k^{\pm1}) = d(k^{\pm1}k_1)\,,\qquad k_0, k_1 \in K\;.$$

This and the definition (2.3.2) imply that $J(r; k_0k', kk_1) = J(r; k', k)$ for all $k, k', k_0, k_1 \in K$ (it is only a matter of testing $\int f(g)\,dg = \int f(k_2^{-1}r_2k)J(r_2; k_2k)\,dr_2dk_2$ on $f_h(\,\cdot\,) = f(h\,\cdot)$ and on $f^h(\,\cdot\,) = f(\cdot\,h)$). The first part of the lemma follows and (2.3.2) takes the form

(2.3.5) $$dg = dr_1\,dk_1 = J(r_2)\,dr_2\,dk_2\,,\qquad g = r_1k_1 = k_2^{-1}r_2k\,.$$

Observe now that since $d^rg = dg^{-1}$, $d^rr = dr^{-1}$, $dk = dk^{-1}$ the lemma of Section 2.1 implies that with the parametrization $g = kr$ ($k \in K$,

$r \in R$) we have $d^r g = d^r r dk$ which together with (2.3.5) allows us to conclude that

$$\begin{aligned} d^r g &= dk_2\, d^r r_2 = dk_2\, m_R(r_2)\, dr_2 \\ &= m_G(g)\, dg = m_G(g)\, J(r_2)\, dr_2\, dk_2\,, \qquad g = k_2^{-1} r_2 k \end{aligned}$$

(2.3.4) follows.

If we use the above lemma in (2.3.3) we finally obtain

$$(2.3.6)\qquad \begin{cases} Tf(rk) = \displaystyle\iint f(rr_1^{-1}k_1)\, M(r_1;k_1,k)\, dr_1 dk_1\,, \\ M(r;k_1,k) = \varphi(k_1^{-1}rk)\, \dfrac{m_R(r)}{m_G(r)} \\ \qquad\qquad = \psi(k_1^{-1}rk)\, m_R(r)\,, \quad r \in R,\ k,k_1 \in K\,, \\ Tf(rk) = \displaystyle\iint f(rr_1^{-1}k_1)\, \psi(k_1^{-1}r_1k)\, d^r r_1\, dk_1\,. \end{cases}$$

The K-bi-invariant case deserves special attention. We say that the operator (2.3.1) is K-bi-invariant if $(Tf)^k = Tf^k$ ($k \in K$, $f^k(g) = f(gk)$). Clearly this is the case if and only if the inner automorphism $I_k : G \to G$, $I_k : g \to k^{-1}gk$ stabilizes the measure μ. By our hypothesis dg is also stable by the action of I_k. Therefore it follows that $\psi(kxk^{-1}) = \psi(x)$ and we can write (2.3.6) in the form

$$(2.3.7)\qquad \begin{aligned} Tf(rk) &= \iint f(rr_1^{-1}k_1^{-1}k)\, \psi(r_1k_1)\, d^r r_1\, dk_1 \\ &= \iint f(rr_1^{-1}k_1)\, \psi(kk_1^{-1}r_1)\, d^r r_1\, dk_1\,, \end{aligned}$$

i.e. as a convolution on the product group $R \times K$.

The point of a K-bi-invariant operator in the above context is that it can be identified with an operator on the homogeneous space $G/K = \{gK : g \in G\}$. We can then identify G/K with R and since left translation by elements of R clearly commutes with the projection $G \to G/K \cong R$ the operator thus obtained on R is a convolution operator

$$(2.3.8)\qquad f \longmapsto f * \tilde{\mu}\,, \qquad f \in C_0(R)\,,$$

where $\tilde{\mu} \in M(R)$. From (2.3.7) we see that we have in fact $d\tilde{\mu}(r) = \tilde{\varphi}(r)\, d^r r$ with

$$\tilde{\varphi}(r) = \int_K \varphi(rk)\, dk\,, \qquad r \in R\,. \tag{2.3.9}$$

The above two formulas (2.3.8) and (2.3.9) are not used in this paper but have the merit of putting the above notions in the correct "perspective" and are relevant in the "semisimple theory" which will be developed elsewhere.

2.4. The composition, the adjoint, $\| \cdot \|_{p\to p}$ norms, amenability and the "local" estimate.

Let us consider

$$T_i = L_i(h, dk) \otimes \{*\mu_{h,k}^{(i)}\}\,, \qquad i = 1, 2, \dots\,,$$

a sequence of R-left invariant operators on X as in Section 2.2. It is clear then that

$$\begin{aligned} T_1 \circ \cdots \circ T_n = \int_{k_1\in K} \cdots \int_{k_{n-1}\in K} & L_1(h, dk_1) \cdots L_n(k_{n-1}, dk) \\ & \otimes \{*(\mu_{h,k_1}^{(1)} * \cdots * \mu_{k_{n-1},k}^{(n)})\} \end{aligned} \tag{2.4.1}$$

with obvious notations. To simply notations let $L(h, dk) = L(h, k)\, dk$ and let

$$T = L(h, k)\, dk \otimes \{*\mu_{h,k}\} \tag{2.4.2}$$

as in (2.2.3, Section 2.2). Let then T^* be the formal adjoint operator with respect to $d^r x = d^r r \otimes dk$ then clearly

$$T^* = L^*(h, k)\, dk \otimes \{*\mu_{h,k}^*\}\,,$$

where $L^*(h, k) = L(k, h)$ and $\mu_{h,k}^* = \check{\mu}_{k,h}$ where for any measure ν on R we adopt throughout the notation $d\check{\nu}(g) = d\nu(g^{-1})$ (*i.e.* $\check{\nu}$ is the image of ν under the mapping $g \mapsto g^{-1}$ on R). This follows trivially from the fact that the formal adjoint with respect to $d^r r$ of the operator $f \mapsto f * \nu$, $f \in C_0(R)$, $\nu \in M(R)$ is $f \mapsto f * \check{\nu}$. If the representation

(2.4.2) is normal it follows that the operator T is self adjoint with respect to $d^r x = d^r r \otimes dk$ if and only if

$$L(h,k) = L(k,h)\,, \qquad \check{\mu}_{h,k} = \mu_{k,h}\ . \tag{2.4.3}$$

Let now $\tilde{dx} = \tilde{dr} \otimes dk$ be some measure on X defined as in Section 2.2, it is then clear from (2.4.1) that the $L^p(X;\tilde{dx}) \to L^q(X;\tilde{dx})$ norm can be controlled by the $*\mu_{k_1,k_2} * \cdots \mu_{k_{n-1},k_n} : L^p(R;\tilde{dr}) \to L^q(R;\tilde{dr})$ convolution norm, *i.e.*

$$\|T_1 \circ \cdots \circ T_n\|_{p\to q} \le \|L_1 \circ \cdots \circ L_n\|_{p\to q} \sup_{k_1\cdots k_n} \|*\mu_{k_1,k_2} * \cdots * \mu_{k_{n-1},k_n}\|_{p\to q}\ ,$$

where $L_1,\dots,L_n$; $L_1\circ\cdots\circ L_n$ are the corresponding operators on K as in Section 2.2. If the representations (2.4.2) are all normal and if $\tilde{dr} = d^r r$, $\tilde{dx} = d^r x$ this means that

$$\begin{aligned} \|T_1 \circ \cdots \circ T_n\|_{p\to p} &\le \|L_1 \circ \cdots \circ L_n\|_{p\to p}\ , \\ \|T^n\|_{p\to p} &\le \|L^n\|_{p\to p}\ , \qquad 1 \le p \le +\infty\,, \end{aligned}$$

we also have the following basic

Lemma. *If R is amenable, the above inequality is actually an equality i.e.*

$$\|T_1 \circ \cdots \circ T_n\|_{p\to p} = \|L_1 \circ \cdots \circ L_n\|_{p\to p}\ , \qquad 1 \le p \le +\infty\,.$$

PROOF. Since "everything is positive" it suffices to show that there exist $0 \le f_m, g_m \in C_0^\infty(R)$ $(m \ge 1)$ such that

$$\|f_m\|_p \le 1\,, \qquad \|g_m\|_q \le 1\,, \qquad m \ge 1\,, \qquad \frac{1}{p} + \frac{1}{q} = 1\,, \tag{2.4.4}$$

$$\langle f_m * \mu_{k_1,k_2} * \cdots * \mu_{k_{n-1},k_n}, g_m\rangle_{L^2(R;d^r r)} \underset{m\to\infty}{\longrightarrow} 1 \tag{2.4.5}$$

and that the limit in (2.4.5) uniform when $k_j \in C \subset\subset K$ $(j = 1,\dots,n)$ and C is a compact subset. In fact, to avoid unnecessary complications in this proof we shall make an additional hypothesis that will always be verified for us: We shall assume that for all $\varepsilon > 0$ exists $C \subset R$ such that $\mu_{h,k}(R\sim C) < \varepsilon$ $(h, k \in K)$. Otherwise the next few lignes have to be "handled with care"!

If we denote by $\pi = \mu^{(1)}_{k_1,k_2} * \cdots * \mu^{(n)}_{k_{n-1},k_n}$ we see that

$$\begin{aligned}\langle f * \pi, g\rangle_{L^2(R;d^r r)} &= \int g(x)\, d^r x \int f(xy^{-1})\, d\pi(y)\\ &= \int \left(\int f(xy^{-1})\, g(x)\, d^r x\right) d\pi(y)\\ &= \int \left(\int \tilde{f}(yx^{-1})\, g(x)\, d^r x\right) d\pi(y) = \langle \tilde{f} \overset{*}{r} g, \pi\rangle\,,\end{aligned}$$

where $\tilde{f}(x) = f(x^{-1})$ and where the defintion of the right convolution $\overset{*}{r}$ of two functions is "given" by the last equality. (2.4.4), (2.4.5) will therefore follow if we can choose $f_m, g_m \in C_0^\infty$ that satisfy (2.4.4) and such that

$$\tilde{f}_m \overset{*}{r} g_m \underset{m\to\infty}{\longrightarrow} 1 \qquad \text{uniformly on compacta}\,. \tag{2.4.6}$$

The well informed reader recognizes here one of the many consequences and definitions of the amenability (*cf.* [16] where it is proved in its dual form $\|f\|_p, \|g\|_q \leq 1$ where the $\|\ \|$ are taken with respect to the left measure dr, and $f_m * \tilde{g}_m \to 1$. Observe also that one way to avoid the uniformity hypothesis on the measures $\mu_{h,k}$ is to impose some kind of monotonicity on the limit (2.4.6)).

A very important conclusion can be drawn from the above consideration. Let us start from the assumption that for some $0 \leq \theta(n) \underset{n\to\infty}{\to} 0$ and for every fixed $\varphi, \psi \in C_0^\infty(R)$ we have

$$\sup_{k_1,\ldots,k_n} \langle \varphi * \mu^{(1)}_{k_1,k_2} * \cdots * \mu^{(n)}_{k_{n-1},k_n}, \psi\rangle_{L^2(R;d^r r)} = O(\theta(n))\,.$$

It then follows that for fixed $F = \varphi_1 \otimes \varphi_2$, $\Psi = \psi_1 \otimes \psi_2$, $\varphi_1, \psi_1 \in C_0^\infty(R)$, $\varphi_2, \psi_2 \in C_0^\infty(K)$ we have

$$\langle T^n F, \Psi\rangle = O(\theta(n)\, \|L^n\|_{2\to 2})\,,$$

where for simplicity we assume that $T_1 = T_2 = \cdots = T$. It follows that if R is amenable we have the local estimate

$$\langle T^n F, \Psi\rangle = O(\theta(n)\, \|T^n\|_{2\to 2})\,. \tag{2.4.7}$$

One final remark is in order. Let $\alpha(r) > 0$ be an arbitrary continuous positive function and $T_\alpha = \alpha^{1/2} T \alpha^{-1/2}$ the corresponding conjugated operator. It is clear that $T_\alpha^n = \alpha^{1/2} T^n \alpha^{-1/2}$ and that the $2 \to 2$ norm of T_α with respect to $\tilde{d}x = \tilde{d}r \otimes dk$ is the same as the $2 \to 2$ norm of T with respect to $\alpha^{-1}\tilde{d}x = (\alpha^{-1}\tilde{d}r) \otimes dk$. The local estimate is on the other hand invariant by that conjugation since the α is absorbed in the compactly supported F and Ψ. In other words for arbitrary α as above we have

$$\langle T^n F, \Psi \rangle = O(\theta(n) \, \|T^n\|_{L^2(\alpha d^r x) \to L^2(\alpha d^r x)})$$

and in particular, with $\alpha = m_R =$ the modular function of R, we have

$$\langle T^n F, \Psi \rangle = O(\theta(n) \, \|T^n\|_{L^2(dx) \to L^2(dx)}) \, .$$

The proof of the upper estimate of our main theorem hindges on this observation.

Let us now suppose that the density $L(h,k)$ of the operator L is continuous and strictly positive and that the operator $L : L^2(K) \to L^2(K)$ is compact. In the above estimate we can then replace $\|T\|_{2\to 2}$ by $\|T\|_{\mathrm{sp}}$ the spectral radius of T (since $\|A\|_{\mathrm{sp}} = \lim \|A^n\|^{1/n}$ we clearly have $\|T\|_{\mathrm{sp}} = \|L\|_{\mathrm{sp}}$).

The reason why we can do this is because the operator L admits then $0 < \varphi_0 \in L^2(K)$ a positive eigenfunction (*cf.* [31]) and in the previous argument we can set $\varphi_2 = \psi_2 = \varphi_0$.

Indeed assume for simplicity that K is compact then any eigenfunction of L is continuous and if $\varphi \in C(K)$ is such an eigenfunction with maximal (in modulus) eigenvalue then

$$L|\varphi| \geq |L\varphi| \, ; \quad L|\varphi| \not> (1+\varepsilon)|L\varphi| \qquad \text{for all } \varepsilon > 0 \, ,$$

i.e. the inequality $L|\varphi| > (1+\varepsilon)|L\varphi|$ does not hold for any $\varepsilon > 0$ (indeed if we assume, as we may, that the eigenvalue in question has modulus 1, such an inequality would give $\|L^n\| \geq (1+\varepsilon)^n$ which contradicts the fact that $\|L\|_{\mathrm{sp}} = 1$).

Therefore there exists $k_0 \in K$ such that $L|\varphi|(k_0) = |L\varphi(k_0)|$. But this (because $L > 0$) implies that $\varphi = e^{i\theta}|\varphi|$ (for some fixed $\theta \in \mathbb{R}$). We have therefore $L|\varphi| = \|L\|_{\mathrm{sp}}|\varphi|$ and from this it follows that $\varphi_0 = |\varphi| > 0$.

2.5. Semigroup of operators.

In this section I shall examine

$$T_t = L_t(h, dk) \otimes \{*\mu_{h,k}^{(t)}\}, \qquad t > 0, \tag{2.5.1}$$

a semigroup of positive R-left invariant operators on X, *i.e.* we assume that $T_t \circ T_s = T_{t+s}$. We shall assume that the representation (2.5.1) is normal, by (2.2.4) it then follows that $L_t \circ L_s = L_{t+s}$ (*cf.* Section 2.2) is also a semigroup and that if T_t is symmetric with respect to $d^r r \otimes dk$ then $T_t \gg 0$, *i.e.* is a positive Hilbert space operator with respect to that measure. Clearly also T_t is (sub)markovian if and only if L_t is (sub)markovian.

Example. $T_t = e^{-tA}$ where A is a R-left invariant differential operator on X. To write down A we can fix once and for all left invariant fields on R, Y_1, $Y_2, \dots$ and local coordinates $(k_1, k_2, \dots)$ on K. It follows that Y_j and $\partial/\partial k_i$ commute and that we can write

$$A = \sum a_{ij} Z_i Z_j + \sum a_i Z_i + a, \tag{2.5.2}$$

where each Z_i is either one of the Y_j's or one of the $\partial/\partial k_j$'s and furthermore each coefficient a_{ij}, a_i, a is independent of $r \in R$ (but may depend on $k \in K$). The "projected" operator B on K is then obtained by retaining only the terms of (2.5.2) for which no Y_j field appears and we have $L_t = e^{-tB}$. Observe that in the group case T_t is a "K-bi-invariant" semigroup if A is a K-bi-invariant and that this implies that B is K-right invariant on K. If we are in a group case $X = G$ we can, for instance, take

$$A = \sum b_{ij} X_i X_j + X_0 + b, \tag{2.5.3}$$

where the b's are constant with $(b_{ij}) \gg 0$ and $X_1, X_2, \dots$ are left invariant vector fields on G. Such an operator can clearly be rewritten

$$\Delta = \sum_{i=1}^{n} X_i^2 + X_0 + b \tag{2.5.4}$$

as in Section 0 (for a different, of course, choice of invariant fields $X_0, X_1, \dots$). When $b = 0$, Δ is a markovian generator on G and is formally selfadjoint with respect to $d^r g$ if $X_0 = 0$. But even if $X_0 = 0$, Δ is not in general formally selfadjoint with respect to $d^r r \otimes dk$ and therefore the corresponding operators e^{-tB} are markovian but not symmetric. There is another important property that B inherits from A. First of all if A is elliptic (*i.e.* if the matrix (a_{ij}) is positive definite) the operator B is also elliptic. Let us assume more generally that the fields $X_1, \dots, X_n$ in (2.5.4) generate the Lie algebra of G. The projected operator B can then be written

$$B = \sum_{i=1}^{n} \tilde{X}_j^2 + \tilde{X}_0 + \tilde{b}\,, \tag{2.5.5}$$

where $\tilde{X}_j$ is the corresponding projected field on K. Of course, even if we are in the group case, $\tilde{X}_j$ need not be in general a K-invariant field in any sense whatsoever, but it is certainly true that B is on K a Hörmander operator in the sense that at every point $k \in K$ the fields $\tilde{X}_1, \dots, \tilde{X}_n$ span together with all their successive brackets the tangent space.

Let us now go back to the general semigroup and let us assume that T_t is symmetric with respect to $d^r x$ and that therefore L_t is symmetric with respect to dk (both T_t and L_t are therefore positive operators in the Hilbert space sense). I shall further make the following assumption: A: (respectively A'): there exists $\varphi_0 > 0$, $\lambda_0 > 0$ such that

$$L_t \varphi_0 = e^{-\lambda_0 t} \varphi_0 \qquad (\text{respectively } L_t \varphi_0 \le e^{-\lambda_0 t} \varphi_0)\,.$$

We shall presently elaborate on that condition but first we shall draw the consequences of A and A'. Under the above conditions we shall consider the semigroups

$$\hat{T}_t = e^{\lambda_0 t} \varphi_0^{-1} T_t \varphi_0\,, \qquad \hat{L}_t = e^{\lambda_0 t} \varphi_0^{-1} L_t \varphi_0\,, \tag{2.5.6}$$

where in (2.5.6) φ_0 is identified with an R-left invariant function on X which satisfies $T_t \varphi_0 = e^{-\lambda_0 t} \varphi_0$ (respectively $T_t \varphi_0 \le e^{-\lambda_0 t} \varphi_0$) on X. The semigroups (2.5.6) are therefore markovian (respectively submarkovian) $\hat{L}_t$ is symmetric with respect to the measure $\hat{dk} = \varphi_0^2 dk$ and $\hat{T}_t$ is symmetric with respect to the measure $d^r r \otimes \hat{dk}$.

Let us now go back to the assumption A (and A') and give natural examples under which it is verified. Let us first suppose that we are in the group case that K is compact and that $T_t = e^{-t\Delta}$ with Δ as in

(2.5.4). To simplify matters let us also assume that L_t is symmetric with respect to dk (dk here is some smooth non vanishing measure that need not be the Haar measure of K). The Hörmander condition on (2.5.4) implies that the operators L_t are in the trace-class on $L^2(K; dk)$. This is because the kernel $L_t(h,k)$ is C^∞ and thus Hilbert-Schmidt (and $L_{t/2} \circ L_{t/2} = L_t$). We have therefore

$$L_t = \sum_{j=0}^{\infty} e^{-\lambda_j t} \varphi_j(h) \varphi_j(k) \,,$$

where $\lambda_0 \leq \lambda_1 \leq \cdots$ with $\sum_{j=0}^{\infty} e^{-\lambda_j t} < \infty$ $(t > 0)$, and $\varphi_j \in C^\infty_{\mathbb{R}}(K)$, $\|\varphi_j\|_2 = 1$, $j = 0, 1, 2, \ldots$.

By the positivity of the operators involved we have $L_t|\varphi_0| \geq |L_t\varphi_0| = e^{-\lambda_0 t}|\varphi_0|$. Also, since $\|L_t\|_{2\to 2} = e^{-\lambda_0 t}$, we have $\|L_t|\varphi_0|\|_2 \leq e^{-\lambda_0 t}\|\varphi_0\|_2$ and therefore $L_t|\varphi_0| = e^{-\lambda_0 t}|\varphi_0|$. It follows that we can renumber the eigenfunctions $\varphi_0, \varphi_1, \ldots$ in such a way that $0 \leq \varphi_0 \in C^\infty(K)$.

The next step is to show that φ_0 never vanishes $\varphi_0 > 0$ $(k \in K)$ and that therefore the condition (A) is verified. This of course is an immediate consequence of the eigenvalue property

$$L_t\varphi_0(h) = \int L_t(h,k)\varphi_0(k)\, dk = e^{-\lambda_0 t}\varphi_0(h)$$

and of the more general fact that for any non identically zero $0 \leq \varphi \in C^\infty(K)$ we have

$$L_t\varphi(h) > 0 \,, \qquad h \in K \,. \tag{2.5.7}$$

To see this we observe that $L_t = e^{-t(B-\alpha)}e^{-t\alpha}$ for any $\alpha > 0$ and B as in (2.5.5). $B - \alpha$, on the other hand, for $\alpha \geq 0$ large, generates a "hypoelliptic" diffusion. This means that the kernel $e^{t\alpha}L_t(h,k)$ never vanishes for $t > 0$ (*cf.* [4]), (2.5.7) follows.

The above situation can be generalized as follows. We shall drop the assumption that K is compact but assume that there exists Γ some discrete group that acts discontinuously on K and in such a way that K/Γ is compact. We shall also assume that T_t the semigroup (2.5.1) is stable by the natural Γ-action and induces thus a corresponding R-left invariant semigroup on $X/\Gamma = R \times K/\Gamma$. If we assume that the corresponding semigroup L_t on K/Γ has all the above properties (so that the existence of φ_0 on K/Γ with the required properties is guaranteed) then we can define the analogous φ_0 on K by taking on K the corresponding Γ-automorphic function. We see in particular that the assumption A is verified in our group case vi) of Section 2.1.

2.6. The symmetric Laplacian in the group case.

Let us suppose that we are in the group case $G = X = R \times K$ as in Section 2.1 vi) and that $\Delta = -\sum X_j^2$, $T_t = e^{-t\Delta}$ is as in Section 0.1. The modified semigroup $\tilde{T}_t = m_G^{1/2} T_t m_G^{-1/2}$ is then symmetric with respect to $d^\ell g = d^\ell r \otimes dk$ and therefore

$$\check{T}_t = m_R^{-1/2}\, \tilde{T}_t\, m_R^{1/2} = (m_G/m_R)^{1/2} e^{-t\Delta} (m_G/m_R)^{-1/2} \tag{2.6.1}$$

is symmetric on X with respect to $d^r x = d^r r \otimes dk$ as was needed for the considerations of Section 2.5 to go through. Observe that when $K = \{e\}$, $m_R \equiv m_G$ and $\check{T}_t = T_t$.

In this section I shall make a number of explicit computations related to the above semigroup. Let G be a simply connected Lie group and let $G = Q \rtimes S$, $S = NAK$, $R = QNA$, $Z \subset K$ have the same meaning as in Section 2.1 so as to have the identification $G = R \times K$. It is clear that $m_G(k) \equiv 1$ ($k \in K$) (and more generally $m_G|_S = 1$) so it suffices to analyse m_R and $m_G|_R$. Since Q is a normal subgroup of G we have $m_G|_Q = m_R|_Q = m_Q$ so that $m_G/m_R|_Q \equiv 1$. Since S is semisimple and $m_G|_S \equiv 1$ it follows that

$$(m_G/m_R)^{1/2}(x) = m_R^{-1/2}(x)\,, \qquad x \in AN\,.$$

Now since all the automorphisms induced on Q by inner automorphisms by elements of S are unimodular (S being semisimple) we have (*cf.* [25])

$$m_R(x) = m_{AN}(x)\,, \qquad x \in AN\,.$$

So with the obvious abuse of notation we have

$$(m_G/m_R)^{1/2} = m_{AN}^{-1/2}\,,$$

where $G = (Q \rtimes AN)K$. For the semisimple group S the quantity m_{AN} is a very familiar creature $m_{AN} = e^{2\rho}$ where $\rho = (\sum \rho_j)/2$ is the 1/2-sum of the roots (*cf.* [9]). In particular it only depends on the

A-coordinate. The "moral" is that the conjugating factor in (2.6.1) is an "old friend".

To obtain the symmetric markovian semigroup $\hat{T}_t$ on G then we define as in Section 2.5 the Z-automorphic function on K, φ_0 and λ_0 the corresponding eigenvalue which is then given by $e^{-\lambda_0 t} = \|\check{T}_t\|_{2\to 2}$ on $L^2(X; d^r x)$ (*cf.* Section 2.4). But clearly also $e^{-\lambda_0 t} = \|\tilde{T}_t\|_{2\to 2}$ on $L^2(G; dg)$ and $\lambda_0 = \lambda$ is just the spectral gap (*cf.* Section 0.1) of Δ. The semigroup

$$(2.6.2)\quad \hat{T}_t = e^{\lambda t}\varphi_0^{-1}\check{T}_t\varphi_0 = e^{\lambda t}\varphi_0^{-1}(m_G/m_R)^{1/2}e^{-t\Delta}(m_G/m_R)^{-1/2}\varphi_0$$

is thus markovian and symmetric with respect to $d^r r\otimes\varphi_0^2 dk = d^r r\otimes\hat{d}k$.

Let now $\hat{\phi}_t(x_1,x_2)$ be the kernel of the semigroup $\hat{T}_t$ with respect to $dr\otimes\hat{d}k$ we shall show then that we have

$$(2.6.3)\qquad \begin{aligned}\hat{\phi}_{2t}(e,e) &\approx e^{2\lambda t}\int_G \phi_t(g)\,\phi_t(g^{-1})\,dg\\ &= e^{2\lambda t}\int \phi_t(k^{-1}r^{-1})\,\phi_t(rk)\,dr\,dk\,,\end{aligned}$$

where $\phi_t(y^{-1}x)$ is the convolution kernel of $T_t = e^{-t\Delta}$ with respect to dg. Indeed quite generally if $k_t(\omega_1,\omega_2)$ is the kernel with respect to $d\omega$ of a general semigroup K_t on $L^2(\Omega; d\omega)$ where $(\Omega; d\omega)$ is some measure space then $k_t^{(\alpha)}(\omega_1,\omega_2)$ the kernel of the conjugated semigroup $\alpha K_t\alpha^{-1}$ with respect to $d\omega$ (where $\alpha(\omega)\neq 0$ is some non zero function) is $k_t^{(\alpha)}(\omega_1,\omega_2) = k_t(\omega_1,\omega_2)(\alpha(\omega_1)/\alpha(\omega_2))$. This in particular implies

$$k_t^{(\alpha)}(\omega_1,\omega_2)\,k_t^{(\alpha)}(\omega_2,\omega_1) = k_t(\omega_1,\omega_2)\,k_t(\omega_2,\omega_1)\,,\qquad t>0\,.$$

Similarly the kernel of K_t with respect to a new measure $\beta(\omega)\,d\omega$ is

$$k_t^{\langle\beta\rangle}(\omega_1,\omega_2) = \frac{1}{\beta(\omega_2)}\,k_t(\omega_1,\omega_2)\,.$$

If we apply these observations in our context where $\beta = \varphi_0^2$ is bounded from above and below we deduce that

$$(2.6.4)\quad C^{-1}\phi_t(x)\,\phi_t(x^{-1}) \le e^{-2\lambda t}\hat{\phi}_t(e,x)\,\hat{\phi}_t(x,e) \le C\,\phi_t(x)\,\phi_t(x^{-1})\,.$$

(2.6.3) follows then from

$$\hat{\phi}_{2t}(e,e) = \int_X \hat{\phi}_t(e,x)\,\hat{\phi}_t(x,e)\,dr\,\hat{d}k\,.$$

Observe now that the symmetry of $\hat{T}_t$ with respect to $d^r r \otimes \hat{d}k$ implies that

$$\hat{\phi}_t(x, x_1)\, m_R^{-1}(r_1) = \hat{\phi}_t(x_1, x)\, m_R^{-1}(r)\,, \qquad x = (r, k),\ x_1 = (r_1, k_1)\,,$$

and therefore also

$$P(t) = \int \hat{\phi}_t(e, x)\, \hat{\phi}_t(x, e)\, d^r r\, \hat{d}k = \int \hat{\phi}_t^2(e, x)\, dr\, d\hat{k}\,. \tag{2.6.5}$$

From (2.6.4) it follows also that

$$C^{-1} P(t) \leq e^{2\lambda t} \int \phi_t(rk)\, \phi_t(k^{-1} r^{-1})\, d^r r\, \hat{d}k \leq C\, P(t)\,. \tag{2.6.6}$$

We shall now show that for large $t \gg 1$ both $\hat{\phi}_{2t}(e, e)$ and $P(t)$ are "comparable" with the quantity

$$Q(t) = e^{2\lambda t} \int \phi_t(r)\, \phi_t(r^{-1})\, dr = e^{2\lambda t} \int \phi_t(r)\, \phi_t(r^{-1})\, d^r r \tag{2.6.7}$$

in the sense that

Lemma. *If K is compact, there exists $C > 0$ such that*

$$C^{-1} Q(t-1) \leq \hat{\phi}_{2t}(e, e) \leq C\, Q(t+1)\,, \qquad t \geq 10\,,$$

$$C^{-1} Q(t-1) \leq P(t) \leq C\, Q(t+1)\,, \qquad t \geq 10\,.$$

It follows in particular that

$$\begin{aligned} C^{-1} \int \hat{\phi}_{t-2}^2(e, x)\, dr\, \hat{d}k &\leq \hat{\phi}_{2t}(e, e) \\ &\leq C \int \hat{\phi}_{t+2}^2(e, x)\, dr\, d\hat{k}\,, \qquad t \geq 20\,. \end{aligned} \tag{2.6.8}$$

PROOF. If K is compact by the standard local Harnack estimate (*cf.* [1], [4]) it follows that

$$C^{-1} \phi_{t-1/2}(g k_2) \leq \phi_t(g) \leq C\, \phi_{t+1/2}(g k_1)\,, \qquad t > 1,\ g \in G,\ k_1, k_2 \in K,$$

where $C > 0$ is independent of t, g and k_1, k_2.

Combining these with the fact that

$$\phi_t(g^{-1}) = \phi_t(g)\, m_G(g)\,, \qquad g \in G\,.$$

We deduce that

$$C^{-1}\phi_{t-1}(k_3 g k_4) \le \phi_t(g) \le C\,\phi_{t+1}(k_1 g k_2)\,, \tag{2.6.9}$$

for $t > 10$, $g \in G$, $k_i \in K$, $1 \le i \le 4$. If K is compact the integrals in both (2.6.3) and (2.6.6) are comparable (in the above sense) with $\int \phi_t(r)\phi_t(r^{-1})\,dr$ and our lemma follows. If K is not compact we shall choose K_0 some relatively compact fundamental domain of the covering map $K \to K/Z$ so that

$$K = \bigcup_{z\in Z} zK_0\,, \quad z_1K_0 \cap z_2K_0 = \varnothing\,, \qquad z_1, z_2 \in Z,\ z_1 \ne z_2\,.$$

What replaces (2.6.9) is then the estimate

$$C^{-1}\phi_{t-1}(zk_3 g k_1) \le \phi_t(zg) \le C\,\phi_{t+1}(zk_1 g k_2)\,, \qquad t > 20\,,$$

where $z \in Z$ is central in G. The above argument therefore works provided that in (2.6.7) we now set

$$\begin{aligned} Q(t) &= e^{2\lambda t}\int_R \sum_{z\in Z} \phi_t(zr)\,\phi_t(z^{-1}r^{-1})\,dr \\ &= e^{2\lambda t}\int_R \sum_{z\in Z} \phi_t(zr)\,\phi_t(z^{-1}r^{-1})\,d^r r\,. \end{aligned}$$

We conclude therefore that (2.6.8) is valid in full generality.

In all the above considerations we used the measure $\hat{dk} = \varphi_0^2 dk$ on K and the corresponding measure $\hat{dk} = d^r r \otimes \hat{dk}$ on X with respect to which the semigroup $\hat{T}_t$ in (2.6.2) is symmetric. It turns out that if we invoque a result of J. Moser [17] we can in fact replace $\hat{dk}$ by the Haar measure dk. J. Moser's result says that when K is compact and orientable there exists a diffeomorphism $\alpha : K \to K$ that takes the measure $\hat{dk}$ to dk. If we use this diffeomorphism and conjugate $\hat{T}_t$ with $\alpha =$ identity$\otimes\alpha$ on X *i.e.* $(f \circ \alpha) \mapsto (\hat{T}_t f) \circ \alpha$ (for all $f \in C_0^\infty(X)$) we obtain a new semigroup that I shall still denote by $\hat{T}_t$ which is markovian and symmetric with respect to $d^r r \otimes dk$. The same thing of course holds in the general case (*i.e.* K is not compact) provided that we can lift the diffeomorphism from the compact manifold K/Z on K. This diffeomorphism lifts automatically when K is simply connected.

We shall finish this section with a probabilistic interpretation of the lower estimates in the Theorem A. Towards that we shall consider $\Omega = \{x(t) \in X; t > 0\}$ the path space of the diffusion on X generated by the semigroup $\hat{T}_t$. In other words

$$\hat{T}_t f(x) = \int \mathbb{P}_x[x(t) \in dy]\, f(y)\,, \qquad f \in C_0^\infty(X)\,,$$

$$\mathbb{P}_e[x(t) \in A] = \int_{x=(r,k)\in A} \hat{\phi}_t(e,x)\, dr \otimes \hat{d}k\,.$$

If we bare in mind that $c^{-1}dk \le \hat{d}k \le c\, dk$ and combine this with our main estimate (2.6.8) we see that

$$\mathbb{P}[x(t) \in A] \le C\,(\hat{\phi}_t(e,e))^{1/2}(dx\text{-measure }(A))^{1/2}\,. \tag{2.6.10}$$

The $dx = dr \otimes dk$ measure of $A \subset G = R \times K$ is of course the left Haar measure on G. If we use however the involution $* : (r,k) \to (r^{-1},k)$ we see from the symmetry of $\hat{T}_t$ with respect to $d^r r \otimes \hat{d}k$ that if A is of the form $A = B \times K$ $(B \subset R)$ then

$$\mathbb{P}[x(t) \in A] = \mathbb{P}[x(t) \in A^*]$$

and since $*$ interchanges the two measures $d^r r \otimes \hat{d}k$ and $dr \otimes \hat{d}k$ we see finally that in (2.6.10), if we so wish, we can replace the dx-measure by any of the measures $d^r r \otimes dk$, $d^r r \otimes \hat{d}k$, $d^\ell r \otimes \hat{d}k$. The estimate (2.6.10) allows us to formulate the following criterion.

Criterion. *Let us assume that for all $n = 1,2,\dots$ we can find a set $X_n = B_n \times K \subset X$ $(B_n \subset R)$ such that*

i) measure $(X_n) \le C\, n^C$, $\qquad n = 1,2,\dots$

ii) $\mathbb{P}[x(n) \in X_n] \ge C^{-1} n^{-C}$, $\qquad n = 1,2,\dots$

where $C > 0$ and "measure" *stands for any of the above measures. Then there exists $C > 0$ such that ϕ_t, the convolution kernel of $e^{-t\Delta}$, satisfies*

$$\phi_t(e) \ge C^{-1} t^{-C} e^{-\lambda t}\,, \qquad t \ge 1\,.$$

The standard local Harnack principle (*cf.* [1]) has to be used of course here to fill in the gaps between the integer values $t = 1, 2, \dots$.

We shall also need (for the lower estimate in Theorem A_1) a modified version of the above criterion: If X_n is as above but instead of i) and ii) we can only assert that

i') measure $(X_n) \leq C\, e^{cn^{1/3}}$

ii') $\mathbb{P}[x(n) \in X_n] \geq c^{-1} e^{-cn^{1/3}}, \qquad n = 1, 2, \dots$

Then we can conclude instead that

$$\phi_t(e) \geq C^{-1} e^{-ct^{1/3}} e^{-\lambda t}\,.$$

2.7. The projection of the infinitesimal generator.

In this section I shall preserve all our previous notations and assume that $N \subset R$ is some closed normal subgroup. We can define then $\pi : X = R \times K \to X/N = R/N \times K$ the quotient spaces by the induced left action by N and if $T = L \otimes \{*\mu_{h,k}\}$ is a positive left invariant on X as in Section 2.1 the above projection induces $T_{X/N} = L \otimes \{*\check{\pi}(\mu)_{h,k}\}$ a positive left invariant operator on X/N ($\check{\pi}(\mu)$ denotes here the image of the measure μ by π).

It is clear then that if T is self adjoint with respect to the measure $d^r r \otimes dk$ then $T_{X/N}$ is self adjoint with respect to $d^r_{R/N} r \otimes dk$ (We can use the criterion (2.4.3) to see this).

We shall now give an important example of the above situation. We shall assume that R is a simply connected soluble Lie group and that N is the nilradical so that $R/N \cong \mathbb{R}^n$. The right measure on R/N is then the Lebesgue measure dx. We shall further assume that we are in the group case and that the left invariant operators considered are the $\hat{T}_t$ defined in Section 2.6 which will be self adjoint with respect to $d^r r \otimes dk$ where dk is now assumed to be the Haar measure on K (*cf.* end of Section 2.6). We clearly have $\hat{T}_t = e^{-t\hat{A}}$ where $\hat{A}$ is a sum of squares (with drift) operator that satisfies the Hörmander condition. We shall project as explained above and obtain $\hat{T}_t = e^{-tD}$ a symmetric (with respect to $dx \otimes dk$) markovian semigroup on $X/N = \mathbb{R}^n \times K$ and we shall analyze more closely D the generator that is a subelliptic differential operator.

Let $x_1, \dots, x_n$ be the standard coordinates on $\mathbb{R}^n$ and let $X_1, \dots, X_s$ be a basis of right invariant fields on K. It is then clear by the $\mathbb{R}^n$-left invariance that

$$-D = D_R + M + P + D_K \,, \tag{2.7.1}$$

where

$$D_R = \sum_{i,j=1}^{n} a_{ij}(k) \frac{\partial^2}{\partial x_i \partial x_j} \,, \tag{2.7.2}$$

where $(a_{ij}(k))$ is a symmetric non negative matrix, $k \in K$.

$$M = 2 \sum_{i=1}^{n} \sum_{\alpha=1}^{s} b_{\alpha,i}(k) X_\alpha \frac{\partial}{\partial x_i} \,, \qquad P = \sum_{i=1}^{n} \delta_i(k) \frac{\partial}{\partial x_i} \,, \tag{2.7.3}$$

and where D_K can be identified with the canonical "projected operator" on K. That operator is self adjoint subelliptic and can thus be written in the form

$$D_K = \sum_{\alpha,\beta=1}^{s} X_\alpha \, \gamma_{\alpha,\beta}(k) \, X_\beta \,, \tag{2.7.4}$$

where $(\gamma_{\alpha,\beta}(k))$ is a symmetric non negative matrix. The constant term is zero because D is a markovian generator. What is also clear is that D_R in (2.7.2) is uniformily elliptic on $\mathbb{R}^n$ *i.e.* that

$$(a_{ij}(k)) \geq \varepsilon_0 I \tag{2.7.5}$$

for some $\varepsilon_0 > 0$ provided that the original operator Δ on G and therefore D on $\mathbb{R}^n \times K$ is actually elliptic.

The formal self adjointness of D with respect to $dx \otimes dk$ implies that

$$\delta_i(k) = \sum_{\alpha=1}^{s} X_\alpha b_{\alpha,i}(k) \,, \qquad i = 1, 2, \dots, n \,,$$

and therefore that

$$\int_K \delta_i(k) \, dk = 0 \,. \tag{2.7.6}$$

This is equivalent to the "formal" statement $\langle D(x_i \otimes 1), 1\rangle = 0$.

Of course D is K-bi-invariant if and only if all its coefficients are constant. This is the only reason why we choose the fields $X_1, \dots, X_s$ to be right invariant rather than left invariant. If we use the canonical projection $\mathbb{R}^n \times K \to \mathbb{R}^n$ and project D we obtain then D_R on $\mathbb{R}^n$. It follows in particular that then D_R is elliptic as soon as D is subelliptic.

Let us finally examine the convolution kernel. Let us go back to the original semigroup $e^{-t\Delta} f = f * \mu_t$ with $d\mu_t(g) = \phi_t(g)\, d^r g$ then the corresponding left invariant operator on $R \times K$ is (*cf.* (2.3.6))

$$\begin{aligned} T_t f(rk) &= \iint f(rr_1^{-1}k_1)\, \phi_t(k_1^{-1}r_1 k)\, d^r r_1\, dk_1\,, \\ T_t &= e^{-t\Delta} = L(h, dk) \otimes \{*\mu_{h,k}\}\,. \end{aligned} \tag{2.7.7}$$

But then clearly with $\tilde{M} = \varphi_0(k)(m_R(r)/m_G(r))^{1/2} = \varphi_0 M$ we obtain

$$\tilde{M}^{-1} T_t\, \tilde{M} = \varphi_0^{-1}(h)\, \varphi_0(k)\, L(h; dk) \otimes \{*M^{-1}\mu_{h,k}\}\,,$$

which means that

$$\begin{aligned} \hat{T}_t f(rk) = e^{\lambda t} \varphi_0^{-1}(k) \iint \varphi_0^{-1}(h)\, f(rr_1^{-1}h)\, \phi_t(h^{-1}r_1 k) \\ \cdot\, m_G^{1/2}(r_1)\, m_R^{1/2}(r_1)\, dr_1\, \hat{d}h\,. \end{aligned}$$

Observe also that, with our previous notations, when $N \subset R$ is the nilradical of R and $R/N = \mathbb{R}^n$ if we project the operator (2.7.7) on $\mathbb{R}^n \times K$ we obtain

$$T_t f(x,k) = \iint f(x - x_1, k_1) \left(\int_N \phi_t(k_1^{-1} n x_1 k)\, dn \right) dx_1\, dk_1\,.$$

2.8. Left invariant Markov chains and the semidirect product decomposition.

We shall consider here $\{x_n \in X : \ n = 1, 2, \dots\}$ a left invariant Markov chain as in Section 2.2 and assume that $R = N \rtimes H$ is a semidirect product with $N \subset R$ a normal subgroup as in Section 2.7. We can identify here $X/N = R/N \times K$ with $Y = H \times K$ and $X = N \times Y$. Let us denote by $\pi : X \to Y$ the canonical projection and by $\mathcal{Y} = \{y_n = \pi(x_n) \in Y : \ n = 1, 2, \dots\}$ the corresponding left invariant chain on Y. With the above identifications we set $x_n = (z_n, y_n)$ ($z_n \in N,\ y_n \in Y, n = 1, \dots$). We shall examine closely the process

$$\mathcal{Z} = \{z_n \in N :\ n = 1, 2, \dots\}.$$

The process $\mathcal{Z}$ is not in general markovian but if we condition on the paths $(y_1, y_2, \dots)$ of $\mathcal{Y}$ $\mathcal{Z}$ becomes a Markov chain. This is a very important fact for us and we shall analyse it here in detail.

To help the reader see what is happening, let us first look at the special case when $N = R$, $Y = K$. If we use a normal representation

$$T_j = L_j(h; dk) \otimes \{*\mu_{h,k}^{(j)}\} \tag{2.8.1}$$

of the transition operator, we see that conditionally on $(k_1, k_2, \dots)$ ($k_j \in K$) being fixed, the process $\{z_n \in R :\ n \geq 1\}$ is the Markov chain on R with transition operators

$$f \longmapsto f * \mu_{k_{j-1},k_j}^{(j)}\,, \qquad j = 2, 3, \dots$$

It is this idea that we generalize when $R = N \leftthreetimes H$. The key fact here is that any probability measure μ on R can be disintegrated

$$\mu = \int_H \lambda_x \, d\nu(x)\,, \qquad \nu \in \mathbb{P}(H),\ \lambda_x \in \mathbb{P}(xN),\ x \in H\,.$$

For simplicity again let us assume that $K = \{e\}$ is the one point set (this is the basic case treated in [13] and it will help the reader at this point to consult that references). The transition operator are then

$$T_j = *\mu^{(j)}$$

for probability measures on R

$$\mu^{(j)} = \int_H \lambda_y^{(j)} \, d\nu^{(j)}(y)\,.$$

The measures $\lambda_y^{(j)}$ can (for every fixed $y \in H$) be identified to $\lambda_y^{(j)} \in \mathbb{P}(N)$ by $zy \leftrightarrow z$ ($z \in N$) and since now $x_n = z_n y_n$ we easily see that with a fixed $(y_1, y_2, \dots)$ the process $\{z_1, z_2, \dots\}$ is a Markov chain on N with transition operators

$$f \longmapsto f * \pi_j\,,$$

where $\pi_1 = \lambda_{y_1}$ and $\pi_j = (\lambda^{(j)}_{y_j})^{y_1 \cdots y_{j-1}}$ $(j \geq 2)$ with the notation λ^x $(\lambda \in \mathbb{P}(R)\ x \in R)$ for the image of λ by the inner automorphism $g \mapsto xgx^{-1}$. An alternative way of viewing the above situation is to observe that if we consider arbitrary measures $\lambda_j \in M(N)$ $(j = 1, \ldots, n)$ and place then on the cosets Ny_j by the identification $z \leftrightarrow zy_j$ then the convolution (in R) of these measures (that are placed on the cosets) lies in the coset $y_1 y_2 \cdots y_n$ and corresponds to the measure $\lambda_1 * \lambda_2^{y_1} * \lambda_3^{y_1 y_2} * \cdots * \lambda_n^{y_1 \cdots y_{n-1}}$ where now the convolution is taken in N.

The above two special cases ($K = \{e\}$ and $N = R$) can now be put in the general context: we identify $X = R \times K = N \times Y = N \times H \times K$ so that $x_n = (r_n, k_n) = (z_n, h_n, k_n)$, $y_n = (h_n, k_n) \in Y$ and with fixed $(y_1, y_2, \ldots)$ we disintegrate

$$\mu_{k_{j-1}, k_j} = \int_H \lambda^{(j)}_h \, d\nu^{(j)}(h)\,,$$

$(z_1, z_2, \ldots)$ is then a Markov chain on N with transition operators

$$f \longmapsto f * (\lambda^{(j)}_{h_j})^{h_1 \cdots h_{j-1}}\,, \qquad j = 2, \ldots$$

2.9. Bi-invariant operators revisited.

Nothing in this section is very new but I felt that it was appropiate to close this chapter by making the connection with know and standard ideas related to K-bi-invariant operators on semisimple groups.

Let G be some Lie group that can be written $G = R \cdot K$, $R \cap K = \{e\}$ for two closed subgroups with $Z \subset K$ as in Section 2.1 so that $m_G|_K \equiv 1$. I shall consider on G a differential operator Δ without constant term (*i.e.* $\Delta 1 \equiv 0$) that is G-left invariant K-right invariant and is in particular formally self adjoint and positive with respect to right measure $d^r g$. What we want is "somehow" to identify Δ with an operator on R. To do this we first conjugate Δ to $\tilde{\Delta} = m_G^{1/2} \Delta m_G^{-1/2}$ to make it formally self adjoint with respect to $dg = d^\ell g$. This of course creates a constant term $\tilde{\Delta} 1 = C$ which in general is not zero. Let us consider $D = \tilde{\Delta} - C$ which is now a G-left, K-right invariant operator without constant term that is formally self invariant with respect to dg. From now onwards we shall consider operators D that have the above properties. When $G = NAK$ is semisimple D, the classical K-bi-invariant Laplacian, has the above properties.

Quite generally an operator on G that has the above properties can be identified with a G-invariant operator on the homogeneous space $G/K = \{gK; g \in G\}$ (When $G = NAK$ is semisimple G/K is the symmetric space and the most important example of the above situation is that of the Laplace-Beltrami operator on G/K). The homogeneous space G/K can be identified with R, we obtain thus an identification of D with an operator $D_{G/K}$ on G/K and D_R on R. D_R is clearly R-left invariant has no constant term and since the G-invariant measure on G/K (which always exists since $m_G|_K \equiv 1 \equiv m_K$) can be identified with the left Haar measure of R, D_R is formally self adjoint with respect to $dr = d^\ell r$ (indeed $D_{G/K}$ is clearly formally self adjoint with respect to the invariant measure on G/K). It follows therefore that

$$D_R = m_R^{1/2}\Big(-\sum X_j^2 + C_R\Big)m_R^{-1/2}\,,$$

where $X_1, \dots, X_n$ are left invariant fields on R. The only issue here is to determine the constant C_R. To do this let λ_D be the spectral gap of operator $D_{G/K}$ on G/K. The operator $D_R - \lambda_D$ has then zero spectral gap on $L^2(R; dr)$ and therefore $\tilde{D}_R = -\sum X_j^2 + C_R - \lambda_D$ has zero spectral gap on $L^2(R; d^r r)$. If we assume, as is the case in all the interesting examples, that R is soluble, and therefore amenable, the spectral gap of $-\sum X_j^2$, which is a markovian generator is 0. It follows that the spectral gap of $\tilde{D}_R$ is $C_R - \lambda_D$ and that $C_R = \lambda_D$. The conclusion is that $D_{G/K} - \lambda_D$ can be identified with $m_R^{1/2}(-\sum X_j^2)m_R^{-1/2}$.

An alternative way to compute C_R is to observe that m_R is multiplicative and therefore that $X_j m_R = \lambda_j m_R$ ($\lambda_j \in \mathbb{R}$, $j = 1, 2, \dots$) $X_j m_R^\alpha = \alpha\lambda_j m_R^\alpha$, $(\sum X_j^2)m_R^\alpha = \alpha^2 \sum \lambda_j^2 m_R^\alpha$ and that therefore the constant term of $m_R^{1/2}(-\sum X_j^2)m_R^{-1/2}$ is $(-\sum \lambda_j^2)/4$. This gives, in view of the fact that D_R has no constant term, that

$$C_R = \frac{1}{4}\sum \lambda_j^2 = \frac{1}{4}\,\rho^2\,.$$

In the case of the Laplace-Beltrami operator on a symmetric space the above considerations amount to the standard way of computing the spectral gap in terms of the roots. Observe finally that by an easy calculation we have

$$D_R\, m_R^\alpha = \rho^2(1/4 - (\alpha - 1/2)^2)\, m^\alpha\,.$$

This shows that $m_R^{1/2} = \phi_0$ is an eigenfunction of D_R with $D_R\phi_0 = \lambda_D\phi_0$ *i.e.* that ϕ_0 is the "ground state" of the Laplace-Beltrami operator.

Let now G be an arbitrary real Lie group and let $K \subset G$ be an arbitrary compact subgroup or more generally a subgroup K that contains $Z \subset K$ a central subgroup such that K/Z is compact. It is then very easy to see that K/Z acts by inner automorphism ($I_{\dot{k}} : x \to kxk^{-1}$ $k \in \dot{k} \in K/Z$) on G. It follows that if Δ is an arbitrary on G then $\tilde{\Delta} = \int_{K/Z} dI_{\dot{k}}(\Delta)\, d\dot{k}$ is K-bi-invariant.

A similar analysis can be done for K-bi-invariant convolution operators $f \mapsto f * \mu$ on G (*i.e.* when μ is stable by the action of I_k, $k \in K$).

Finally when $G = R \cdot K$ as in Section 2.1 the above considerations show that for K-bi-invariant Laplacians and K-bi-invariant convolution operators both Theorem A and Theorem B reduce to the analogous theorems on R. When R is soluble and the spectral gap is zero these results have been proved in [13].

3. Gaussian measures on groups.

3.1. Elementary facts on the geometry of groups.

Let G be a connected real Lie group and let $X_1, \dots, X_k$ be left invariant fields that satisfy the Hörmander condition. These fields define therefore a left invariant distance $d(\cdot,\cdot)$ on G, *cf.* [1]. We shall always denote by $|g|_G = |g| = d(e,g)$. The thing to remember is that "at infinity" $|\ |_G$ only depends on G and is independent of the particular choice of the fields $X_1, \dots, X_k$. More precisely for every $e \in \Omega$ Nhd of the identity and for a new choice $X_1^*, \dots, X_s^*$ of fields as above we have

$$C^{-1}|g|^{\text{old}} \le |g|^{\text{new}} \le C|g|^{\text{old}}, \qquad g \in G \backslash \Omega .$$

It is clear of course that $|hg| \le |h| + d(h,hg) = |h| + |g|$ and that $|g^{-1}| = d(e,g^{-1}) = d(g,gg^{-1}) = |g|$. It follows in particular that $|\,|hg| - |g|\,| \le |h|$, $|\,|gh| - |g|\,| \le |h|$ $(g,h \in G)$ and therefore also that $|\,|h_1 g h_2| - |g|\,| \le |h_1| + |h_2|$ $(g, h_1, h_2 \in G)$.

We shall also denote by

$$B(r) = B_G(r) = \{g \in G : \ |g| \le r\}$$

the corresponding r-ball.

Let now $H \subset G$ be some closed subgroup and let m_H denote either the left or the right Haar measure of H. There exists $c > 0$ then such that

$$m_H\{h \in H :\ |h|_G \le r\} \le e^{cr}\,, \qquad r > 0\,. \tag{3.1.1}$$

Observe that the above set $H_r = \{h \in H :\ |h| \le r\}$ is not "equivalent" with the $B_H(r)$ the r-ball in H. Observe also that since the involution $h \to h^{-1}$ $(h \in H)$ interchanges the left and right Haar measure on H the statement (3.1.1) need only be proved for the right measure m_H. The proof of (3.1.1) is easy. Indeed the left distance on G induces $|\cdot|_{G/H}$ a distance of the homogeneous space $\{Hg :\ g \in G\}$ and if for every $\dot g \in G/H$ with $|\dot g|_{G/H} \le r$ we fix as we may some $g \in \dot g$ with $|g| \le 2|\dot g|$ we clearly have

$$\bigcup_{|\dot g| \le r} H_r g \subset B_G(3r)\,. \tag{3.1.2}$$

It is clear also that we can "disintegrate" $m_G = m_H \otimes m_{G/H}$ for some appropiate C^∞-non vanishing measure on G/H so that (3.1.2) gives

$$m_H(H_r) \cdot m_{G/H}(B_{G/H}(r)) \le m_G(B_G(3r))$$

with obvious notations. $m_{G/H}$ is the Haar measure of G/H if H is normal but in general it does not have to be G-invariant. What however always holds is that $m_{G/H}(B_{G/H}(r)) \ge \varepsilon_0 > 0$ $(r \ge 1)$ and (3.1.1) follows from the well known and obvious fact (*cf.* [11]) that

$$\gamma(r) = m_G(B_G(r)) \le C\, e^{cr}\,, \qquad r > 0\,. \tag{3.1.3}$$

What is clear also is that for any closed analytic subgroup $H \subset G$ we have $|h|_G \le C|h|_H$ $(h \in H)$ the best estimate the other way around is (*cf.* [18], [13], [43])

$$|h|_H \le C \exp(c\,|h|_G)\,, \qquad h \in H\,. \tag{3.1.4}$$

The proof of (3.1.4) is non trivial. If G is algebraic (3.1.4) follows from general considerations (*cf.* [18]). If G is simply connected soluble and $H = N$ is in the nilradical (3.1.4) was proved in [13] (*cf.* also Section 1.6). This is the only case that will be needed in this paper. In the special case when we can write $G = H \cdot K$ where $K \Subset G$ is a compact subset we have $|h|_H \approx |h|_G$ $(h \in H)$. (This is because for any $h, h' \in H$ we can find $h = h_1, \dots, h_n = h' \in H$ such that $d_G(h_j, h_{j+1}) \le C$, $n \le Cd(h, h'))$. When $G = H \cdot K$ where K is a closed group that contains $Z \subset K$ a discrete central (in G) subgroup such that K/Z is compact (as in Section 2.1) and $H \cap Z = \{0\}$, we again have $|h|_H \approx |h|_G$, provided that the image of H in G/Z is closed. Indeed if we denote by $\pi : G \to G/Z = H \cdot (K/Z)$ the canonical projection we have from the above remark $|h|_H \approx |h|_{G/Z}$ but quite generally we also have $|\pi(g)|_{G/Z} \leq |g|_G$ ($g \in G$) and our result follows.

Observe finally that the above remark together with the structure theorems of Lie groups allows us to reduce the proof of (3.1.4) to the case when G is soluble. That reduction is however non trivial (*cf.* Section 4.8 and [43]). For a soluble group G which we can further assume to be simply connected, the proof of (3.1.4) is done by the use of the "exponential coordinates of the second kind". One first proves that when the group G of [6, Theorem 3.18.11] is nilpotent, then the coordinates $(t_1, \dots, t_m)$ of $g \in G$ are $O(|g|^N)$, this is easily done by induction. We shall then choose the basis $X_1, \dots, X_m$ in [6, Theorem 3.18.11] in such a way that $X_1, \dots, X_n$ (for some $n \leq m$) is a basis of the nilradical. A simple use of the above special case and the results of [13] show then that, in general, the coordinates satisfy $|(t_1, \dots, t_m)| = O(\exp(c|g|))$. From this and the proof of Theorem 3.18.12 in [6] our assertion (3.1.4) follows. The details will be left for the reader.

3.2. Functions and measures on a group.

Let G be some real connected Lie group and let $\varphi(g) \in C^\infty(G)$. We shall say that φ is an Ex-function (Ex- for "Exponential") if there exists $C > 0$ such that

$$C^{-1}\exp(-C|g|) \leq \varphi(g) \leq C\exp(C|g|)\,, \qquad g \in G\,,$$

and if for any sequence of left invariant fields $X_1, \dots, X_k, \dots$ there exist $C_k, \tilde{C}_k > 0$ ($k \geq 1$) such that

$$|X_1 X_2 \cdots X_k\, \varphi(g)| \leq \tilde{C}_k \exp(C_k\, |g|)\,, \qquad g \in G\,. \tag{3.2.1}$$

Similarly we shall say that φ is a Gs-function (Gs for "Gaussian") if there exist $C_\pm, \tilde{C}_\pm > 0$ such that

$$\tilde{C}_- \exp(-C_-\, |g|^2) \leq \varphi(g) \leq \tilde{C}_+ \exp(-C_+\, |g|^2)\,, \qquad g \in G\,, \tag{3.2.2}$$

and for any sequence $X_1, \ldots, X_k, \ldots$ there exist $C_k, \tilde{C}_k > 0$ $(k \geq 1)$ such that

$$(3.2.3) \qquad |X_1 \cdots X_k\, \varphi(g)| \leq \tilde{C}_k \exp(-C_k\, |g|^2)\,, \qquad g \in G\,.$$

We shall sometimes say that φ is a strict Gs-function if for any $\varepsilon > 0$ in the above estimates we can choose

$$C_{\pm} = \frac{1}{4 \pm \varepsilon}\,, \qquad C_k = \frac{1}{4 + \varepsilon}\,,$$

and where $\tilde{C}_{\pm}(\varepsilon)$, $\tilde{C}_k(\varepsilon)$ depend on $\varepsilon > 0$. In the rest of this section we shall examine closely the above notations.

First of all it is clear that if $m_1, m_2 \in$ Ex (*i.e.* are Ex-functions) if $\varphi_1, \varphi_2 \in$ Gs (*i.e.* are Gs-functions) if $\alpha_1, \alpha_2 \in \mathbb{R}$; $n_1, n_2 = 1, 2, \ldots,$ $m_1^{\alpha_1} m_2^{\alpha_2} \in$ Ex, $\varphi_1^{n_1} \varphi_2^{n_2} \in$ Gs, $m\varphi \in$ Gs.

Typically any positive character (*e.g.* the modular function m_G) is an Ex-function. More generally when $(m_{ij}) = M : G \to GL_n(\mathbb{R})$ is a group homomorphism then each matrix coefficient m_{ij} is $O(\exp(C|g|))$ and satisfies (3.2.1) (This is because $m_{ij}(gx) = \sum m_{i\alpha}(g) m_{\alpha_j}(x)$ and the fields X_k are left invariant).

It follows in particular that if

$$(3.2.4) \qquad d\mu = \varphi\, dg = \psi\, d^r g$$

is a positive measure on G then $\varphi \in$ Gs if and only if $\psi \in$ Gs. A measure μ as in (3.2.4) with $\varphi \in$ Gs will be called a Gs-measure.

Let now Y be a right invariant field. It is clear then that $Y(g) \in T_g(G)$ (*i.e.* the value of the field at $g \in G$) coincides with $X(g)$ the value at g of the left invariant field X for which $X(e) = \operatorname{Ad} g Y(e)$). The upshot is that $Y(g) = M(g)(X_1, \ldots, X_n)^T$ where $M(g) \in GL_n(\mathbb{R})$ is as above and $(X_1, \ldots, X_n)$ is a basis of left-invariant fields. From this and our previous remarks we see that in the above definition of Ex or Gs-functions we can replace left invariant fields by right invariant fields. If we use the notations

$$\check{f}(g) = f(g^{-1})\,, \quad f^h(g) = f(gh)\,, \quad f_h(g) = f(hg)\,, \qquad g, h \in G\,.$$

The above considerations show that $\check{\varphi} \in$ Ex (respectively $\check{\varphi} \in$ Gs) if and only if $\varphi \in$ Ex (respectively $\varphi \in$ Gs). Also if $\varphi \in$ Ex (respectively $\varphi \in$ Gs) and $k \in G$ then $\varphi^k, \varphi_k \in$ Ex (respectively $\in$ Gs) and that this is so uniformly (*i.e.* with uniform constant) as $k \in K \Subset G$ runs through the compact subsets of G. I shall leave the reader with the task to work out which of the above observation extend to strict Gs-functions and which do not.

We shall finally need to extend the above notations to the product space $X = R \times K$ as in Section 2.1. We shall say that $\varphi \in C^\infty(X)$ is an Ex- (respectively Gs-) function on X if and only if the functions $\varphi_k(r) = \varphi(r, k)$ are Ex- (respectively Gs-) functions uniformly in $k \in K$ (*i.e.* with uniform constants).

A typical example of a Gs-function on $R \times K$ is supplied by the convolution operator in Section 2.3 where the convolution measure $d\mu = \varphi(g)\, dg$ is Gs. The formula (2.3.6) for $M(r; k_1, k)$ and our previous remarks show that $M(r; k_1, k)$ is Gs on R uniformly in k, k_1 provided that K is compact or more generally, uniformly when $k_1^{-1} k \in K_0 \Subset K$ where K_0 is some compact subset of K. Indeed $m_R(r) \in \text{Ex}$ on R and $\psi(g) = \varphi(k_1^{-1} g k) m_G^{-1}(g) \in \text{Gs}$ on G. To show that $\psi|_R$ and therefore M is Gs on R it suffices therefore to use Section 3.1 and the fact that when K is compact we have

$$C^{-1} |r|_G \le |r|_R \le C\, |r|_G \,, \qquad r \in R,\ |r|_G \ge C\,. \tag{3.2.5}$$

Due to the fact that Z is a central subgroup, the estimate (3.2.5) also holds when K is not compact (*cf.* end of Section 3.1) provided that $k_1^{-1} k \in K_0$.

Another notion that will be used is that of a Gs R-left invariant positive operator T on $X = R \times K$ as in Section 2.2. We shall write

$$T = L(h, dk) \otimes \{ * \mu_{h,k} \}$$

in normal form as in Section 2.2 and we shall say that T is Gs on X if the measures $\mu_{h,k} \in \text{Gs}$ on R uniformly in $h, k \in K$. It follows that when K is compact then the operator T that corresponds to a convolution operator on G by a Gs measure is Gs in the above sense.

3.3. Subgroups and quotients.

Let $H \subset G$ be as in Section 3.1 (or at least some closed analytic subgroup for which (3.1.4) is known to hold), and let $\varphi \in \text{Gs}$ on G. I shall consider the restricted function $\tilde\varphi = \varphi|_H \in C^\infty(H)$ by Section 3.1 it is clear that

$$\begin{aligned} \tilde\varphi(h) &\le \tilde C_+ \exp\left(-C_+ \log^2(|h|+1)\right), \\ |X_1 \cdots X_k\, \tilde\varphi(h)| &\le C_k \exp\left(-C_k \log^2(|h|+1)\right), \qquad h \in H\,. \end{aligned} \tag{3.3.1}$$

In general however $\tilde{\varphi}$ is not a Gs-function. Let us now assume throughout this section that H is normal and let $\pi : G \to G/H$ be the canonical projection (For the applications that we have in mind $H \cong \mathbb{R}^n$, the distinction that we make below of $d^\ell h$ and $d^r h$ is therefore inessential). Let $m \in G/H$ be an Ex-function on G/H, then $m \circ \pi$ is an Ex-function on G. The analogous statement is in general false for Gs-functions. Quite generally for any $0 \leq \varphi \in C^\infty(G)$ we shall define (possibly $= +\infty$)

$$\begin{aligned} \varphi_\ell(g) = \varphi_\ell(\dot{g}) &= \int_H \varphi(gh)\, d^\ell h\,, \\ \varphi_r(\dot{g}) &= \int \varphi(hg)\, d^r h\,, \qquad g \in \dot{g} = gH \in G/H\,. \end{aligned} \tag{3.3.2}$$

We have $(\check{\varphi})_\ell = (\varphi_r)^\vee$. In what follows it suffices therefore to examine one of the two transforms $\varphi \to \varphi_\ell$ or $\varphi \to \varphi_r$. We shall need the following

Lemma. *Let $H \subset G$ be as above. Then for every $c > \varepsilon > 0$ there exists $C = C(c, \varepsilon)$ such that*

$$\int \exp\left(-c\, |gh|_G^2\right) d^\ell h \leq C \exp\left(-(c-\varepsilon)\, |\dot{g}|_{G/H}^2\right), \tag{3.3.3}$$

for all $g \in G$, $g \in \dot{g} \in G/H$.

PROOF. By Section 3.1 it is clear that we can estimate the above integral by

$$\int_{|\dot{g}|}^\infty \exp\left(-c\,|\xi|^2 + C\,\xi + C\,|\dot{g}|\right) d\xi\,.$$

Indeed by (3.1.1) it is only matter of splitting the integral along the intersection of gH with the shells $\{x \in G : \ |x|_G \in d\xi\} \subset G$. The lemma follows.

In the above lemma we can replace $|gh|_G$ and $d^\ell h$ by $|hg|_G$ and $d^r h$ and the same conclusion holds (indeed we pass from one to the other by the involution $x \to x^{-1}$ in G). With the above notations let us assume that $\varphi \in \mathrm{Gs}$ and that $X_1, X_2, \ldots, X_k$ are left invariant fields and let us denote

$$\varphi_\ell^{(k)}(g) = (|X_1 \cdots X_k\, \varphi(g)|)_\ell\,, \qquad \varphi_r^{(k)}(g) = (|X_1 \cdots X_k\, \varphi(g)|)_r\,.$$

It is clear from the lemma that

$$\varphi_\ell^{(k)}(\dot g)\,,\quad \varphi_r^{(k)}(\dot g)\le \tilde C_k \exp\left(-C_k\,|\dot g|^2\right),\qquad \dot g\in G/H\,. \tag{3.3.4}$$

However it is also true that

$$\varphi_\ell(\dot g)\,,\quad \varphi_r(\dot g)\ge \tilde C_- \exp\left(-C_-\,|\dot g|^2\right),\qquad \dot g\in G/H\,. \tag{3.3.5}$$

Indeed for fixed $\dot g\in G$ let $g\in\dot g$ be chosen so that $|g|_G\le|\dot g|_{G/H}+\varepsilon$ and since with $h\in H$, $|h|_H\le 1$ we have (*cf.* Section 3.1)

$$\big|\,|gh|_G-|g|_G\,\big|\,,\qquad \big|\,|hg|_G-|g|_G\,\big|\le 1\,.$$

(3.3.5) follows by restricting the integration in (3.3.2) to the ball $|h|\le 1$.

Let now X and Y be a left invariant and a right invariant field respectively on G and let $\dot X$, $\dot Y$ the corresponding projected fields on G/H. It is evident (from the definition $Xf(g)=\lim(f(ge^{tX})-f(g)))/t$; $Yf(g)=\cdots$) that

$$\dot X\varphi_r(\dot g)=\int_H (X\varphi)(hg)\,d^r h\,,$$
$$\dot Y\varphi_\ell(\dot g)=\int (Y\varphi)(gh)\,d^\ell h\,.$$

The analogous expressions for the "multiple derivatives" $\dot X_1\dot X_2\cdots\dot X_k\varphi_r$ also hold. If we use this remark together with (3.3.3), (3.3.4), (3.3.5) we coclude that φ_r, φ_ℓ are both Gs-functions and that furthermore they are strict Gs-functions, for the quotient metric, if φ is.

It is clear that the above considerations generalize to Gs-functions on $X=R\times K$ where for each $\varphi\in C^\infty(X)$ and $H\subset R$ a closed normal subgroup the corresponding functions $\varphi_r,\varphi_\ell\in C^\infty(R/H\times K)$ are defined in the obvious way for every slice $\varphi(\cdot,k)$ separetly.

We shall now consider more closely the restriction of Gs-function on a subgroup or more generally on a coset gH. Motivated by (3.3.1) we shall say quite generally that for any Lie group H, $f\in C^\infty(H)$ is an Sp-function (superpolynomial) with constants $c;C_0,C_1,\dots>0$ if

$$|X_1\cdots X_k f(h)|\le C_k\exp\left(-c\log^2(|h|+1)\right),\qquad h\in H,\ k\ge 0\,. \tag{3.3.6}$$

It is thus clear that the restriction to $H\subset G$ of a Gs-function on H is Sp on H.

More generally let $f \in \mathrm{Gs}$ on G and let us define

$$f_g(h) = \Big(\int_H f(gh)\, d^\ell h\Big)^{-1} f(gh)\,, \qquad g \in G,\ h \in H\,.$$

(One should observe that for all our applications H will be in fact unimodular and $d^r h = d^\ell h$. More generally by choosing a global analytic section of $G \to G/H$, which always exists in the simply connected case, we can find an Ex-function on G that allows us to pass from the $d^r h$ measure to the $d^\ell h$ measure. We shall have no use of this fact however and therefore we shall not elaborate further). Just as before if $X_1, \dots, X_k$ are left invariant fields on H, which can be identified to left invariant fields on G, we clearly have

$$(X_1 \cdots X_k f)_g(h) = X_1 \cdots X_k (f_g)(h)\,, \qquad g \in G,\ h \in H\,,$$

and if $f \in \mathrm{Gs}$ on G by (3.3.5) we have

$$\int f(gh)\, d^\ell h \geq C \exp\left(-C\, |\dot{g}|^2\right).$$

The upshot of the above consideration is that

$$\begin{aligned} |X_1 \cdots X_k (f_g)(h)| &\leq \exp\left(C_1 |g|_G^2 - C_2 |gh|_G^2\right) \\ &\leq \exp\left(c_3 |g|_G^2 - c_4 |h|_G^2\right). \end{aligned} \tag{3.3.7}$$

If we combine this with (3.1.4) we conclude that for every $g \in G$ the function $f_g \in C^\infty(H)$ is an Sp-function with a constant $c > 0$ in (3.3.6) that only depends on f and where

$$C_k = C_k(g) \leq \tilde{C}_k \exp\left(c_k |g|^2\right), \qquad k = 0, 1, \dots$$

The constants $c, \tilde{C}_k, c_k > 0$ clearly only depend on the constants of the definition (3.2.2)-(3.2.3), and the estimates (3.3.7) are uniform for a family of functions f that are uniformly Gs on G.

3.4. Mass escape at infinity of the convolution product.

Let $\mu_j \in \mathbb{P}(G)$ $(j = 1, 2, \dots)$ be a sequence of probability measures on G that are Gs-measures uniformly in $j = 1, 2, \dots$ If we bare in mind (3.1.3) we see that this implies that

$$\mu_j\{g \in G:\ |g| > R\} \leq C \exp\left(-c\, R^2\right), \qquad R > 0,\ j = 1, 2, \dots,$$

where $C, c > 0$ are independent of $R > 0$ and $j = 1, 2, \dots$ If we take the convolution products $\mu^{(n)} = \mu_1 * \cdots * \mu_n$ we deduce

$$\begin{aligned}\mu^{(n)}\{g \in G : \ |g| > R\} &\leq \sum_{j=1}^{n} \mu_j \{g \in G : \ |g| \geq R/n\} \\ &\leq C\, n \exp\left(-c\,(R/n)^2\right), \qquad R > 0,\ n = 1, 2, \dots\end{aligned}$$

We have in particular

$$\mu^n\{g \in G : \ |g| \geq n^{1+\varepsilon}\} \leq C \exp\left(-c\, n^{2\varepsilon}\right), \qquad n = 1, 2, \dots,\ \varepsilon > 0\,.$$

Similarly we can consider probability measures $d\mu_j(g) = f_j(g)\, dg \in \mathbb{P}(G)$ where $f_j \in \mathrm{Sp}$, $j = 1, 2, \dots$ and where for simplicity we shall assume that G is unimodular. More precisely we shall demand that there exist $c > 0$, $C_1, C_2, \dots > 0$ such that

$$f_j(g) \leq C_j \exp\left(-c \log^2(|g| + 1)\right), \qquad j = 1, 2, \dots,\ g \in G\,.$$

We shall assume further that G is a group of polynomial growth *i.e.* that

$$\gamma(r) = \text{Haar measure of } B_G(r) \leq C\,(r+1)^A, \qquad r > 0\,.$$

It then follows that

$$\mu_j\{g \in G : \ |g| \geq R\} \leq C\, C_j \exp\left(-c \log^2(R+1)\right),$$

for $j = 1, 2, \dots$, $R > 0$, and therefore, as before, the convolution product $\mu^{(n)} = \mu_1 * \cdots * \mu_n$ satisfies

$$\mu^{(n)}\{g \in G : \ |g| \geq R\} \leq C \left(\sum_{j=1}^{n} C_j\right) \exp\left(-c \log^2(R/n + 1)\right),$$

for $R > 0$, $n = 1, 2, \dots$, with $R = n^{1+\varepsilon}$ we have in particular

$$\mu^{(n)}\{g \in G : \ |g| \geq n^{1+\varepsilon}\} \leq C \sup_{1 \leq j \leq n} C_j \exp\left(-c \log^2 n\right),$$

where $C, c > 0$ are independent of n.

3.5. The Heat kernel.

Let $\Delta_0 = -\sum X_j^2$ be a driftless subelliptic Laplacian and let $\phi_t(g)$ be the corresponding convolution kernel as in Section 0. For every fixed $t > 0$ the function $\phi_t(g)$ is then a Gs-function on G (*cf.* [1]). In fact $\phi_1(g)$ is a strict Gs-function. The strict upper estimate is contained in [1]. The strict lower estimate is (implicitely) contained in [19] (especially Section 2.4, [19] II). Since we shall be able to complete the proofs of our theorems without the strict estimates, we shall not give the details here.

Let now $\Delta = -\sum_{j=1}^n X_j^2 + X_0$ be a general subelliptic Laplacian (*i.e.* X_0 need not be zero). The convolution kernel ϕ_t is again, for every fixed $t > 0$, a Gs-function. The proof of the upper estimate has been written out in a much more general context in [20]. For an alternative simple proof, (*cf.* Section A.4). The lower estimate when $X_0 = \sum \lambda_i X_i + \sum \lambda_{ij}[X_i, X_j]$ is an easy consequence of the scaled Harnack estimate (*cf.* also [21]). For a general drift however this lower Gaussian estimate is difficult to prove (*cf.* Section A.4).

From the above and the considerations at the end of Section 3.2 we see that T the left invariant operators on $R \times K$ that corresponds to the semigroup $T_t = e^{-t\Delta}$ on G as in sections 2.5, 2.6, 2.7 are Gs-operators when K is compact. This statement remains true in general, even when K is not compact, but this statement is not trivial to prove. Since we shall be able to do without this general case we shall not give this proof here.

4. Upper estimates.

4.1. Gaussian measures on a special class of groups.

In this section we shall consider a real Lie group G and $H \subset G$ a closed normal subgroup that satisfy the following conditions: $H \cong \mathbb{R}^n$ and $G/H \cong V \times S$ where $V \cong \mathbb{R}^m$ and S is compact. We shall summarize this information in the exact sequence

$$(4.1.1) \qquad 0 \longrightarrow H \cong \mathbb{R}^n \longrightarrow G \underset{p}{\longrightarrow} G/H \cong V \times S \cong \mathbb{R}^m \times S \longrightarrow 0\,.$$

The above situation is not as special as it looks. Indeed let $\overline{G}$ be simply connected group and let $N \subset Q \subset \overline{G}$ be its radical and nilradical. Let further $[N,N]$ the analytic subgroup that corresponds to $[\mathfrak{n},\mathfrak{n}]$ where $\mathfrak{n} \subset \overline{\mathfrak{g}}$ is the nilradical of the Lie algebra of $\overline{G}$. Then the group $G = \overline{G}/[N,N]$ satisfies the above conditions (4.1.1) with $H = N/[N,N]$, $V = Q/N$ and $S = \overline{G}/Q$. S is a semisimple group. When $\overline{G}$ is soluble we have $S = \{e\}$. When $\overline{G}$ is amenable (*e.g.* when $\overline{\mathfrak{g}}$ is an R-algebra) then S is compact. Observe that if we assume in addition that $\overline{\mathfrak{g}}$ is an algebraic algebra (*i.e.* that it is the Lie algebra of some algebraic Lie group) in the above situation we have $\overline{G}/[N,N] = G = H \leftthreetimes (V \times S)$ (*cf.* [7]) and if $\overline{G} = Q$ is soluble we have $Q/[N,N] = H \leftthreetimes V$.

The basic thing to observe is that under the condition (4.1.1) the group G/H ($\cong \mathbb{R}^m \times S$) can be made to act naturally on H so as to have

$$\pi : G/H \longrightarrow GL(H)\,. \tag{4.1.2}$$

This is of course true in general (and trivially so) when $G \cong H \leftthreetimes G/H$ (*e.g.* G simply connected and $H = Q$, the radical of G) but here the action (4.1.2) is obtained from inner automorphisms because H is abelian. Indeed for $x \in G/H$ we choose some $g \in G$ such that $p(g) = x$ and then the action $h \to g^{-1}hg$ is independent of the particular choice of g.

The Lie algebra of V and H will be identified with V and H respectively and we shall consider

$$d\pi : V \longrightarrow \mathfrak{gl}(H)\,. \tag{4.1.3}$$

We shall also consider the roots of the action (4.1.3) which are $\lambda \in \mathrm{Hom}_{\mathbb{R}}[V;\mathbb{C}]$ and are defined by

$$(d\pi(v) - \lambda(v))w = 0\,, \qquad v \in V\,,$$

and some $0 \neq w \in H \otimes_{\mathbb{R}} \mathbb{C}$. The corresponding root spaces $U_\lambda \subset H \otimes_{\mathbb{R}} \mathbb{C}$ are defined accordingly.

I shall defined then $L_1, L_2, \dots, L_p$ all the distinct real parts ($L = \mathrm{Re}\,\lambda$) of these roots (contrary to what was done in Chapter 1 the zero real part is also admitted here). If then $H_j \subset H$ is defined by the fact that $H_j \otimes \mathbb{C} = \sum U_\lambda$ for all the λ's such that $\mathrm{Re}\,\lambda = L_j$ we obtain $H = H_1 \oplus \cdots \oplus H_p$ a decomposition of H as a direct sum of subspaces. All the subspaces H_j are stable by the representation (4.1.2) and are such that if $G/H \ni \dot{g} = (v,s)$ ($v \in V$, $s \in S$) we have

$$\det(\pi_j(\dot{g})) = e^{d_j L_j(v)}\,, \tag{4.1.4}$$

where $d_j = \dim H_j$ and $\pi_j = \pi|H_j$ (This is because the determinant is real and its modulus is clearly given by (4.1.4)). Observe also that in the relevant cases S is compact or semisimple and therefore $\det \pi(s) = 1$, $s \in S$. All the above facts are consequences of elementary linear algebra and will thus be left for the reader.

It is of course clear that if G satisfies (4.1.1) then we can represent

$$0 = \sum_{j=1}^{p} \alpha_j L_j \,, \qquad \alpha_j \geq 0 \,, \tag{4.1.5}$$

non trivially (*i.e.* not all $\alpha_j L_j = 0$ in the above sum) if and only if $\mathfrak{g}$ is a C-algebra.

Proposition. *Let G be a real Lie group that satisfies the conditions* (4.1.1) *and let us assume that the Lie algebra $\mathfrak{g}$ satisfies the* C*-condition. Let $\mu_j \in \mathbb{P}(G)$, $j = 1, 2, \ldots$ be a sequence of probability measures and let us assume that $\mu_j \in$* Gs *on G uniformly in j (i.e. with constants that are independent of j). Let further $\mu^{(n)} = \mu_1 * \cdots * \mu_n$ be the corresponding convolution products.*

Then there exists $c > 0$ such that for every $f \in C_0^\infty$ we have

$$\langle \mu^{(n)}, f \rangle = O(e^{-cn^{1/3}}) \,, \qquad n \geq 1 \,. \tag{4.1.6}$$

In fact we have $d\mu^{(n)}(g) = \varphi^{(n)}(g)\, dg$ where $\varphi^{(n)} \in C^\infty(G)$ (and even $\varphi^{(n)} \in$ Gs on G but this is irrelevant) and "morally" what the estimate (4.1.5) actually says is

$$\varphi^{(n)}(g) = O(e^{-cn^{1/3}}) \,, \qquad g \in G \,.$$

The proof of (4.1.6) will be given in Section 4.5.

4.2. The Fourier transform.

$G \supset H$ and all other notations will be as in Section 4.1. We shall consider $d\mu(g) = \varphi(g)\, dg$ a Gs-probability measure on G and define

$$\theta_{\dot g}(h) = \left(\int_H \varphi(gh)\, dh \right)^{-1} \varphi(gh) \,, \qquad h \in H,\ g \in \dot g \in G/H \,,$$

the above function will be identified with a function $\theta_{\dot g}(x) \in C^\infty(\mathbb{R}^n)$ uniquely defined up to translation $\cdot \mapsto \cdot + \alpha$ on $\mathbb{R}^n$ ($\alpha \in \mathbb{R}^n$). The modulus of the Fourier transform

$$f_{\dot g}(\xi) = |\hat\theta_{\dot g}(\xi)|$$

is therefore uniquely defined. For the sequence of measures given in the proposition of Section 4.1 we shall consider $d\mu_j(g) = \varphi_j(g)\,dg$ ($j = 1, 2, \dots$) and the corresponding $f_g(\xi) = f_{j,\dot g}(\xi) = |\hat\theta_g(\xi)|$ (for typographical reasons we shall drop the $j = 1, \dots$ and the "dot" above the g). We shall need the following

Lemma. *Let $\hat H = \hat H_1 \oplus \cdots \oplus \hat H_p$ be the dual decomposition of $H = H_1 \oplus \cdots \oplus H_p \cong \mathbb{R}^n$ and let $\xi = (\xi_1, \dots, \xi_p) \in \hat H$, $\xi_i \in \hat H_i$, $i = 1, \dots, p$ be the corresponding coordinates. Then (uniformly in $j = 1, 2, \dots$) we can find functions $f_g^{(1)}, \dots, f_g^{(p)}$, ($g \in G/H$) (i.e. these functions are independent of $j = 1, 2, \dots$) that satisfy*

$$f_g(\xi) \le f_g^{(1)}(\xi_1) \cdots f_g^{(p)}(\xi_p)\,, \qquad g \in G/H,\ \xi = (\xi_1, \dots, \xi_p) \in \hat H\,,$$

$$(4.2.1) \qquad 0 \le f_g^{(i)} \le 1\,, \qquad \int_{\hat H_i} f_g^{(i)}(\xi)\,d\xi \le C e^{c|g|^2}\,,$$

for all $g \in G/H$, $i = 1, 2, \dots, p$.

PROOF. It is clear that $0 \le f_g(\xi) \le 1$ (for all $j = 1, \dots$) and we shall presently show that for all $N \ge 1$ there exists $C, c > 0$ such that (again for $j = 1, \dots$)

$$(4.2.2) \qquad f_g(\xi) \le C\, e^{c|g|^2} |\xi|^{-N}\,, \qquad \xi \in \hat H,\ g \in G/H\,.$$

Let N be so large that there exist $0 \le f^{(i)}(\xi_i) \le 1$, $i = 1, 2, \dots, p$ such that

$$\min\{1, |\xi|^{-N}\} \le f^{(1)}(\xi_1) \cdots f^{(p)}(\xi_p)\,, \qquad \xi = (\xi_1, \dots, \xi_p) \in \hat H\,,$$

$$\int_{\hat H_i} f^{(i)}(\xi)\,d\xi < +\infty\,, \qquad i = 1, 2, \dots, p\,.$$

We can for instance take $f^{(i)}(\xi) = \min\{1, |\xi|^{-2d}\}$ where $d = \sum d_i$ and $N = 2pd$. Then by rescaling we have

$$\min\{1, Ce^{c|g|^2}|\xi|^{-N}\} \le f^{(1)}(Ce^{-c|g|^2}\xi_1)\cdots$$

(with a different C and c) and our lemma follows with $f_g^{(i)}(\xi) = f^{(i)}(Ce^{-c|g|^2}\xi)$.

The estimate (4.2.2) is clearly implied by

$$\int \left| \frac{\partial^k}{\partial x_{i_1}\cdots\partial x_{i_k}} \theta_g(x)\, dx \right| \le C\, e^{c|g|^2}, \tag{4.2.3}$$

$g \in G/H$ uniformly in j, with $k = 1, \dots$ (and $C = C_k$, $c = c_k$). To prove (4.2.3) let $X_1, \dots, X_s$ be a basis of left invariant fields on G. It is clear then, by induction on k, that $(\partial^k/\partial x_{i_1}\cdots\partial x_{i_k})\,\varphi(gx)$ ($x \in H \cong \mathbb{R}^n$, $g \in G$) is a linear combination of expressions of the form $(X_{j_1}\cdots X_{j_r}\varphi)(gx)$ ($x \in H$, $g \in G$) our estimate (4.2.3) therefore follows from results in Section 3.3.

The above proof shows in fact that if the original φ is a strict Gs-function on G then in (4.2.3) (and therefore in (4.2.2) and in (4.2.1)) we can choose $c = \varepsilon > 0$ arbitrary small provided that the $C = C_k$ of (4.2.3) (and the other corresponding C's in (4.2.2) and (4.2.1)) are made to depend on $\varepsilon > 0$. When we are considering several φ_j ($j \ge 1$) the above strict-Gs property can of course to be made uniform in j.

4.3. The disintegration of the kernel.

In this section I shall follows closely [13, I, Section 3]. I shall consider $H \subset G$ as in Section 4.1 with $H \cong \mathbb{R}^a$ (notice that to avoid a possible confusion with notations that I followed, I have changed here the dimensions of H into a: $\dim H = a$ and not n as in the previous sections) and shall assume that $V \cong \mathbb{R}^m$ is a vector subgroup and S is compact. I shall disintegrate μ_j for $j = 1, 2, \dots$

$$\mu_j = \int_{G/H} \nu_{\dot g}^{(j)}\, d\check\mu_j(\dot g),$$

where $\nu_{\dot g}^{(j)}$ are probability measures on the fibers $gH = \dot g \in G/H$ (all the other notations are as before). From this it clearly follows that

$$\mu^{(n)} = \int_{G/H} \cdots \int_{G/H} \nu_{g_1} * \cdots * \nu_{g_n}\, d\check\mu_1(g_1)\cdots d\check\mu_n(g_n), \tag{4.3.1}$$

where the $*$ indicates convolution in G. I shall now identify, as I may, ν_g with a measure on H (by $H \leftrightarrow gH$ up to translation on H), and, for any $\nu \in \mathbb{P}(H)$ and $g \in G/H$, I shall denote by $\nu^g \in \mathbb{P}(H)$ the image of ν by the action $\pi : G/H \to \mathrm{Aut}(H)$ on H (induced by inner automorphisms as in Section 4.1). It is clear then that the integrand of (4.3.1) which, up to translation, can be identified to a measure on the coset $\dot{g}_1 \cdots \dot{g}_n \subset G$, can also be identified up to translation with

$$\nu(g_1, \dots, g_n) = \nu_{g_1}^{s_1} * \nu_{g_2}^{s_2} * \cdots * \nu_{g_n}^{s_n} \in \mathbb{P}(H)\,,$$

where $s_j = g_1 \cdots g_j \in G/H$ and where the convolution product is now taken in H. (This identification is now done for the "right product" identification $H \leftrightarrow Hs_n$). Now the measures $\nu(g_1, \dots, g_n)$ can be identified to a $L^\infty(H)$ functions of H and, since convolution goes by Fourier transforms to pointwise product, we have

$$\|\nu(g_1, \dots, g_n)\|_\infty \le \int_{\hat{H}} f_{g_1}(\pi(s_1)^*\xi) \cdots f_{g_n}(\pi(s_n)^*\xi)\, d\xi\,. \tag{4.3.2}$$

Note that, to simplify notations I have dropped throughout from the ν's and the f's the $j = 1, 2, \dots$ coming from μ_j. To estimate the integral in (4.3.2) I shall first use the decomposition $\hat{H} = \hat{H}_1 \oplus \cdots \oplus \hat{H}_p$ coming from $L_j = \mathrm{Re}\,\lambda$ $(j = 1, \dots, p)$ the real parts of the roots of the representation $\pi : G/H \to GL(H)$ as in Section 4.1. For the above decomposition and with the obvious notation $\xi = (\xi_1, \dots, \xi_p) \in \hat{H}$, I shall apply the Lemma of Section 4.2 and estimate

$$|f_g(\xi)| \le f_g^{(1)}(\xi_1) \cdots f_g^{(p)}(\xi_p)\,, \qquad g \in G/H\,.$$

This estimate will be inserted in the integrand of (4.3.2). It follows that the right hand side of (4.3.2) can be estimated by

$$\inf \int_{\hat{H}} f_{g_{j_1}}^{(1)}(\pi(s_{j_1})^*\xi_1)\, f_{g_{j_2}}^{(2)}(\pi(s_{j_2})^*\xi_2) \cdots f_{g_{j_p}}^{(p)}(\pi(s_{j_p})^*\xi_p)\, d\xi\,,$$

where the infimum is taken over all choices $1 \le j_i \le n$ $(i = 1, \dots, p)$. The integral under the above inf splits in $\hat{H}_1 \oplus \cdots \oplus \hat{H}_p$ and each integral $\int_{\hat{H}_j}$ can be explicitely computed by a change of variable whose determinant is known by (4.1.4).

Let us introduce the following notation $s_j = (b_j, \sigma_j)$, $g_j = (X_j, \tilde{\sigma}_j)$ $\in V \times S$, $(j = 1, \dots, n)$ and for each $g = (u, \sigma) \in G/H$ let us observe that $|g|_{G/H} \approx |u|_V$ = the norm in V (provided that $|g| \geq 1$) let us further denote by

$$A_n(L_i) = \inf_{1\leq j\leq n} \exp\left(c\,|X_j|^2 - d_i L_i(b_j)\right). \tag{4.3.3}$$

With these notations if we combine all the above estimates we obtain

$$\|\nu(g_1,\dots,g_n)\|_\infty \leq C A_n(L_1)\cdots A_n(L_p). \tag{4.3.4}$$

4.4. The probabilistic estimate.

All the notations introduced up to now will be preserved. The $X_j \in V \cong \mathbb{R}^m$ $(j = 1,2,\dots)$ in the definition of $A_n(L_i)$ will be independent (not necessarily equidistributed) random variables such that the corresponding density functions $\mathbb{P}[X_j \in dx] = \psi_j(x)\,dx$ are Gs-functions on $\mathbb{R}^n$ uniformly in $j = 1,2,\dots$. We have then $b_t = X_1 + \cdots + X_t$. In Section B of the appendix we shall prove the estimate

$$\mathbb{E}(A_n(L_1)\cdots A_n(L_p)) = O(\exp(-c\,n^{1/3})) \tag{4.4.1}$$

for some $c > 0$ provided that the real roots $L_1,\dots,L_p$ satisfy the C-condition (*cf.* (4.1.5)). This estimate was proved in [13] when all the X_j's are equidistributed centered Gaussian variables (so that $b_t = b(t) = X_1 + \cdots + X_t \in \mathbb{R}^m$ is brownian motion) and when the constant $c > 0$ appearing in the definition of $A_n(L_i)$ (*cf.* (4.3.3)) is small enough. Here again, if we are prepared to use the fact that for a driftless Laplacian the heat kernel on G is a strict Gs-function, we can suppose that the $c > 0$ in (4.3.3) is as small as we like. In the appendix however we shall prove (4.4.1) without that restriction.

4.5. The proof of the Proposition of Section 4.1.

All our previous notations are preserved. Let $0 \leq \varphi \in C_0^\infty(G)$ and let

$$\sup_{g\in G}\int_H \varphi(hg)\,dh = C_0\,.$$

(I implicitely use here the right identification $H \leftrightarrow Hg$!). Then the basic formula (4.3.1) that expresses $\mu^{(n)}$ as a barycenter of measures sitting in the various cosets Hg gives

$$\langle \mu^{(n)}, \varphi \rangle \leq C_0 \int \|\nu(g_1, \dots, g_n)\|_\infty \, d\check{\mu}_1(g_1) \cdots d\check{\mu}_n(g_n) \,,$$

where by Section 3.3 $\check{\mu}_j \in \mathbb{P}(G/H)$ $(j = 1, \dots)$ are Gs-measures on G/H (uniformly in j). Since $G/H = V \times S$, we can project $\check{\mu}_j$ on V by the canonical $G/H \to V$ and obtain a sequence of probability measures $\psi(x)\,dx$ $(j = 1, 2, \dots)$ on $\mathbb{R}^m$ that are uniformly Gs on $\mathbb{R}^m$. A sequence $X_j \in \mathbb{R}^m$ $(j = 1, 2, \dots)$ of independent random variables can then be defined by $\mathbb{P}[X_j \in dx] = \psi_j(x)\,dx$. The corresponding $A_n(L_i)$ can thus be constructed and because of (4.3.4) we clearly have

$$\begin{aligned} \int \|\nu(g_1, \dots, g_n)\|_\infty \, d\check{\mu}_1(g_1) \cdots d\check{\mu}_n(g_n) \\ \leq C\, \mathbb{E}(A_n(L_1) \cdots A_n(L_p)) \,. \end{aligned} \tag{4.5.1}$$

The estimate (4.1.6) follows from (4.4.1) and (4.5.1).

4.6. The Proposition for an arbitrary soluble group.

In this section I shall prove the following

Proposition. *Let Q be a connected soluble group that satisfies the* C-*condition. Let $\mu_j \in \mathrm{Gs}(Q) \cap \mathbb{P}(Q)$ $(j = 1, 2, \dots)$ uniformly in j and let $\mu^n = \mu_1 * \cdots * \mu_n$. There exists then $c > 0$ such that*

$$\langle \mu^{(n)}, f \rangle = O(\exp(-c\, n^{1/3})) \,, \qquad f \in C_0^\infty(Q) \,.$$

We shall need the following

Lemma. *Let G be an arbitrary connected real Lie group and let $K \subset G$ be some closed subgroup.*

i) *If the conclusion of the proposition is valid for G/K then it is also valid for G.*

ii) *Conversely if we assume that K is compact and assume that the conclusion of the proposition is valid for G it is also valid for G/K.*

The proof of the lemma is evident and will be left to the reader.

The first step in the proof of the proposition is to reduce the proof to the case when the center of $\mathfrak{q}$ is 0. To see this let $\mathfrak{z} \subset \mathfrak{q}$ and let more generally $\mathfrak{z} = \mathfrak{z}_0 \subset \mathfrak{z}_1 \subset \cdots \subset \mathfrak{z}_j \subset \cdots \subset \mathfrak{q}$ be defined inductively by $\mathfrak{z}_j = \pi_j^{-1}$ (the center of $\mathfrak{q}/\mathfrak{z}_{j-1}$) where $\pi_j : \mathfrak{q} \to \mathfrak{q}/\mathfrak{z}_{j-1}$ is the canonical projection. Then clearly $\mathfrak{p} = \cup \mathfrak{z}_j$ (in fact $\mathfrak{z}_{k+1} = \mathfrak{z}_k$ for some k) is a nilpotent ideal and $\mathfrak{p} \subset \mathfrak{n}$ and, by its construction, $\mathfrak{q}/\mathfrak{p}$ has trivial center. An easy composition series argument on $\mathfrak{q}$ shows that $\mathfrak{q}/\mathfrak{p}$ is a C-algebra.

Let $Z_j \subset Q$ be the analytic subgroup that corresponds to $\mathfrak{z}_j$. One easily sees by induction that these are closed subgroups. Indeed quite generally if Z is the analytic subgroup that corresponds to the center of the algebra Z is closed for $\overline{Z}$, its closure, is connected and the subalgebra $\bar{\mathfrak{z}}$ that corresponds to $\overline{Z}$ is central. To make the required reduction therefore it suffices to consider P the analytic subgroup that corresponds to $\mathfrak{p}$ and to consider Q/P. Our reduction then follows from the lemma.

Let now $\theta : \tilde{Q} \to Q$ be the universal covering map and let $\tilde{N} \subset \tilde{Q}$ be the nilradical. Our hypothesis that the center of $\mathfrak{q}$ is trivial implies then that

$$\operatorname{Ker} \theta \cap \tilde{N} = \{e\}\,. \tag{4.6.1}$$

Indeed $\Theta = \tilde{N} \cap \operatorname{Ker} \theta$ is a discrete central subgroup of $\tilde{Q}$ and therefore $\Theta \subset Z_N$, the center of $\tilde{N}$ which can be identified with a vector space $Z_N \cong \mathbb{R}^c$ $(c \geq 1)$. The Ad action induces $\mathrm{Ad} : \tilde{Q} \to GL(Z_N)$ and if we denote by $V_Z \subset Z_N$ the vector subspace generated by Θ we have $\mathrm{Ad}(Q)|_{V_Z} = \mathrm{Id}$. This means that $V_Z \subset \mathfrak{q}$ is central and therefore $V_Z = \{0\}$ by our hypothesis. (4.6.1) follows.

To finish the proof it suffices to make one further reduction. Indeed let $\tilde{N} \subset \tilde{Q}$; $N \subset Q$ be the corresponding closed nilradicals, *i.e.* the analytic subgroups that correspond to the nilradical $\mathfrak{n} \subset \mathfrak{q}$. By (4.6.1) the mapping $\theta_N : \tilde{N} \to N$ is then (1-1), continuous and onto. It therefore is a homeomorphism. N is therefore simply connected and therefore $N_2 \subset N$, the analytic subgroups that correspond to $[\mathfrak{n}, \mathfrak{n}]$, is closed. By our lemma we can reduce the proof of our proposition to the group $Q/N_2 = G$. This new group satisfies (4.1.1) with $H = N/N_2 \cong \tilde{N}/\tilde{N}_2 \cong \mathbb{R}^a$ (by 4.6.1). Indeed $G/H \cong Q/N$ is a homomorphic image of $\tilde{Q}/\tilde{N} \cong \mathbb{R}^d$ and has therefore the required form $G/H \cong \mathbb{R}^m \times \mathbb{T}^b =$

$\mathbb{R}^m \times S$ as in (4.1.1). The condition (4.1.5) is clearly verified by the C-condition on our original group and the proof of our proposition follows from the lemma and the proposition in Section 4.1.

The above proof gives "for free", so to speak, something slightly stronger. What it shows is that the conclusion of the proposition also holds for any amenable group (*i.e.* when $\mathfrak{g}/\mathfrak{q} = \mathfrak{s}$ is a compact semisimple algebra). This proposition can therefore be viewed as a generalization of the results of [13]. Indeed let G be such a group and let Q, $S \subset G$ be the radical and some Levi subgroup respectively. Then since S is compact we can form the canonical semidirect product and the canonical covering map $\pi : Q \leftthreetimes S = \tilde{G} \to G$ which is now an isogeny (*i.e.* $\operatorname{Ker} \pi$ is a finite subgroup). By our lemma again, it suffices to prove the proposition for the group $\tilde{G}$. For the group $\tilde{G} = Q \leftthreetimes S$ if we repeate our previous argument we reduce the proof to the case where $\mathbb{R}^n \cong N = H \subset \tilde{G}$ and $\tilde{G}/N = Q/N \times S = \mathbb{R}^m \times S$ (*cf.* [13]). This completes the proof.

4.7. General Lie groups.

The key to the proof of Theorem B for a general connected real Lie group is to show that with the machinery that we have developed we can give a proof of that theorem for groups of the form $G = Q \leftthreetimes S$ where Q is soluble and connected and where S is semisimple. Indeed for such a G, as we already pointed out (*cf.* Section 2.1), there exists $Z \subset S$ a discrete subgroup that is central in G and of finite index in the center of S. Let $G_1 = G/Z = Q \leftthreetimes S_1 = Q \leftthreetimes (S/Z)$. Then G_1 is a similar group but has the additional property that the center of S_1 is finite. We can therefore write $G_1 = QNAK = RK$ where $NAK = S_1$ is the Iwasawa decomposition of S_1, K is compact and R is soluble.

The proof of Theorem B for the group G_1 is contained in Section 2.4. Indeed if we identify $f \mapsto f * \mu$ with an R-left invariant operator on $X = R \times K$ we see that we have our theorem as long as we can show that (2.4.7) in Section 2.4 holds with $\theta(n) = \exp(-c\, n^{1/3})$. But modulo Section 3.5 this is exactly what was proved in Section 4.6.

To complete the proof of Theorem B for the group G we shall use the following general observation. Let quite generally $\pi : G \to G_1$ be some covering map between two arbitrary Lie groups and let $d\mu(g) = \phi(g)\, dg$ be some Gs-probability measure on G. Let the corresponding image measure be $\mu_1 = \check{\pi}(\mu) \in \mathbb{P}(G_1)$. μ_1 as we saw in Section 3, is a Gs-measure on G_1 and can be written

$$\begin{aligned} &d\mu_1(g_1) = \phi_1(g_1)\, dg_1 \\ &\phi_1(g_1) = \sum_{z\in \operatorname{Ker}\pi} \phi(gz)\,, \qquad g_1 = g \operatorname{Ker}\pi \in G_1\ . \end{aligned} \tag{4.7.1}$$

The obvious observation is that if Theorem B holds for μ_1 on G_1 then it also holds for μ on G. This is because of the amenability of $\operatorname{Ker}\pi \subset G$, which implies that if we denote by $\|\cdot\|_{2\to 2}$ the $f \mapsto f * \cdot$ convolution norms on $L^2(G; d^r g)$ and $L^2(G_1; d^r g_1)$ we have

$$\|\mu\|_{2\to 2} = \|\mu_1\|_{2\to 2}\ . \tag{4.7.2}$$

(4.7.2) is very well known. Observe also that Section 2.4 in fact contains a proof (4.7.2).

Let now G be an arbitrary connected real Lie group not necessarilly of the form $Q \leftthreetimes S$ and let $Q \subset G$ be its radical let further $S \subset G$ be some Levi subgroup that is an analytic but not necessarily closed subgroup of G. It is clear that $Q \cap S$ is a closed subgroup of Q and a central subgroup of S (Indeed $Q \cap S$ is a normal and discrete subgroup of S for the intrinsic Lie topology of S). As already pointed out twice before there exists then $Z_1 \subset Q \cap S$ a discrete central subgroup of G that is of finite index in $Q \cap S$. We shall quotient by Z_1 and obtain $G_1 = G/Z_1$. This group has a Levi decomposition $G_1 = Q_1 S_1$ as before with the additional property that $Q_1 \cap S_1$ is finite. By what was said just above, if we can prove our theorem for G_1 then we also have it for G.

Using the canonical action of S_1 on Q_1 we can then construct the semidirect product $\tilde{G} = Q_1 \leftthreetimes S_1$ where the kernel of the canonical projection $\tilde{G} \to G_1$ is finite. Since we already know that the Theorem holds for $\tilde{G}$ and since the summation in (4.7.1) is finite, it follows that the Theorem holds also for G_1 (here we make essential use of the fact that Theorem can be stated equivalently either as $\mu^{(n)}(e) = O(\cdots)$ or $\mu^{(n)}(g) = O(\cdots)$ for any $g \in G$). The proof of Theorem B is complete.

4.8. The Iwasawa radical revisited.

It is interesting to observe that the techniques of the previous section prove the following

Proposition. *Let G be a connected real Lie group. Then $R \subset G$, the analytic subgroup that corresponds to $\mathfrak{r} = \mathfrak{q} + \mathfrak{n}_S + \mathfrak{a} \subset \mathfrak{g}$, an Iwasawa radical of the Lie algebra (cf. Chapter 1), is closed.*

Indeed let $Q \subset G$ the (closed) radical of G and let $\Sigma \subset G$ some analytic (but not necessarily closed) Levi subgroup. Let $\Sigma = NAK$ be some Iwasawa decomposition of Σ. Then $Z_0 = Q \cap \Sigma \subset \Sigma$ is a discrete (for the intrinsic Lie-group topology of Σ) central subgroup of Σ therefore $Z_0 \subset Z(\Sigma) \subset K$ where $Z(\Sigma)$ is the center of Σ. It follows in particular that $Q \cap AN = \{e\}$. Let us form $Q \leftthreetimes \Sigma = \tilde{G}$ the semidirect product and let $\theta : \tilde{G} \to G$ be the canonical covering map. Let $\pi : \tilde{G} \to \Sigma$ be the canonical projection so that

$$\pi(\operatorname{Ker}\theta) = Z_0 \,.$$

The subgroup $R = QAN \subset G$ is the image by θ of the subgroup $\tilde{R} = Q \leftthreetimes AN \subset \tilde{G}$. Clearly

$$\tilde{R} \cap \operatorname{Ker}\theta = \{e\} \tag{4.8.1}$$

and to show that R is closed it suffices to show that if $k_n \in \operatorname{Ker}\theta$ ($n \geq 1$) is a a sequence that satisfies $d_{\tilde{G}}(k_n, \tilde{R}) \to 0$ then $k_n = e$ for all $n \geq n_0$ large enough. The proof of this is easy. Indeed we have $d_\Sigma(\pi(k_n), AN) \to 0$ and therefore (since $\pi(k_n) \in Z(\Sigma)$ which is a discrete subgroup of Σ) $\pi(k_n) = e$, $n \geq n_0$. Our assertion therefore follows from (4.8.1) and the fact that $\operatorname{Ker}\pi = Q$.

We shall say that the subgroup $R \subset G$ is an Iwasawa radical of G. As we already pointed out for an arbitrary group G we can find $Z \subset G$ some central discrete subgroup such that $G_1 = G/Z$ is such that Q_1 its radical and Σ_1 some Levi subgroup have a finite intersection (*i.e.* $|Q_1 \cap \Sigma_1| < +\infty$). By quotienning further by $Z_1 \subset G_1$ another central discrete subgroup we can obtain $G_2 = G_1/Z_1 = Q_2\Sigma_2$ where Q_2 is the radical of G_2 and Σ_2 is semisimple with finite center. But then clearly $G_2 = RK$ where R is an Iwasawa radical and K is a compact subgroup.

5. The proof of the Theorem (NB) and the lower estimates.

5.1. The proof of Theorem $\mathbf{A}_2$ for a special class of groups.

In this section G will be a real Lie group that can be written in the form $G = R \cdot K$ where the closed subgroup R is a simply connected soluble NC-group and K is a compact subgroup such that $R \cap K = \{e\}$. We shall identify G with $X = R \times K$ as in Section 2.1 and then decompose

$$R = N \leftthreetimes Q$$

as in sections 1.3-1.6 where N is a simply connected nilpotent subgroup and Q is a simply connected R-group. I shall furthermore systematically use the following notation for the "coordinates" in X

$$x = (r, k) = (n, q, k)\,, \tag{5.1}$$

for $x \in X = G$, $r \in R$, $n \in N$, $q \in Q$, $k \in K$. I shall fix $\Delta = -\sum X_j^2$ some driftless Laplacian on G and $\lambda \geq 0$ will denote the corresponding spectral gap. On the space X, I shall consider the semigroup $\hat{T}_t$ defined in Section 2.6 and denote by $\Omega = \{x(t) \in X : \ t > 0\}$ the path space of the corresponding diffusion. For that path space we shall show that the criterion at the end of Section 2.6 holds. This will complete the proof of theorem for the above group. We shall adopt the following notation

$$x(s) = r_s k_s \in G\,, \tag{5.2}$$

$$r_s = \gamma_1\gamma_2\cdots\gamma_s \in R\,, \qquad s \geq 1,\ \gamma_j \in R,\ j = 1, 2, \ldots, s\,, \tag{5.3}$$

where we use group multiplication in both (5.2) and (5.3), but where, unless $K = \{e\}$, the $\gamma_1, \gamma_2, \cdots \in R$ are not independent random variables. As we pointed out in Section 2.8 however if we fix $k^\infty = (k_j)_{j=1}^\infty \in K^\infty$ (some path in K) and condition with respect to that path the variables $\gamma_1, \gamma_2, \ldots$ become independent with uniformly Gaussian densities on R (*cf.* sections 3.2 and 3.5). It follows that under that condition $r_1, r_2, \ldots$ becomes a time inhomogeneous random walk on R.

The following events $A_1, \ldots, B_1, \cdots \subset \Omega$ will now be considered

$$A_s = [|\gamma_j|_G \leq C \log s\,;\ j = 1, 2, \ldots, s]\,, \qquad s = 1, 2, \ldots\,, \tag{5.4}$$

where $C > 0$ will be chosen appropiately at the end. By the above (uniform in j) Gaussian estimate on the variables γ_j we have the following estimate on the conditional expectations (uniformly in k^∞)

$$\mathbb{P}[\sim A_s // k^\infty] \le \exp(-c \log^2 s)\,, \qquad s = 1, 2, \dots\,,$$

where $\sim$ stands as usual, for complement. Therefore we have

$$\mathbb{P}[\sim A_s] \le \exp(-c \log^2 s)\,.$$

Observe that with the notations of (5.1) on the event A_s we have (*cf.* 3.1.4)

$$|q_j|_Q \le c \log s,\ \ |n_j|_N \le C \exp(c \log s) = C s^C, \qquad \gamma_j = (n_j, q_j)\,,$$

for $j = 1, 2, \dots, s$. Let now

$$B_s = [L_k(q_1 \cdots q_j) \le C\,;\ j = 1, \dots, s\ ;\ k = 1, \dots, n]\,. \tag{5.5}$$

Here $C > 0$ and $L_1, \dots, L_n$ are the real roots attached to the semidirect product $N \leftthreetimes Q$ as defined in Section 1.6. The basic fact that follows from Section D in the appendix (*cf.* D.2) is that

$$\mathbb{P}[B_s] \ge c\, s^{-C}, \qquad s = 1, 2, \dots\,,$$

for appropiate constants $C, c > 0$. When the operator Δ is elliptic the analogous even stronger statement (with the continuous time parameter) is a consequence of A(1) which was proved with considerably less cost in Section A of the appendix.

Let us now define the set

$$X_s = \{x = (n, q, k) \in X\ :\ |n|_N \le C s^C\ ,\ |q|_Q \le C s^C\}\,.$$

It is then clear from the above and from Section 1.6 that

$$A_s \cap B_s = \Omega_s \subset [x(s) \in X_s]$$

and therefore that $\mathbb{P}[x(s) \in X_s] \ge c\, s^C$.

On the other hand we clearly have

$$d^r r \otimes dk\text{-measure}[X_s] \le C s^C$$

and therefore our criterion of Section 2.6 is verified and we are done.

The following remark is worth making. We have used here the fact that in our criterion we can use indiscriminately either the $d^r r \otimes dk$ or the $d^\ell r \otimes dk$ measure to measure the set X_s. There is a very simply way to avoid this. Towards that let us define

$$C_s = [|x(s)| \leq C \log s]\,,$$

then clearly by the Gaussian estimate on the Heat kernel on G

$$\mathbb{P}[\sim C_s] \leq C \exp(-c \log^2 s)$$

and

$$|m_G(x)^{\pm 1}|\ ,\ |m_R(x)^{\pm 1}| \leq C s^C\,, \qquad |x| \leq C \log s\,.$$

This means that if we replace X_s by $X_s \cap [x \in X; |x|_G \leq C \log s]$ and Ω_s by $A_s \cap B_s \cap C_s$ we obtain a new X_s that satisfy the criterion as before and that furthermore on these new sets X_s the two measures $d^r r \otimes dk$ and $d^\ell r \otimes dk$ are equivalent up to a constant that grows at most polynomically in s. Because of this it follows that it does not matter which of the two measures we consider.

5.2. General NB-groups.

From the above special case I shall deduce here the lower estimate (0.2) of Theorem A for a general group. Let G be an arbitrary real NB-Lie group and let $\tilde{G} \to G$ be the simply connected cover of G. It clearly is enough to prove the NB-theorem for $\tilde{G}$ for then by the standard local Harnack principle the theorem also holds for G. We have that $\tilde{G} = Q \leftthreetimes S$ where Q is the radical (simply connected) and S is a simply connected semisimple group. By considering $S = NA\tilde{K}$ the Iwasawa decomposition of S we can write then $\tilde{G} = R\tilde{K}$ with $R = QNA$ but where $\tilde{K}$ is not necessarily compact. By general considerations however (*cf.* [15]) there exists $Z \subset \tilde{K}$ a discrete central (in $\tilde{G}$) subgroup such that $K = \tilde{K}/Z$ is compact. We have $\tilde{G}/Z = RK$ and therefore the lower estimate in (0.2) holds for the group $\tilde{G}/Z$. We shall now show how one deduces from this the same lower estimate for $\tilde{G}$ and therefore also for G.

We start with the following definition. Let G be a compactly generated locally compact group and let $H \subset G$ be a closed compactly generated subgroup (*e.g.* G a real Lie group and $H = \Gamma$ some discrete subgroup). As we already pointed out in Section 3.1 for any $h \in H$ the two distances $|h|_H$ and $|h|_G$ are not in general equivalent (we use here the more general notion of $|\cdot|_H$ valid for non connected groups, *cf.* [1]). We shall say that H is a O-distortion subgroup if for all $\Omega \subset H$ neighbourhood of $e \in H$, there exists $C > 0$ such that

$$C^{-1}|h|_G \leq |h|_H \leq C|h|_G\,, \qquad h \in H\backslash\Omega\,.$$

The important thing to observe is that the central subgroup $Z \subset S \subset \tilde{G} = Q \leftthreetimes S$ considered above is a O-distortion subgroup of $\tilde{G}$. This fact is easy to prove and the details were outlined in [15]. The fact that in the lower estimate (0.2) we can pass from $\tilde{G}/Z$ to $\tilde{G}$ is therefore a consequence of the following

Lemma. *Let $\pi : \check{G} \to G$ be a covering map and let $\Gamma = Ker\pi \subset \check{G}$ be a* O*-distortion finitely generated subgroup. Let us further consider $\Delta_0 = -\sum X_j^2$ some driftless sublaplacian on G which can be identified with a sublaplacian on $\check{G}$. Let $\phi_t(g)$, $\check{\phi}_t(g)$ be the corresponding Heat diffusion kernels and let $\lambda = \tilde{\lambda}$ be the corresponding spectral gap as in Section* 0 (*cf.* (4.7.2))*. We have*

$$\phi_t(e) = \sum_{\gamma \in \operatorname{Ker}\pi} \check{\phi}_t(\gamma)\,, \qquad t \geq 0\,, \tag{5.6}$$

and there exists $C > 0$ such that

$$\phi_t(e) \leq C\, \check{\phi}_t^{1/2}(e)\, e^{-\lambda t/2} t^C\,, \qquad t \geq 1\,. \tag{5.7}$$

The reader could observe that Γ is automatically finitely generated but this point is here irrelevant.

Proof. $\Gamma \subset \check{G}$ is a central subgroup it follows therefore that $\check{\phi}_t(\gamma)$ is a positive definite function on Γ and therefore $\check{\phi}_t(\gamma) \leq \check{\phi}_t(e)$, $\gamma \in \Gamma$. We clearly also have $m_G(\gamma) = 1$, $\gamma \in \Gamma$. By ([1, Chapter 9, Section 1]) and the O-distortion property we also have

$$\check{\phi}_t(\gamma) \leq C\, e^{-\lambda t} \exp\Big(-\frac{|\gamma|^2}{ct}\Big)\,, \qquad \gamma \in \Gamma,\ t \geq 1\,,$$

for some $C > 0$, and therefore also

$$\check{\phi}_t(\gamma) \leq C\, \check{\phi}_t^{1/2}(e)\, e^{-\lambda t/2} \exp\Big(-\frac{|\gamma|^2}{ct}\Big)\,, \qquad \gamma \in \Gamma,\ t \geq 1\,,$$

for a different c, $C > 0$. If we apply the summation (5.6), which is trivial to prove (*cf.* (4.7.1)), our estimate (5.7) follows.

REMARK. An adaptation of the above method, with the use of the i'), ii') version of the criterion in Section 2.6, gives the proof of the lower estimate in Theorem A_1). What changes is the geometry and the "section" that is used (*cf.* [13, II]).

If we use the global structure theorem for (not necessarily simply connected) NC-groups from [13, III] we can adapt the method of Section 5.1 to general NB-groups. The Section 5.2 becomes then redundant.

Appendix.

Guide to the appendix.

For the upper estimate of Theorem A_1) and for Theorem B one only needs Section B of this appendix. My advise to the reader in a first reading is to go straight for Section B and simply refer back for the notations.

Sections A.0 and A.1 suffice for the lower estimate of Theorem A_2) in the case when Δ is an elliptic operator. In my mind this should be the next thing that the reader should study. To do this one should also study (or at least believe) Section C. Section C is elementary calculus but a certain amount of ingenuity is already needed. The estimate A(2) in Section A.3 is needed for the lower estimate of Theorem A_1). Had it not been for the non elliptic Laplacians Δ we would stop there and then. The discrete formulation and the discretisation presented in Section D and Section E are only needed to cope with this subelliptic (but not elliptic) situation, and Section F stands towards Section D what Section C stoud for Section A.1. More explicitely for the (non elliptic) lower estimate of Theorem A_2 one needs the first half of Section F, Section E.2 and Section D.1 (*i.e.* D(2) for $p = +\infty$). The property D(1) can be used as an alternative to A(2) for the proof of the lower estimate of Theorems A_1).

Both Section D and Section F are non trivial (in fact they are, technically, quite difficult) and they present an independent interest. I intend to come back in the future to the problems involved in sections D and F and examine them systematically for their own sake. The reader who is not particularly interested in these problems should simply not waste time and energy in these sections. Indeed an alternative approach (more sophisticated, but technically simpler) to the subelliptic Laplacians will be given in a second instalment of this work.

A. The continuous time diffusion.

A.0. Statement of the results.

We shall consider here the space $X = \mathbb{R}^n \times K$ where K is a some compact C^∞-manifold assigned with some smooth non vanishing measure dk. On X we shall further consider D some subelliptic formally self adjoint (with respect to $dx = dz \otimes dk$ where dz is Lebesgue measure on $\mathbb{R}^n$) second order operator with constant term $D1 \equiv 0$. D will be assumed invariant under the left action of $\mathbb{R}^n$ (*cf.* Section 2.7). To simplify notations (and since this is the only case that we shall use) we shall further assume here that K is some compact group and dk is the Haar measure. The general case when K is an arbitrary C^∞-manifold can be treated with identical methods.

Let us denote by $e = (1, 0, \ldots, 0) \in \mathbb{R}^n$ and by $C_\alpha = \{z \in \mathbb{R}^n : \langle z, e\rangle \geq |z| \cos\alpha\}$ $(0 < \alpha < \pi/2)$ the corresponding conical region. Let us further denote by $C_{\lambda,\alpha} = C_\alpha - \lambda e$ $(\lambda > 0)$ the above conical region translated backwards so as to contain the origin $0 \in C_{\lambda,\alpha}$.

We shall now consider the continuous time diffusion on X

$$\Omega = \{x(t) = (z(t), k(t)) \in X = \mathbb{R}^n \times K\,;\ t > 0\}$$

controlled by the differential operator D. I shall denote as usual by $\mathbb{P}_x$ $(x \in X)$ the corresponding probability measure on Ω with $\mathbb{P}_x[x(0) = x] = 1$. We shall show that for any $0 < \alpha < \pi/2$ and $\lambda > 0$ there exists $c > 0$ such that

$$\text{(A.1)} \qquad \mathbb{P}_0[x(s) \in C_{\lambda,\alpha} \times K\,;\ 0 < s < t] \geq t^{-c}\,, \qquad t > c\,,$$

i.e. the diffusion stays in the conical region "polynomially long".

We shall also show that there exist $C > 0$ such that

(A.2)

$$\mathbb{P}_0[|x(s)| \leq M\,;\ 0 < s < t] \geq C^{-1} \exp\Big(-C\,\frac{t}{M^2}\Big)\,, \qquad t, M \geq 1\,,$$

where for $x = (z, k) \in \mathbb{R}^n \times K$ we denote $|x| = |z|$.

Let us observe straight away that $M = t^{1/3}$ in (A.2) gives

$$\mathbb{P}_0[|x(s)| \leq t^{1/3}\,;\; 0 < s < t] \geq C^{-1}\exp(-C\,t^{1/3})\,, \qquad t \geq 1\,.$$

A.1. The differential operator D.

In this section we shall consider the operators $-D = D_R + M + P + D_K$ on $X = \mathbb{R}^n \times K$ as in sections 2.7 and A.0 and preserve all the notations introduced there.

Let $U(x) \in C^\infty(\Omega)$ ($\Omega \subset \mathbb{R}^n$, open) and let $\psi_j(k) \in C^\infty(K)$, ($j = 1, \dots, n$) be arbitrary. I shall denote by $F(x,k) = U(x_1 + \psi_1(k), \dots, x_n + \psi_n(k))$. We have then

$$D_K F = \sum_{j=1}^{n} D_K(\psi_j)\frac{\partial U}{\partial x_j} + \sum_{i,j=1}^{n}\Big(\sum_{\alpha,\beta=1}^{s}\gamma_{\alpha,\beta}(X_\alpha\psi_i)(X_\beta\psi_j)\Big)\frac{\partial^2 U}{\partial x_i \partial x_j}\,,$$

$$MF = 2\sum_{i,j=1}^{n}\Big(\sum_{\alpha=1}^{s} b_{\alpha,i}X_\alpha\psi_j\Big)\frac{\partial^2 U}{\partial x_i\partial x_j}\,,$$

$$PF = \sum_{j=1}^{n}\delta_j\frac{\partial U}{\partial x_j}\,.$$

By (2.7.6) it then follows (this is standard Fredholm theory *cf.* [23], [24], [39] when D is elliptic and the result easily generalizes to subelliptic operators) that for $j = 1, 2, \dots, n$ we can choose $\psi_j \in C^\infty(K)$ so that

$$D_K(\psi_j + c_j) + \delta_j = 0\,, \qquad (c_1, \dots, c_n) \in \mathbb{R}^n\,.$$

With that choice of the ψ's we have therefore $-DF = (LU)(x + \psi)$ where

$$L = \sum_{i,j=1}^{n} R_{ij}(k)\frac{\partial^2}{\partial x_i\partial x_j}\,,$$

$$R_{i,j} = a_{i,j} + \sum_{\alpha,\beta=1}^{s}\gamma_{\alpha,\beta}c_{\alpha,i}c_{\beta,j} + 2\sum_{\alpha=1}^{s} b_{\alpha,i}c_{\alpha,j}\,, \qquad i,j = 1, \dots, n\,,$$

where $c_{\alpha,i} = X_\alpha \psi_i$ ($i = 1, \dots, n$, $\alpha = 1, \dots, s$). Let us use standard matrix notations and set $A = (a_{ij}) \in M_{n\times n}$, $B = (b_{\alpha,i}) \in M_{s\times n}$, $\Gamma = (\gamma_{\alpha,\beta}) \in M_{s\times s}$. The characteristic form of $-D$ then is

$$\mathbf{a} = \begin{pmatrix} A & B^T \\ B & \Gamma \end{pmatrix} \in M_{n+s,n+s} \tag{A.1.1}$$

and the caracteristic form of L is

$$F(\mathbf{a}; C) = A + C^T\Gamma C + B^T C + C^T B\,, \tag{A.1.2}$$

where $C = (c_{\alpha,i}) \in M_{s\times n}$. ($T$ stands for the matrix transposition operator).

The matrix $\mathbf{a}$ in (A.1.1) is non negative (*cf.* Section 2.7). This implies that $F(\mathbf{a}; C) \gg 0$ and therefore in particular if we assume that $\mathbf{a} \gg \varepsilon I$ for some $\varepsilon > 0$ (this is the order relation of symmetric matrices) we also have $F \gg \varepsilon I$. The proof of these facts is elementary linear algebra. Indeed assume first that A, Γ are the identity matrices then $\mathbf{a} \gg 0$ implies that $\Lambda^T \mathbf{a} \Lambda \geq 0$ ($\Lambda^T = (\lambda^T, \mu^T)$, $\lambda \in \mathbb{R}^n$, $\mu \in \mathbb{R}^s$) *i.e.* $\lambda^T\lambda + \lambda^T B^T \mu + \mu^T B\lambda + \mu^T\mu \geq 0$, setting $\mu = -B\lambda$ we obtain the required result $\lambda^T\lambda - \lambda^T B^T B\lambda \geq 0$. In general, by standard perturbation, we can assume that A, Γ are invertible. We set then $C = \Gamma^{-1/2} D$ and obtain

$$F = A - B^T\Gamma^{-1}B + (D^T + B^T\Gamma^{-1/2})(D + \Gamma^{-1/2}B)\,.$$

This means that it suffices to show that $A - B^T\Gamma^{-1}B \gg 0$. Towards that by conjugating $\mathcal{L}^T \mathbf{a}\, \mathcal{L}$ with

$$\mathcal{L} = \begin{pmatrix} A^{-1/2} & 0 \\ 0 & \Gamma^{-1/2} \end{pmatrix}$$

we can reduce the problem to the case where A and Γ are the identity matrices which is the special case that has just been treated.

In terms of our differential operators the above says that L is a second order operator on $\mathbb{R}^n$ with positive characteristic and that L is uniformly elliptic on $\mathbb{R}^n$ if D is uniformly elliptic on X. It is an unfortunate fact that we cannot replace ellipticity by subellipticity in the above statement (example: $n = 1$, $K = \mathbb{T}$, $D = \partial^2/\partial x^2 + \partial^2/\partial\theta^2 \pm 2\cos\theta\, \partial^2/\partial x\partial\theta$, then $\partial^2\psi/\partial\theta^2 = \mp \sin\theta$ and therefore $\partial\psi/\partial\theta = \pm\cos\theta = c(\theta)$ and $L = \sin^2\theta\, \partial^2/\partial x^2$).

One important consequence of the above transformation where we set $U(x) = x_i$ $(i = 1, \dots, n)$, the n coordinate functions, is that the process

$$y(t) = z(t) + \psi(k(t)) \in \mathbb{R}^n$$

is a vector valued martingale. Equivalently this says that for any linear function $U(x) = \sum u_j x_j$ and any $C \in \mathbb{R}^n$ the function $F = U(x + \psi(k) + C)$ is D-harmonic on X.

It is in general impossible to construct explicitely any other D-harmonic function on X. When D is elliptic however it is very easy to give an explicit construction of an important family of D-subharmonic functions F *i.e.* functions that satisfy

$$DF(x) \leq 0\,. \tag{A.1.3}$$

More precisely let $\alpha > 0$ be small and let $\tilde{C} = C_\alpha \times K$ be the conical region in X as defined in Section A.0. Then a subharmonic function F as in (A.1.3) can be constructed to have the following additional properties:

$$F \in C^k(X) \text{ for some (suitably high) } k > 0\,, \qquad F \geq 0\,,$$

$$F \equiv 0 \text{ on } X \backslash \tilde{C}\,, \qquad F \not\equiv 0 \text{ on } \tilde{C}\,, \tag{A.1.4}$$

$$\begin{aligned} &F(x) = O(|x|^A) \text{ for some } A > 0\,,\\ &F(\mu e, e_K) \longrightarrow \infty \text{ as } \mu \to \infty\,,\\ &e_K = \text{ identity in } K\,. \end{aligned} \tag{A.1.5}$$

The construction of F is easy. Indeed we start with $F_{r,k} \in C^{k+10}(\mathbb{R}^n)$ as in Section C and for an appropiate choice of ν, k we set $U(x) = F_{\nu,k}$ in our previous construction. It is clear then that if we choose appropiately the constants $C = (C_1, \dots, C_n) \in \mathbb{R}^n$ the function

$$F(x,k) = U(x_1 + \psi_1 + C_1, \dots, x_n + \psi_n + C_n) \tag{A.1.6}$$

has all the required properties. The role of the constants C is simply to translate the value of the argument (on $\partial C_\alpha \times K$) outside ∂C_α where U is $\equiv 0$.

A.2. The proof of (A.1).

All our notations will be preserved. We shall first prove the following

Lemma. *There exist $C, c > 0$ such that*

$$\mathbb{P}_e[\max_{0<s<t} |z(s)| \geq C\,t] \leq e^{-ct}\,, \qquad t \geq 1\,.$$

In fact for our purposes we only need the following weaker statement

$$\mathbb{P}_e[\max_{0<s<t} |z(s)| \geq c\,t] = O(t^{-A})\,, \qquad t \geq 1\,, \tag{A.2.1}$$

for any $A > 1$. The proof of the lemma is easy. Let us define $T_0 = 0 < T_1 < \cdots$ a sequence of stopping times by

$$T_{j+1} = \inf\{t > T_j \;:\; |z(t) - z(T_j)| \geq 1\}\,.$$

For every $j \geq 0$, $y_j(t) = y(t + T_j) - y(T_j)$, $(t > 0)$, with $y(t)$ as in Section A.1, is then a martingale and it is easy to verify that $S_j(t)$, the S-function of this martingale, satisfies (*cf.* [27, Section 5])

$$\mathbb{E}(\exp(\alpha S_j^2(t \wedge \xi_j))//\mathcal{T}_{T_j}) \leq C\,, \qquad \xi_j = T_{j+1} - T_j,\; t > 0,\; j \geq 0\,,$$

for $\alpha > 0$ small enough. Since clearly by the stochastic integral representation of that martingale we have $c_1 t \leq S_j^2(t) \leq c_2 t$ $(0 < t < \xi_j$, $j \geq 0$ and some $c_1, c_2 > 0)$ it follows that there exists $c > 0$ such that

$$\mathbb{P}[T_{j+1} - T_j = \xi_j > \lambda//\mathcal{T}_{T_j}] = O(e^{-c\lambda})\,, \qquad \mathbb{E}(\xi_j) \geq c > 0\,.$$

One can then use Bernstein's inequality for the sum of independent random variables (*cf.* [22]) which works in this more general context and deduces that

$$\mathbb{P}_e[\max_{0<s<t} |z(s)| \geq c\,n] = \mathbb{P}[\xi_1 + \cdots + \xi_n \leq t]\,, \qquad n \sim t\,,$$

has the correct bound. This proves the lemma.

One can also prove (A.2.1) directly using the Doob maximal theorem on the martingale $y(s)$ $(s \geq 0)$ constructed in Section A.1. This however requires the estimate $\|y(s)\|_p = O(s^{2/3})$ (for p-large enough). This final estimate is a consequence of the Gaussian decay

$$\mathbb{P}[|z(s)| \geq M] = O\Big(\exp\Big(-c\,\frac{M^2}{s}\Big)\Big)$$

which although correct is not trivial to prove [19]. The subellipticity of D is needed for that estimate to hold. In our context the above Gaussian estimate can also be picked up by the corresponding Gaussian estimate on the original group G (*cf.* [1]).

Finally, for yet another approach to prove (A.2.1) one can use $S(t)$ the S-function of the martingale $y(t)$. Using the Itô (stochastic integral) approach of the construction of the diffusion $x(t)$ $(t > 0)$ one sees inmediately that $\|S(t)\|_\infty = O(t^{1/2})$. The only complication here is of course the fact that X is not $\mathbb{R}^M$ but a manifold and the construction has to be done in "patches", *cf.* [28]. The estimate (A.2.1) follows again by the standard martingale inequality $\|S\|_p \approx \|\max\|_p$ (*cf.* [27]). The advantage of this approach is that again no subellipticity is used.

Let now $\tilde{C}'$ and $F(x)$, $x \in X$, be as in A.1 and satisfy the conditions (A.1.4)-(A.1.5). We shall start diffusion at $O = (\lambda e, e_K) \in C \times K$ some large $\lambda > 0$ and denote

$$C_R = \tilde{C} \cap [a \leq |x| \leq R] \subset \mathbb{R}^n\,, \qquad \tau = \tau_R = \inf\{t\ :\ x(t) \in \partial C_R \times K\}\,.$$

The standard submartingale property of the process $\{F(x(t))\ ;\ t > 0\}$ implies then that

$$F(O) \leq \mathbb{E}\{F[x(\tau)]\} \leq C_0 + \mathbb{P}[|x(\tau)| = R]\,R^C\,, \tag{A.2.4}$$

where the $C_0 > 0$ is independent of R and comes from the fact that $x(t)$ could exist at some small $x(\tau)$ on which $F(x(\tau)) > 0$. If we choose $\lambda > 0$ large enough however we are going to have $F(O) > 2\,C_0$ and therefore

$$\mathbb{P}[|x(\tau)| = R] \geq c\,R^{-C}\,, \qquad R \geq 1\,. \tag{A.2.5}$$

Our lemma on the other hand implies

$$\mathbb{P}[\tau < R;\ |x(\tau)| > c\,R] = O(R^{-A}) \tag{A.2.6}$$

for all $A > 0$. If we put (A.2.6) and (A.2.5) together we conclude that

$$\mathbb{P}[\tau > R] \geq c\, R^{-C}\,, \qquad R \geq 1\,.$$

This clearly implies (A.1).

A.3. The proof of (A.2).

The estimate (A.2) can easily be transformed to a standard "1^{rst} eigenvalue" estimate. Indeed let us consider the operator D on the M-ball of X

$$X_M = \{x = (z,k)\ :\ |z| \leq M\}$$

with Dirichlet boundary conditions (*i.e.* we "kill" the diffusion at the boundary) and let $\lambda > 0$ be the first eigenvalue and $0 \leq \psi \in C^\infty(X_M)$, $\|\psi\|_2 = 1$ the corresponding eigenfunction. Then clearly

$$\|e^{-tD}1\|_2 \geq \langle 1, e^{-tD}\psi\rangle = e^{-t\lambda}\int_{X_M}\psi\, dx\,. \tag{A.3.1}$$

Using standard methods we shall presently see that

$$\int \psi\, dx \geq c > 0\,, \qquad \lambda \leq CM^{-2}. \tag{A.3.2}$$

It follows thus that for each $t > 0$ there exists some $x_0 \in X_M$ such that

$$\begin{aligned} Q(x_0,t) &= \mathbb{P}_{x_0}[|x(s)| \leq M\,;\ 0 < s < t] \\ &\geq C\exp\Big(-c\,\frac{t}{M^2}\Big)\,, \qquad t > 0\,. \end{aligned}$$

By the left action of $\mathbb{R}^n$ on X we can assume that $x_0 = (0,k_0) \in K$. To show that we can assume that $x_0 = 0$ we can use the parabolic Harnack estimates that are verified by $Q(x,t)$ (these use the subellipticity of D). Otherwise (without the use of the above Harnack) we have automatically from (A.3.1)

$$\sup_{x_0\in K}\mathbb{P}_{x_0}[|x(s)| \leq M\,;\ 0 < s < t] \geq C\exp\Big(-c\,\frac{t}{M^2}\Big)\,, \qquad M,t > 0\,.$$

The estimate $\lambda \leq CM^{-2}$ in (A.3.2) is easy enough and is an immediate consequence of the fact that the function

$$\varphi(z,k) = (M - |z|)^+\,, \qquad (z,k) \in X\,,$$

appropiately smoothed for $|z|$ near 0 and M satisfies

$$\|\varphi\|_2^2 \approx M^{n+2}, \qquad (D\varphi,\varphi) \approx M^n.$$

The first estimate in (A.3.2) is a trifle more subtle (but also very standard). The subellipticity of D implies (*cf.* [1]) that

$$\text{(A.3.3)} \qquad \|f\|_p \leq C[(Df,f)^{1/2} + \|f\|_2]\,, \qquad f \in C_0^\infty(X)\,,$$

for some $p > 2$. The estimate (A.3.3) applied to ψ implies (since $M \geq 1$) that $\|\psi\|_p \leq C\|\psi\|_2$ which by standard convexity gives the required estimate. A less sophisticated method to see that $\int \psi > C$ is to combine directly the fact $\|\psi\|_2 = 1$ with the Harnack estimate. Subellipticity is again essential for this approach (if $\int \psi \sim 0$, $\|\psi\|_2 = 1$ then there exists $x_0 \in X_M$ such that $\psi(x_0) \gg 0$ also by standard elliptic estimates we may suppose that $\operatorname{dist}\{x_0, \partial X_M\} \geq c_0 > 0$. Harnack applies on ψ and does the rest).

A.4. Gaussian estimates for a Laplacian with a drift.

The Gaussian estimate for the heat kernel of a Laplacian with a drift term $\Delta = -\sum X_j^2 + X_0$ is not quite standard and we shall outline the proof here. The upper estimates are contained in [20] but here the proof does not need the rather difficult technology of [20] and this proof is already implicit in [19,I]. Indeed if $u(t,x)$ is a solution of $\partial/\partial t - \Delta$ then $v(t,x) = u(t, xe^{tX_0})$ is a solution of the (time dependent) evolution equation

$$\frac{\partial}{\partial t} - \sum(\mathrm{Ad}(e^{tX_0})X_j)^2 = 0\,.$$

Let $\{T_{s,t} : 0 < s < t\}$ be the corresponding time inhomogeneous semigroup [29] and let $\varphi \in C^\infty$ be such that $|\nabla_r \varphi| \leq 1$ where ∇_r denotes the gradient of some fixed left invariant Riemmanian structure on G. By the standard argument (*cf.* [19,I]) we then see that

$$\|e^{\lambda\varphi}\, T_{t,s}\, e^{-\lambda\varphi}\|_{2\to 2} \leq \exp(-c\,(t-s)\lambda^2)\,.$$

To give the proof of the upper Gaussian estimate of the corresponding Heat kernel (and of all its derivatives) we simply use the local Harnack principle just as in [19,I;1]. One should simply observe that the canonical distance on G induced by the fields $X_1, \dots, X_n$ and the above Riemmanian distance are "equivalent" at infinity.

It is of independent interest to observe that the following maximal Gaussian estimate also holds (and we do not need subellipticity for this estimate)

$$P_t(\gamma) = \mathbb{P}[\max_{0<s<t} |z(s)| \geq \gamma] \leq C\exp(-c\,\gamma^2/t)\,, \qquad 0 < t < 1,\ \gamma/t > 1\,,$$

and that there here exists an alternative more direct proof of this fact. This proof relies on a standard Laplace transform argument (*cf.* [21]) and the estimate $P_t(1) \leq c\,e^{-c/t}$ $(t > 0)$ which is equivalent to

$$\text{(A.4.1)} \qquad \mathbb{P}[T = \inf[s\,;\ |z(s)| \geq 1] < t] \leq c\,e^{-c/t}, \qquad t > 0\,.$$

This last estimate is non trivial. The only way I know how to prove it is by considering in local coordinates the semimartingale expression of $t^{-1/2}z(s \wedge t \wedge T)$, $(s > 0)$ for fixed $t \leq 1$. It is easy to see then that the S-function of that semimartingale satisfies $\|S\|_\infty \leq C$. This implies (well known: we time change the martingale part of the semimartingale and make it brownian motion) that the maximal function

$$M^* = t^{-1/2} \sup_{0<s<t} |z(s \wedge T)|$$

satisfies $\|\exp(\alpha(M^*)^2)\|_\infty \leq C$ and our estimate (A.4.1) follows from the fact that on the set $[T < t]$ we have $M^* = t^{1/2}$.

The proof of the lower Gaussian estimate (unless the drift is of special form, *cf.* [19]) is as far as I can tell considerably more difficult to prove. The pivot of the proof is the estimate

$$\text{(A.4.2)} \qquad \mathbb{P}_x[d(z(t), y) \leq 10^{-10}] \geq C\exp(-\frac{c}{t})\,,$$

for $0 < t < 1$, $x, y \in G$, $10^{-10} \leq d(x,y) \leq 10^{10}$. This estimate for a Laplacian with a drift is essentially the Varadhan-Ventcel-Freidlin large deviations estimates for the Heat kernel (*cf.* [30]). The details are rather formidable to write out. This has been done in [19, II] (esp. Section 4.3). In that reference the drift had a special form but the proof given there works for a general drift. From (A.4.2) the lower Gaussian estimate follows by standard methods (*e.g.* [19,II], Section 2.4).

B. The large deviation estimate.

In this section we shall preserve all the notations of Section 4.4 and we shall prove the estimate (4.4.1). The proof is done in two steps. The first step consists in modifying

$$\tilde{A}(L_i) = \inf_{0\le j\le n} \exp(-d_i L_i(b_j)) \tag{B.1}$$

where we set $b_0 = 0$ and in showing that

$$\mathbb{E}(\tilde{A}_n(L_1)\cdots\tilde{A}_n(L_p)) = O(\exp(-c\,n^{1/3}))\,. \tag{B.2}$$

The second step consists in deducing (4.4.1) from (B.2).

To simplify notations I shall also assume throughout that $d_i = 1$ $(i = 1,\dots,p)$. At any rate in both (4.4.1) and (B.1) we can also absorb the d_i with the L_i and consider $\tilde{L}_i = d_i L_i$ instead.

PROOF OF THE STEP 1. By the C-condition we can fix $\ell \ge 2$, $1 \le i_1 < \cdots < i_\ell \le p$ and $\alpha_s > 0$ $(1 \le s \le \ell)$ such that

$$\alpha_s L_{i_s} \ne 0,\ 1 \le s \le \ell\,, \qquad \sum_{s=1}^{\ell} \alpha_s L_{i_s} = 0\,.$$

It is then clear from the geometry of the situation that there exists $C > 0$

$$|L_{i_1}(x)| \le C \sum_{s=1}^{\ell} L_{i_s}^{+}(x), \qquad x \in V\,,$$

therefore since

$$\tilde{A}_n(L_1)\cdots\tilde{A}_n(L_p) \le \exp(-\sup_{1\le j\le s} \sum_{s=1}^{\ell} L_{i_s}^{+}(b_j))$$

we conclude that (B.2) will follow as soon as we can prove that for any $a > 0$ we have

$$\mathbb{E}\big(\exp(-a \sup_{1\le j\le n} |L_{i_1}(b_j)|)\big) = O(e^{-c\,n^{1/3}})\,, \tag{B.3}$$

where $c > 0$.

Observe now that $U_k = L_{i_1}(X_k) \in \mathbb{R}$, $(k = 1, 2, \dots)$ is a sequence of real random variables and that the density functions ψ_k of these variables $\mathbb{P}[U_k \in dx] = \psi_k(x)\,dx$ clearly are Gs-functions on $\mathbb{R}$ (by Chapter 3) uniformly in $j = 1, 2, \dots$. Our estimate (B.3) is thus a consequence of the following lemma

Lemma 1. *Let $U_1, U_2, \dots$ be a arbitrary sequence of independent random variables that satisfies the above condition. Let $S_n = U_1 + \cdots + U_n$ and $V_n = \sup_{1\le j\le n} |S_j|$. Then for every $a > 0$ there exists $c > 0$ such that*

$$\mathbb{E}\,(e^{-aV_n}) = O(e^{-cn^{1/3}})\,. \tag{B.4}$$

Lemma 2. *Let $U_1, U_2, \dots$ be as in Lemma 1. Then there exists $\varepsilon > 0$ and an integer n_0 such that*

$$\mathbb{P}\,[|U_1 + U_2 + \cdots + U_{n_0 m^2}| \le 2m] \le 1 - \varepsilon\,, \qquad m = 1, 2, \dots\,.$$

Lemma 1 follows from Lemma 2. Indeed from Lemma 2 it is clear that

$$\mathbb{P}\,[V_{p\,n_0\,m^2} \le m] \le (1-\varepsilon)^p\,, \qquad m, p = 1, 2, \dots\,,$$

and therefore that

$$\mathbb{P}\,[V_n \le m] \le C \exp\Bigl(-c\,\frac{n}{m^2}\Bigr)\,, \qquad n, m = 1, 2, \dots\,,$$

and (B.4) follows by integration.

Proof of Lemma 2. Let $\varphi_k(\xi) = \hat{\psi}_k(\xi)$, $\xi \in \mathbb{R}$, denote the characteristic function of the variable U_k, $k = 1, 2, \dots$. The uniform lower estimate of the Gs-condition $\psi_k(x) \ge C\exp(-cx^2)$ implies that $\psi_k(x) = \alpha G(x) + (1-\alpha)\tilde{\psi}_k(x)$, where $G(x)$ is a Gaussian distribution and $\tilde{\psi}_k$ some other probability distribution and there exist thus $0 < \alpha < 1$, $c > 0$ such that

$$|\varphi_k(\xi)| \le \alpha\exp(-c\,\xi^2) + 1 - \alpha\,, \qquad k = 1, 2, \dots\,. \tag{B.5}$$

Now again the uniform Gs-condition on the ψ_j's implies that $\varphi_k \in C^\infty(\mathbb{R})$ uniformly in k and therefore since

$$\varphi_k(0) = 1,\ |\varphi_k(\xi)| < 1\,, \qquad \xi \ne 0\,,$$

it follows that there exists $c > 0$ such that

$$|\varphi_k(\xi)| \leq 1 - c\,|\xi|^2\,, \qquad |\xi| < c,\ k = 1, 2, \dots\,. \tag{B.6}$$

Putting (B.5) and (B.6) together we conclude that there exists $c, \eta > 0$ such that

$$|\varphi_k(\xi)| \leq \begin{cases} 1-\eta & \text{if } |\xi| \geq 1\,, \\ e^{-c|\xi|^2} & \text{if } |\xi| \leq 1\,, \end{cases} \qquad k = 1, 2, \dots$$

Let now

$$0 \leq \chi, \hat{\chi} \in C_0(\mathbb{R}) \cap L^1(\mathbb{R}), \quad \hat{\chi}(\xi) = \int e^{ix\xi} \chi(x)\, dx\,.$$

Then for any $m = 1, 2, \dots, r > 0$ we have

$$\begin{aligned} &\int \psi_1 * \cdots * \psi_{m^2}(x)\, \chi\Big(\frac{x}{rm}\Big)\, dx \\ &= r\,m \int (\varphi_1 \cdots \varphi_{m^2})(\xi)\, \hat{\chi}(r\,m\,\xi)\, d\xi \\ &\leq r\,m \int e^{-cm^2\xi^2}\, \hat{\chi}(rm\xi)\, d\xi + (1-\eta)^{m^2} \chi(0) \\ &\leq \int e^{-c\xi^2/r} \hat{\chi}(\xi)\, d\xi + (1-\eta)^{m^2} \chi(0) \\ &\leq c\sqrt{r}\, \hat{\chi}(0) + (1-\eta)^{m^2} \chi(0) \leq 1 - \delta\,, \end{aligned} \tag{B.7}$$

where the last estimate holds if r is small enough and m large enough. For an appropiate choice of χ we have $\chi(x) \geq 1$, $|x| \leq 1$ and thus (B.7) gives

$$\mathbb{P}\,[|U_1 + \cdots + U_{m^2}| \leq r\,m] \leq 1 - \delta$$

and Lemma 2 follows.

Proof of step 2. Let us fix $N \geq 1$ (to be chosen later) and denote $I_\alpha = \{\alpha N + 1, \alpha N + 2, \dots, (\alpha+1)N\} \subset \mathbb{N}$, $\alpha = 0, 1, 2, \dots$. Let $Y_\alpha = \inf_{j \in I_\alpha} |X_j|$ and let $j_\alpha \in I_\alpha$ be the first integer $j \in I_\alpha$ for which $|X_j| = Y_\alpha$. Let us further define

$$B_n(L_i) = \inf_{\alpha; \alpha N \leq n+N} \exp(c\,Y_\alpha^2 - L_i(b_{j_\alpha}))\,,$$

$$C_n(L_i) = \inf_{\alpha; \alpha N \leq n+N} \exp(-L_i(b_{j_\alpha}))\,,$$

$$\zeta_n = \sup_{\alpha;\alpha N \le n+3N} \left(\sum_{k\in I_\alpha} |X_k| \right), \qquad \xi_n = \sup_{\alpha;\alpha N\le n+3N} (Y_\alpha)\,.$$

It is then clear that

$$A_{n+N}(L_i) \le B_n(L_i) \le C_n(L_i)\, e^{c\xi_n^2}, \tag{B.8}$$

$$C_n(L_i) \le \tilde{A}_n(L_i)\, e^{C\zeta_n}, \tag{B.9}$$

provided that $C > 0$ is large enough.

By the independence and the Gaussian decay of the variables involved, we see on the other hand that

$$\mathbb{P}\left[\sum_{k\in I_\alpha} |X_k| \ge \lambda \right] \le C \exp(-c\,\lambda^2)\,, \qquad \alpha = 1,2,\dots\,,$$

$$\mathbb{P}\,[\zeta_n \ge \lambda] \le C\, n \exp(-c\,\lambda^2)\,,$$

where C, c are independent of $n, \alpha = 1,2,\dots,$ (but may depend on N) and that

$$\mathbb{P}\,[Y_\alpha \ge \lambda] \le C_0 \exp(-c_0\, N\, \lambda^2)\,, \qquad \alpha = 1,2,\dots\,,$$

$$\mathbb{P}\,[\xi_n \ge \lambda] \le C_0 n\, \exp(-c_0\, N\, \lambda^2)\,,$$

where C_0, c_0 are independent of $n, \alpha, N = 1,2,\dots$. It follows in particular that

$$\|e^{c\zeta_n}\|_p = O(n^{1/p})\,, \qquad 1 \le p \le +\infty\,, \tag{B.10}$$

and that for every given $k > 0$ and $1 \le p < +\infty$ there exists an $N \ge 1$ large enough for which

$$\|e^{k\xi_n^2}\|_p = O(n^{1/p})\,. \tag{B.11}$$

The proof of the step 2 is then a consequence of (B.2), (B.8), (B.9), (B.10), (B.11) and a simple use of Hölder's inequality.

C. The conical function and the Hessian.

Let $e = (1, 0, \ldots, 0) \in \mathbb{R}^n$ and let $r = |x|$, let $\varphi_0 = \theta$ be the latitude with respect to the north pole e and $(\varphi_1, \ldots, \varphi_{n-2})$ be an appropiate set of local coordinates on S_{n-2} (the $n-2$ sphere), so that $(r, \theta, \varphi_1, \ldots, \varphi_{n-2})$ are a set of "polar" coordinates on $\mathbb{R}^n$.

We shall now fix some small $0 < \theta_0$ and define a function $F = F_{\nu,k}$ on $\mathbb{R}^n$. First of all $F \equiv 0$ if $\langle x, e\rangle \leq |x| \cos\theta_0$ *i.e.* $F \equiv 0$ outside the region C_α (*cf.* Section A.0) with $\alpha = \theta_0$. Next we require that $F(x) = r^\nu u(\theta)$ for $x \neq 0$ for some large $\nu = 1, 2, \ldots$ and $0 \leq u(\theta) \in C^k$. The function $u(\theta)$, $0 \leq \theta \leq \pi$ will have the following properties $u(\theta) > 0$ for $-\theta_0 < \theta < \theta_0$; $u(\theta) \equiv 1$ for $-\varepsilon < \theta < \varepsilon$ (for some small ε) and $u(\theta) = (|\theta| - |\theta_0|)^k$ for $|\,|\theta| - |\theta_0|\,|$ small where $k = 2, 4, \ldots$ is an appropiate even integer. In this section we shall analyze the Hessian $H_{\nu,k} = \mathrm{Hess}(F) = (h_{i,j})$.

Quite generally let us denote by $\mathcal{S}$ the set of symmetric real $n \times n$ matrices and by $\mathcal{P} \subset \mathcal{S}$ the cone of non negative matrices. Let $\ell = (\ell_1, \ldots, \ell_n) \in \mathbb{R}^n$, we shall denote by $\ell \otimes \ell = (\ell_i \ell_j) \in \mathcal{P}$. It is clear that any $s \in \mathcal{S}$ can be written

$$s = \sum_{j=1}^{k} \lambda_j^+ \ell_j^+ \otimes \ell_j^+ - \sum_{i=k+1}^{n} \lambda_i^- \ell_i^- \otimes \ell_i^- \,, \tag{C.1}$$

where $\lambda_j^\pm \geq 0$ are the characteristic roots and $\ell_j^\pm$ $(j = 1, 2, \ldots)$ are the corresponding orthonormal set of eigenvectors. We shall finally define the scalar product in $\mathcal{S}$

$$\langle S^{(1)}, S^{(2)} \rangle = \sum_{i,j} S_{ij}^{(1)} S_{ij}^{(2)} \,, \qquad S^{(k)} = (S_{ij}^{(k)}) \in \mathcal{S} \,.$$

The following two notations will be needed

$$\mathcal{P}_a = \{p = (p_{ij}) \in \mathcal{P} \ : a \sum |\mu_j|^2 \geq \sum p_{ij} \mu_i \mu_j \geq a^{-1} \sum |\mu_j|^2 \text{ for all } (\mu_1, \ldots, \mu_n) \in \mathbb{R}^n \} \,, \qquad a \geq 1 \,,$$

$$\mathcal{S}_a = \{ s \in \mathcal{S} \ : \ \langle p, s \rangle \geq 0 \text{ for all } p \in \mathcal{P}_{a^{1/2}} \} \,, \qquad a \geq 1 \,.$$

The connection between the above two definitions is described in the following elementary.

Lemma. *Let $s \in \mathcal{S}$ and $a \geq 1$. Then $s \in \mathcal{S}_a$ if and only if we can write s in* (C.1) *with*

$$\sum \lambda_j^+ \geq a\Big(\sum \lambda_j^-\Big)\,. \tag{C.2}$$

Proof. Indeed $p \in \mathcal{P}_{a^{1/2}}$ if and only if in (C.1) it can be written

$$p = \sum \mu_j \pi_j \otimes \pi_j \tag{C.3}$$

with $a^{-1/2} \leq \mu_j \leq a^{1/2}$ and $\pi_j \in \mathbb{R}^n$ some orthonormal basis. If we bare in mind that $\langle \ell \otimes \ell, \pi \otimes \pi\rangle = \langle \ell, \pi\rangle^2$ for the standard scalar product on $\mathbb{R}^n$ we see that with s as in (C.1) and p as in (C.3) we have

$$\langle s, p\rangle = \sum_{j,\alpha} \lambda_j^+ \mu_\alpha \langle \ell_j^+, \pi_\alpha\rangle^2 - \sum_{j,\alpha} \lambda_j^- \mu_\alpha \langle \ell_j^-, \pi_\alpha\rangle^2. \tag{C.4}$$

We clearly also have

$$\sum_\alpha \langle \ell, \pi_\alpha\rangle^2 = \|\ell\|_2^2\,, \qquad \ell \in \mathbb{R}^n. \tag{C.5}$$

From (C.4) and (C.5) it follows that $\langle s, p\rangle \geq 0$ if (C.2) is verified. This gives the first half of the lemma. To see the opposite direction for $s \in \mathcal{S}$ as in (C.1) it suffices to test the condition $\langle s, p\rangle \geq 0$ on the matrix

$$p = \sum_j a^{-1/2} \ell_j^+ \otimes \ell_j^+ + \sum_j a^{1/2} \ell_j^- \otimes \ell_j^- \in \mathcal{P}_{a^{1/2}}\,.$$

The significance for us of the above notions lies in the following

Lemma. *Let $a \geq 1$ be given. Then there exists $k_0 = k_0(a)$ such that for all $k \geq k_0$ there exists $u(\theta)$, satisfying the conditions of the definition of $F_{\nu,k}$, and $\nu_0, \nu_0(a,u) \geq 3$ such that*

$$H_{\nu,k} = \mathrm{Hess}\,(r^\nu u(\theta)) \in \mathcal{S}_a\,, \qquad \nu \geq \nu_0\,,$$

at every point of $\mathbb{R}^n$.

The first step is to observe that we have

$$\nabla = \Big(\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n}\Big) = A\Big(\frac{\partial}{\partial r}, \frac{1}{r}\frac{\partial}{\partial \varphi_0}, \dots, \frac{1}{r}\frac{\partial}{\partial \varphi_{n-2}}\Big)$$

valid for $r > 0$ and $|\theta-\theta_0| \leq |\theta_0|/2$ and $(\varphi_1, \dots, \varphi_{n-2})$ in some appropiate patch of local coordinates of S_{n-2}. The matrix $A = (a_{ij}(\varphi))$ has C^∞ coefficients that only depend on $(\varphi_0, \dots, \varphi_{n-2})$ and is independent of r. For every $p \in \mathcal{P}$ and the corresponding differential operator P we have therefore

$$\begin{aligned} PF &= \langle p, H_{\nu,k} \rangle \\ &= \sum p_{ij} \frac{\partial^2 F}{\partial x_i \partial x_j} \\ &= s \frac{\partial^2 F}{\partial r^2} + 2 \frac{1}{r} \sum_{i=0}^{n-2} s_i \frac{\partial^2 F}{\partial r \partial \varphi_i} + \frac{1}{r^2} \sum_{i,j=0}^{n-2} s_{i,j} \frac{\partial^2 F}{\partial \varphi_i \partial \varphi_j} + \alpha \frac{1}{r} \frac{\partial F}{\partial r} \\ &\quad + \frac{1}{r^2} \sum_{i=0}^{n-2} \beta \frac{\partial F}{\partial \varphi_i} , \end{aligned}$$

where the coefficients are C^∞ and where the matrix $(s_{ij}\ ;\ i,j = -1, 0, 1, \dots, n-2) \gg 0$ is positive definite (with $s_{-1,-1} = s$, $s_{-1,i} = s_i$). For $F = r^\nu u$ it follows that

$$\begin{aligned} r^{2-\nu} PF &= \nu(\nu-1) s u + 2\nu s_0 u' + s_{0,0} u'' + \nu \alpha u + \beta_0 u' \\ &= \nu^2 (s + O(1/\nu)) u + 2\nu (s_0 + O(1/\nu)) u' + s_{0,0} u'' . \end{aligned} \tag{C.6}$$

Therefore for our special choice of $u(\theta)$ and θ close to θ_0 ($|\theta| < |\theta_0|$) we have

$$\begin{aligned} r^{2-\nu} PF = \nu^2 &(\theta-\theta_0)^k (s + O(1/\nu)) \\ &+ 2k\nu (\theta-\theta_0)^{k-1} (s_0 + O(1/\nu)) \\ &+ k(k-1)(\theta-\theta_0)^{k-2} s_{0,0} . \end{aligned} \tag{C.7}$$

Given $a \geq 1$ it follows that the discriminant in (C.7)

$$D = k^2 (s_0 + O(1/\nu))^2 - k(k-1) s_{0,0} (s + O(1/\nu))$$

is strictly negative for all

$$|\theta - \theta_0| \leq \varepsilon_0 , \quad \nu \geq \nu_0 , \quad k \geq k_0 , \quad p \in \mathcal{P}_{a^{1/2}} ,$$

where $\nu_0 = \nu_0(a)$, $k = k_0(a)$ only depend on a.

Let us fix some $k \geq k_0$ and some $u(\theta)$ that satisfies the conditions of the definition of F (for that k). Once u has been fixed, it follows from (C.6) that there exists $\nu_1 = \nu_1(u) \geq \nu_0$ (that depends on u) such that $PF \geq 0$ for all $|\theta - \theta_0| \geq \varepsilon_0$ and $\nu \geq \nu_1$. This completes the proof of the lemma.

D. The random walk and the martingale.

D.0. Statement of the results.

Let $S_j = X_1 + \cdots + X_j \in \mathbb{R}^n$ denote an n-dimensional random walk where the variables X_j are independent and centered ($\mathbb{E}(X_j) = 0$, $j = 1, \dots$), but not necessarily identically distribuited, and where there exists $a \geq 1$ such that each covariance matrix satisfies

$$\{\mathbb{E}(X_j^\alpha X_j^\beta)\}_{\alpha,\beta=1}^n \in \mathcal{P}_a\,, \qquad j \geq 1\,, \tag{D.0.1}$$

(here we use the notations of Section C and $X_j = (X_j^1, \dots, X_j^n)$ are the coordinates). We shall also assume that for some $2 \leq p \leq +\infty$ we have

$$\|X_j\|_p \leq C\,, \qquad j = 1, 2, \dots\,. \tag{D.0.2}$$

We shall generalize the above setup and consider a vector valued martingale

$$f_j = d_1 + \cdots + d_j \in \mathbb{R}^n, \qquad j = 1, \dots,\ f_0 = 0\,.$$

The conditions we shall impose on the martingale differences will be a natural generalization of (D.0.1) and (D.0.2)

$$\mathbb{E}(d_j^\alpha d_j^\beta // \mathcal{T}_{j-1}) \in \mathcal{P}_a\,, \quad \mathbb{E}(|d_j|^p // \mathcal{T}_{j-1}) \leq C\,, \qquad j = 1, \dots\,,$$

where $\mathcal{T}_1 \subset \mathcal{T}_2 \subset \cdots$ are the σ-fields of the martingale. Let $(f_j\ ;\ j \geq 0)$ be a martingale as above we shall then show that if $p > 2$ there exists $C, c > 0$ such that

$$\mathbb{P}[\sup_{1 \leq j \leq n} |f_j| \leq M] \geq c \exp\Big(-C\,\frac{n}{M^2}\Big)\,, \qquad n, M \geq C\,. \tag{D.1}$$

We shall also show that (if $p \geq p_0$ large enough) for any $0 < \alpha < 1$ and $\lambda > 0$ large enough there exists $\xi > 0$ such that

$$\mathbb{P}[f_j \in C_{\lambda,\alpha}\ ;\ 1 \leq j \leq m] \geq m^{-\xi}\,, \qquad m \geq \xi\,, \tag{D.2}$$

where $C_{\lambda,\alpha}$ is as in Section A.0.

Both the above estimates are very easy to prove when n = dimension = 1. For $n = 1$, (D.2) is well known. We can take $p = 1+\varepsilon > 1$ arbitrary then, and we do not need the extra covariance condition. To see this we shall consider $u_j = F(f_j)$ where $F(x) \equiv 0$, $x \leq -1$, and $F(x) = x + 1$, $x \geq -1$ which is a submartingale and then apply the "optional stopping time theorem" on $f_{j\wedge T}$ where $T = \inf\{j : f_j \leq -1\}$.

The proof of D.1 for $n = 1$ is not very much harder. We set $T = \inf\{j : |f_j| \leq M\}$ and compose $f_j^* = f_{j\wedge T}$ with the function $F_M(x) = F(x/M)$ where $F \in C^\infty$, $0 \leq F \leq 1$, $F(x) \equiv 0$ for $|x| \geq 1$, $F(x) \equiv 1$ for $|x| \leq 1/2$, $F(x) > 0$ for $|x| < 1$ and $F(x) = C(1-|x|)^{10}$ for $1 - |x| \leq 1/10$. Using the Taylor series of F, it is easy to verify that (*cf.* Section D.2 for details; in fact this verification is entirely trivial if $|d_k| \leq 1$)

$$\mathbb{E}[F_M(f_n^*)//\mathcal{T}_{n-1}] \geq e^{-cM^{-2}} F_M(f_{n-1}^*)\,, \qquad n \geq 1\,.$$

If we iterate this for $n, n-1, \ldots$ we see that (D.1) follows at once.

Both (D.1) and (D.2) are false in higher dimensions without the covariance condition in (D.0.1). It is clear why (D.2) brakes down, it suffices in the random walk $S_j = X_1 + \cdots + X_j$ to consider "singular" variables $X_j \in$ hyperplane perpendicular to the axis of C_α. To see why (D.1) brake down when $n = 2$ we start with $r_j = (r_j, 0) \in \mathbb{R}^2$, where $r_j = \pm 1$ are Rademacher variables, and consider $T_j \in SO_2$ so that the last vector ($\in \mathbb{R}^2$) in the following summation

$$f_j = r_1 + T_1(r_2) + T_2(r_3) + \cdots + T_{j-1}(r_j)$$

is orthogonal in $\mathbb{R}^2$ to the sum of the first $j-1$ terms. The rotations T_j can clearly be made $\mathcal{T}_{j-1}$-measurable. We obtain thus a martingale transform (*cf.* [27]) that satisfies $|f_j| \equiv \sqrt{j}$.

REMARK. It is clear that the first condition (D.0.1) is equivalent to

$$a \sum |\lambda_\alpha|^2 \geq \mathbb{E}\Big(\Big| \sum_{\alpha=1}^{n} \lambda_\alpha d_j^\alpha \Big|^2 // \mathcal{T}_{j-1} \Big) \geq a^{-1} \sum |\lambda_\alpha|^2\,.$$

Therefore when the d_j's admit conditional densities

$$\text{(D.0.3)} \qquad \mathbb{P}[d_j \in dy // \mathcal{T}_{j-1}] = d\mu_{j,\omega}(y)\,, \qquad j = 1, 2, \ldots\,,$$

the condition (D.0.1) is equivalent to

$$\text{(D.0.4)} \qquad \Big(\int y_\alpha y_\beta \, d\mu_{j,\omega}(y) \Big)_{\alpha,\beta=1}^{n} \in \mathcal{P}_a\,, \qquad j = 1, 2, \ldots$$

D.1. A general notion of subharmonicity, the Taylor series and the proof of (D.2).

We shall consider

$$\mu_x^{(\alpha)} \in \mathbb{P}(\mathbb{R}^n),\ x \in \mathbb{R}^n,\ \alpha \in A = \text{some index set}\,,$$

a family of centered (*i.e.* $\int_{\mathbb{R}^n} y_i \, d\mu_x^{(\alpha)}(y) = 0,\ x \in \mathbb{R}^n,\ \alpha \in A,\ i = 1, \dots, n$) probability measures and $F \geq 0$ some upper semicontinuous function on $\mathbb{R}^n$. Let further $\mathcal{R} \subset \mathbb{R}^n$ be some open domain. We shall then say that F is subharmonic in $\mathcal{R}$ with respect to the above family if

$$\text{(D.1.1)} \qquad F(x) \leq F * \mu_x^{\alpha}(x)\,, \qquad x \in \mathcal{R},\ \alpha \in A\,.$$

For simplicity in what follow we shall drop the index $\alpha \in A$ and consider only the case when A reduces to the one point set.

The interest for us of the above definition lies in the following considerations. We shall consider the family of measures (D.0.3). These measures in our applications will be given by $d\mu_x(y) = p(y, x)\, dy$ where the p's are as in Section E. We shall furthermore fix F and $\mathcal{R}$ as above that satisfy (D.1.1) and define further

$$\tau = \inf\{n \ :\ f_n \notin \mathcal{R}\}$$

the first exit time of the process $f_n = y(n)$ defined in Section E. What is important for us is then the following

Lemma. *Let $F, \mathcal{R}, \tau$ be as above. Then the process $u_n = F(f_{n\wedge\tau})$ is a submartingale.*

In our applications the martingale f_n will be one of the two martingales $y(n)$ or $y(T_n)$ of Section E and then we deduce that for any starting probability of the diffusion $(x(t), t > 0)$ in Section A the process $u_n = F(f_{n\wedge\tau})$ is a submartingale.

The proof of this lemma is straight forward and was given in [13, II, Section 4], it will therefore be omitted.

We shall now explain a general procedure that allows us to analyze the convolution $F * (\delta - \mu)$ for any $\mu \in \mathbb{P}(\mathbb{R}^n)$ with $\int x \, d\mu(x) = 0$. Towards that we shall use the Taylor development at $x \in \mathbb{R}^n$ and write

$$F * (\mu - \delta)(x) = 2 \sum_{i,j} \int (1 - t)\, F_{i,j}(x + ty)\, y_i y_j \, d\mu(y)\, dt\,,$$

where $F_{i,j} = \partial^2 F/\partial x_i \partial x_j$, we assume here that F is sufficiently differentiable and that all the above integrals ($y \in \mathbb{R}^n,\ 0 < t < 1$) converge absolutely.

We shall modify the above expression as follows

$$2\sum_{i,j}\int (1-t)F_{i,j}(x+ty)\frac{y_i}{|y|}\frac{y_j}{|y|}\,d\nu(y)dt = 2\sum_{i,j}\int F_{i,j}(x+y)\frac{y_i}{|y|}\frac{y_j}{|y|}\,d\lambda(y)$$

where $d\nu(y) = |y|^2 d\mu(y)$ and $d\lambda$ is the image of the measure $(1-t)dt \otimes d\nu$ by the mapping $(t,y) \to ty$. We shall assume throughout that $\int |y|^2 d\mu(y) < +\infty$ so that λ is a bounded measure. Quite generally for any measure $\lambda \in \mathbb{P}(\mathbb{R}^n)$ and any matrix $(h_{ij}) = H$ we shall introduce the notation

$$\lambda \# H(x) = \int h_{ij}(x+y)\,\frac{y_i}{|y|}\frac{y_j}{|y|}\,d\lambda(y)\,.$$

With the above notation we have therefore

$$F * (\mu - \delta)(x) = 2\,\lambda \# \mathrm{Hess}(F)(x)\,,$$

where $\mathrm{Hess}(F) = (F_{ij})$ denotes the Hessian matrix of F.

For the new measure λ we can no longer assert that it is centered and its baricenter $\overline{\lambda} = \int x\, d\lambda$ may not be 0. Indeed this is not in general true even for the measure ν. Let us make the additional hypothesis that

$$E(\mu) = \Big(\int x_i x_j\, d\mu(x)\ ;\ i,j = 1,\ldots,n\Big) \in \mathcal{P}_a\,, \qquad \int |x|^{2+\delta} d\mu(x) \le C\,,$$

for some $C,\ \delta > 0,\ a \ge 1$. It is then easy to see that there exist $0 < \varepsilon \ll 1$ and $R \gg 1$ such that

$$\text{(D.1.2)}\quad \mu\{x\ :\ |x - x_0| \le R - \varepsilon\} \ge \varepsilon \qquad \text{for all } x_0 \in \mathbb{R}^n,\ |x_0| = R\,.$$

Furthermore R, ε only depend on C, δ, a. This means that the measures $\mu_{x_0,R} = \chi_{\{|x-x_0|<R-\varepsilon\}}\mu \le \mu$ ($|x_0| = R$) all satisfy $\|\mu_{x_0,R}\| \ge \varepsilon$. It follows also that the measures $\lambda_{x_0,R}$ that we can associate to $\mu_{x_0,R}$ by the same procedure satisfy $\|\lambda_{x_0,R}\| \ge \varepsilon^3$. In other words, the property (D.1.2) is "inherited" by λ and can be used as a substitute of $\overline{\lambda} = 0$. This point will be used at the end of Section F below.

Let us now give the proof of (D.2) and to make the argument that follows clearer let us assume first that $\mu \in \mathbb{P}(\mathbb{R}^n)$ as above is compactly supported and that its moment matrix satisfies $E(\mu) = (\int x_i x_j \, d\mu(x) \;;\; i,j = 1,\dots,n) \in \mathcal{P}_a$ for some $a > 1$. The condition that we have imposed on the moment matrix is then equivalent to the fact that there exists $\varepsilon > 0$ such that for any $(n-1)$-dimensional subspace $H \subset \mathbb{R}^n$ we have

$$\mu\{x \,:\, \operatorname{dist}(x,H) \geq \varepsilon\} \geq \varepsilon\,. \tag{D.1.3}$$

Since $\operatorname{supp}\mu$ is compact it is easy to see that the measure λ that corresponds to the above μ has the same property (D.1.3) and therefore it follows that the corresponding moment matrix

$$E(\lambda) = \Big(\int x_i x_j \, d\lambda(x) \;;\; i,j = 1,\dots,n\Big) \in \mathcal{P}_b \tag{D.1.4}$$

for some $b > 1$.

We shall apply the above considerations to the function $F(x) = F_{\nu,k}(x)$ of Section C. By the proposition of Section F and (D.1.4) we deduce that for any μ as above and $|x|$ appropiately large we have

$$F * (\delta - \mu)(x) \leq 0\,,$$

and our lemma applies. An easy adaptation of the argument (A.2.4), (A.2.5), (A.2.6) completes then the proof of the assertion (D.2) for the case $p = +\infty$ in (D.0.2).

There are several ways of getting rid of the compactness of the support in (D.1.4) since we shall not need optimal results, let us proceed as follows: Suppose as above that $E(\mu) \in \mathcal{P}_a$ and that $\operatorname{supp}\mu \subset \{|x| \leq R\}$, $R \gg 1$. It is easy to see (*e.g.* by scaling) that $E(\lambda) \in \mathcal{P}_b$ where $b \sim \mathbb{R}^2$ (if a is fixed). Let then μ be an arbitrary measure that is assumed to admit a high enough moment $E_N = \int |x|^N d\mu < +\infty$ $(N \gg 1)$ and let us denote by $\mu_R = \chi_{\{|x|>R\}}\mu$, the part of μ at ∞, and by λ_R the measure that corresponds to μ_R we have then $E(\lambda_R) = O(R^{-\alpha})$ for an arbitrary large α (provided that N is high enough). We can therefore correct the contribution of λ coming from $\mu^R = \chi_{\{|x|<R\}}\mu$ by $O(R^{-\alpha})$ and obtain that $E(\lambda) \in \mathcal{P}_b$ with $b^{-1} \sim R^{-2} + O(R^{-\alpha})$. For R large enough we obtain thus again (D.1.4) for some b that only depends on a, N and E_N. Working out the exact value of N is not so hard and that exact value of N is not so large either.

The proof of (D.2) for general values of $p < +\infty$ in (D.0.2) can then be completed as before except that we now have to use the second half of Section F. For our applications however the case $p < +\infty$ is not essential. Indeed if we use the martingale $y(T_n)$ constructed in Section E.2 then the supports of all the measures involved are (uniformly) compact and $p = +\infty$.

D.2. The radial function and the proof of (D.1) (The case $p = +\infty$ in (D.0.2)).

Let $0 \leq \varphi \in C^N(\mathbb{R}^n)$ be radial decreasing (*i.e.* $\varphi(x) = \varphi(r)$, $r = |x|$ and N is sufficiently large) and such that $\varphi \equiv 1$ if $|x| < 1/2$ and $\varphi \equiv 0$ if $|x| \geq 1$, $\varphi > 0$ if $|x| < 1$. Let us further assume that $\varphi(x) = (1-r)^\nu$ for $3/4 \leq r \leq 1$ and $\nu = 4, 6, \dots$ some appropiately large even integer. The above function is clearly not convex (if $n \geq 2$ it is not even convex in some Nhd of the unit sphere $r = 1$). Let $\mathrm{Hess}(\varphi) = (\partial^2\varphi/\partial x_i \partial x_j)$ be the corresponding Hessian matrix. This matrix can easily be diagonalized and an easy calculation shows that for $|x| \sim 1$ we have

$$\mathrm{Hess}\,\varphi = \nu(\nu-1)(1-r)^{\nu-2}\rho\otimes\rho - c\,\nu(1-r)^{\nu-1}\theta_j\otimes\theta_j\ ,$$

where $\rho(x)$ is the unit vector along the radius $\overline{Ox}$ and $\theta_j(x)$ ($j = 1, \dots, n-1$) are an orthonormal complement of $\rho(x)$ (tangent to the sphere $\{y : |y| = |x|\}$.

The crucial fact in the structure of the above Hessian is that for each $r_0 < 1$ if we add some appropiately large multiple of φ we obtain a positive matrix

$$\mathrm{Hess}\,\varphi + C\,\varphi\, I \gg 0\,, \qquad |x| < r_0\ .$$

By scaling therefore $\varphi_M(\cdot) = \varphi(\cdot/M)$ ($M > 0$) we obtain

$$\mathrm{Hess}\,\varphi_M + CM^{-2}\varphi_M I \gg 0\,, \qquad |x| < M\,r_0\ .$$

If we use the second order Taylor development of φ_M we obtain therefore that

$$\varphi_M * (\mu - \delta)(x) \geq -CM^{-2}\int \varphi_M(x+y)\,d\lambda(y)\,, \qquad |x| < M\,r_0\ ,$$

where μ and λ are as in Section (D.1) and are compactly supported since $p = +\infty$, and μ satisfies [(D.1.3) of Section D.1]. Then ν and λ

also satisfy (D.1.3) and since $\varphi_M(r)$ is decreasing it follows that there exist $M_0, c_0 > 0$ such that

$$\text{(D.2.1)}\quad \varphi_M*(\mu-\delta)(x) \geq -c_0\, M^{-2}\varphi_M(x)\,, \qquad M \geq M_0\,,\ |x| < M\, r_0\,.$$

The next observation is that for all $a > 1$ there exists ε such that if $1-r_1 < \varepsilon$ then $\operatorname{Hess}\varphi \in \mathcal{S}_a$, $|x| > r_1$ therefore also $\operatorname{Hess}\varphi_M \in \mathcal{S}_a$ ($M > 0$, $|x| > r_1 M$). In informal terms this says that, near the boundary, φ "looks more and more like a convex function". In fact by an elementary calculation, that is best carried out by drawing a few pictures, we see that (D.2.1) holds (with $c_0 = 0$) for $|x| \geq r_1 M$ provided that $1 - r_1$ is small enough and M large enough. The final conclusion is therefore that (D.2.1) holds for all $x \in \mathbb{R}^n$.

From the estimate (D.2.1) we deduce that there exist $M_1, c > 0$ such that

$$\text{(D.2.2)}\qquad \varphi_M * \mu(x) \geq e^{-cM^{-2}}\varphi_M(x)\,, \qquad x \in \mathbb{R}^n,\ M \geq M_1\,.$$

To finish the proof of (D.1) let us set $T = \inf\{j \,:\, f_j \geq M\}$, $f_j^* = f_{j\wedge T}$ and let us apply (D.2.2) with $\mu = \mu_{j,x}$ as in (D.0.3). We obtain

$$\mathbb{E}[\varphi_M(f_j^*)//\mathcal{F}_{j-1}] \geq e^{-cM^{-2}}\varphi_M(f_{j-1}^*)$$

which by iteration gives

$$\text{(D.2.3)}\qquad \mathbb{E}[\varphi_M(f_n^*)] \geq e^{-cn/M^2}\,, \qquad n \geq 1,\ M \geq M_1\,.$$

From (D.2.3) (D.1) follows at once. It has thus been shown that (D.1) holds if $p = +\infty$ in (D.0.2). The above argument can be adapted to deal with $p < +\infty$, the proof will be omitted since this is not essential for us.

E. Discretising the continuous time martingale.

E.1. The deterministic discretisation $t = 1, 2, \ldots$.

We shall preserve all the notations of Section A and recall (Section A.1) that $y(t) = z(t) + \psi(k(t)) + C \in \mathbb{R}^n$ is a continuous time martingale. It follows in particular that $f_j = y(j)$ ($j \geq 0$) is a discrete time martingale (unlike Section D.0 $f_0 = C$ which is not necessarily 0) our purpose is to examine

$$\mathbb{P}\,[f_j - f_{j-1} \in dy // f_1 = x_1, \ldots, f_{j-1} = x_{j-1}] = p_j(y;x)\,dy\,,$$

for $x = (x_1, \ldots, x_{j-1})$, $y \in \mathbb{R}^n$. To be formally correct the above probability density is only defined x almost surely and is a measure $d\mu_x(y)$. The above abusive notation will be justified by what follows. What is clear by the martingale property is that

$$\int y\, d\mu_x(y) = 0\,, \qquad j = 0, 1, \ldots,\ x \in \mathbb{R}^n\,.$$

We shall show then that there exist $C, C_1 > 0$ such that we have (uniformly in x)

$$\text{(E.1.1)}\quad C_1^{-1}\exp(-C|y|^2) \le p_j(y,x) \le C_1 \exp\Big(-\frac{|y|^2}{C}\Big)\,, \qquad y \in \mathbb{R}^n\,.$$

It is essentially this estimate that justifies our previous abusive notation. It is clear that it suffices to prove the same Gaussian estimates for the "finer" conditional probabilities with respect to the fields $\mathcal{T}_j = \mathcal{T}\{z(t)\,;\, t \le j-1\}$. By the Markov property we must therefore consider the conditional properties

$$\mathbb{P}\,[f_j - f_{j-1} \in dy // (z(j-1), k(j-1)) = x] = \tilde{p}_j(y;x)\,dy\,,$$

for $y \in \mathbb{R}^n$, $x = (z,k) \in X = \mathbb{R}^n \times K$. These new Gaussian estimates can be deduced from the Gaussian estimates for the diffusion kernel $q_t(x_1, x_2)$, ($t > 0$, $x_i = (z_i, k_i) \in X$, $i = 1, 2$) of the diffusion Ω (*cf.* Section A.0)

$$\text{(E.1.2)}\quad C_1^{-1}\exp(-C|z_1 - z_2|^2) \le q_1(x_1, x_2) \le C_1 \exp\Big(-\frac{|z_1 - z_2|^2}{C}\Big)\,.$$

To deduce (E.1.1) from (E.1.2) one simply "integrates" along the fibers

$$F_y = \{(z,k) \in X \;:\; z + \psi(k) = y\}\,.$$

The upper Gaussian estimate (E.1.2) is perfectly standard and follows from the more general (C^∞-manifold) upper Gaussian estimates for subelliptic operators and the intrinsic distance that they induce (*cf.*

[1]). The lower estimates makes essential use of the left invariance of D and the corresponding scaled (for "small balls" but spacially uniform) Harnack estimate. The argument is an easy adaptation of [19]. Alternatively, if the reader is not prepared to either believe or verify for himself the above argument, he could refer to [1] where the above lower estimate is explicitely proved for Lie groups G and left invariant operators. The diffusion Ω that we will be considering is non other than the diffusion that in Section 2 is induced on $R \times K$ by the corresponding diffusion in our original group G. The lower Gaussian estimate (E.1.2) can then easily be picked up by the corresponding estimate in that group. The verification will be left to the reader.

The reader should also observe that the above lower Gaussian estimate is not essential for us here. Indeed the reason that we need these estimates is that we have to show that the above martingale $f_j = y(j)$ satisfies the conditions of Section D.0. For this it suffices to have the upper Gaussian estimate (E.1.1) which guarantees the moment condition (D.0.2) and a much weaker lower estimate of the form

$$\tilde{p}_j(y, x) \geq \varepsilon\,, \qquad |y| < \varepsilon\,,$$

for some $\varepsilon > 0$. This is guaranteed by the uniform Harnack estimate on X for the operator D.

E.2. The optional time discretisation.

There is an alternative way to discretise the time parameter of the martingale $y(t)$, $(t > 0)$. Let $T_0 = 0$ and

$$T_1 = \inf\{t/|z(t) - z(0)| \geq C\}\,, \quad T_j = \inf\{t/|z(t) - z(T_{j-1})| \geq C\}\,,$$

for $j = 2, 3, \dots$, and some large $C > 0$. We can set then $f_j = y(T_j)$, $(j = 1, 2, \dots)$ which is now a martingale as in Section D.0 with the additional property that the martingale differences $d_j = f_j - f_{j-1} \in L^\infty$ are uniformly bounded. For this new martingale we shall define again

$$\mathbb{P}\,[f_j - f_{j-1} \in dy//x(T_{j-1}) = (z(T_{j-1}), k(T_{j-1})) = x] = d\tilde{\mu}_x(y)\,,$$

for $j = 1, 2, \dots$, $x \in X = \mathbb{R}^n \times K$, and

$$\mathbb{P}\,[f_j - f_{j-1} \in dy//f_1 = x_1, \dots, f_{j-1} = x_{j-1}] = d\mu_x(y)\,,$$

for $j = 1, 2, \ldots, x_1, \cdots \in \mathbb{R}^n$, and we shall show that the measures μ_x satisfy the covariance condition of Section D.0 uniformly in $j \geq 1$ and $x \in \mathbb{R}^n$. For this it clearly suffices to prove the corresponding covariance condition (D.0.4) for the measures $\tilde{\mu}_x$. To see this fact one simply has to understand what a subelliptic diffusion means. The best way to analyze this situation is to work with the "trajectories" of the diffusion Ω (*cf.* Section A.0).

Indeed if $x, y \in \Theta \subset X$ where Θ is some open set of X then

$$\begin{aligned}\mathbb{P}_0[x(T_{j-1}+t)\in\Theta,\ 0<t<t_0;\ \operatorname{dist}(x(T_{j-1}+t_0),y)<\varepsilon//x(T_{j-1})=x]\\ =\mathbb{P}_x[x(t)\in\Theta,\ 0<t<t_0;\ \operatorname{dist}(x(t_0),y)<\varepsilon]>0\end{aligned}$$

for any t_0 and $\varepsilon > 0$. This is a basic consequence of the subellipticity of the operator D and follows from the smoothness of the heat diffusion kernel and elementary (if lenghty and tedious) considerations that will be left for reader.

For fixed x and j therefore, by appropiately chosing t_0 and y we see that measure $\mu_x(dy)$ charges positively (and in and uniform fashion with respect to x and j) a whole family of small discs around x. Furthermore there are enough of these discs on every direction as we go away from x to guarantee the covariance condition (D.0.4) for $\tilde{\mu}_x$. The details will be left to the reader.

F. The geometry of the Hessian.

F.1. Dimension $= n = 2$.

To see clearly what is involved we shall first consider the case of $\mathbb{R}^2$ (*i.e.* $n =$ dimension $= 2$). We shall preserve all the notations of the previous sections and translate the θ-variable (now of course the polar coordinates are $(x, y) = (r, \theta) \in \mathbb{R}^2$) by $\theta_0 = \alpha$ so that the x-axis becomes one of the two edges of the wedge C_α, and $\{x \geq 0,\ y = 0\} = \{r \geq 0,\ \theta = 0\}$.

For these coordinates we have

$$\frac{\partial}{\partial x} = \cos\theta\,\frac{\partial}{\partial r} - \sin\theta\,\frac{1}{r}\frac{\partial}{\partial\theta}, \frac{\partial}{\partial y} = \sin\theta\,\frac{\partial}{\partial r} + \cos\theta\,\frac{1}{r}\frac{\partial}{\partial\theta}\,.$$

For $\theta \sim 0$ ($\theta > 0$) and $F = F_{\nu,k}$ as in Section C we obtain by a straight forward calculation that

$$
\begin{aligned}
\operatorname{Hess} F &= r^{\nu-2}\theta^{k-2}\begin{pmatrix} \nu^2\theta^2 & \nu k\theta+\nu^2\theta^3 \\ \nu k\theta+\nu^2\theta^3 & k^2+(2\nu k+\nu-k)\theta^2+\nu^2\theta^4 \end{pmatrix} \\
&= r^{\nu-2}\theta^{k-2}\mathcal{H}\,,
\end{aligned}
\tag{F.1.1}
$$

where each coefficient of the matrix $\mathcal{H}$ has to be multiplied in addition by a factor $= 1+O(\theta^2)+O(1/k)+O(k/\nu)$, and in the considerations where $\mathcal{H}$ is used we shall assume throughout that $0<\theta<\varepsilon_0\ll 1$, $1\ll k$, $1\ll \nu/k$.

For a θ that it is not close to 0 and $F=r^\nu u(\theta)$ we also have the following expression of the Hessian

$$
\begin{aligned}
\operatorname{Hess} F &= \nu\,(\nu-1)\,r^{\nu-2}u(\theta)\begin{pmatrix} \cos^2\theta & \cos\theta\sin\theta \\ \cos\theta\sin\theta & \sin^2\theta \end{pmatrix} \\
&= \nu\,(\nu-1)\,r^{\nu-2}u(\theta)\tilde{\mathcal{H}}\,,
\end{aligned}
\tag{F.1.2}
$$

where every coefficient of the matrix $\tilde{\mathcal{H}}$ has to be multiplied by a factor of the form $(1+O(1/\nu))$ with a $O(\cdot)$ that depends of course on the particular chice of u.

Let now quite generally $K=(k_{i,j}(x))\in\mathcal{S}$, $x\in\mathbb{R}^n$, denote an arbitrary matrix and let $\mu\geq 0$ denote some non negative measure on $\mathbb{R}^n$. We shall use then the same notation as in Section D.1

$$
(\mu\#K)(x)=\sum_{i,j}\int k_{ij}(x+z)\,\frac{z_i}{|z|}\,\frac{z_j}{|z|}\,d\mu(z)\,,
$$

where we shall assume that all the above integrals converge absolutely. I shall denote

$$
E_{ij}=\int x_i x_j\,d\mu(x)\,.
$$

We have then

Proposition. *Let μ be as above and let us assume that $E=(E_{ij})\in\mathcal{P}_b$ for some $b>1$ and that μ is a probability measure supported in the unit ball:* $\operatorname{supp}\mu\subset\{x\in\mathbb{R}^n\ :\ |x|\leq 1\}$. *Then there exists a choice of ν,k and some $r_0>0$ such that*

$$
(\mu\#\operatorname{Hess} F_{\nu,k})(x)\geq 0\,,\qquad x\in\mathbb{R}^n,\ |x|\geq r_0\,.
\tag{F.1.3}
$$

The proof is elementary but lenghty. Before we give the proof we shall explicitely state the three basic properties of Hess F that make the things work. The proof of this properties will be left to the reader.

a) Using the notation of Section C.1 we shall diagonalize the matrix $\mathcal{H}$ at every $z \in \mathbb{R}^2$

$$\mathcal{H} = \lambda_1 e_1 \otimes e_1 + \lambda_2 e_2 \otimes e_2 \,,$$

where (e_1, e_2) is an orthonormal basis of $\mathbb{R}^2$ and where $\lambda_1 \geq 1/2$ provided that $0 < \theta < \varepsilon_0$ for ε_0 small enough and $k, \nu/k$ are large enough. Furthermore by direct computations or by the considerations of Section C we see that for all $a > 1$ there exists t_0, k_0 such that $\lambda_1 \geq a|\lambda_2|$ for $\nu/k \geq t_0$, $k \geq k_0$.

b) Let us assume that ν and k are fixed, then $\lambda_i(z)$ and $e_i(z)$ are continuous functions of θ. By the uniform continuity and the fact that $\theta \lesssim y/|z|$ we see therefore that for all $\delta > 0$ there exists $r_0 = r_0(\delta; \nu, k)$ such that

$$|\lambda_i(z_1) - \lambda_i(z_2)| \leq \delta\,, \quad |e_i(z_1) - e_i(z_2)| \leq \delta\,, \qquad |z_1 - z_2| \leq 10,\ |z_i| \geq r_0\,.$$

c) Let us again fix ν and k. Near $\theta = 0$ we have

$$\mu\#\mathrm{Hess}\, F_{\nu,k} = \mu\# r^{\nu-2}\theta^{k-2}\mathcal{H}|_{\theta=0} + \mathrm{Error}\,. \tag{F.1.4}$$

Let $A > 0$ be fixed for $z_1 = (x_1, y_1)$ with $|y_1| \leq A$ and x_1 large enough. We have

$$\begin{aligned}(\mu\# r^{\nu-2}\theta^{k-2}\mathcal{H}|_{\theta=0})(z_1) &= k^2 \int r^{\nu-2}\theta^{k-2}\,\frac{y^2}{|z|^2}\,d\mu(z)\\ &= x_1^{\nu-2-k}(c_0 + O(1/x_1))\,,\end{aligned}$$

where $c_0 > 0$. To prove that $c_0 \neq 0$ we use already the hypothesis $E \in \mathcal{P}_b$. By an easy calculation on the other hand one sees that the "error" in (F.1.4) is $O(x_1^{\nu-3-k})$ because all but the k^2 terms of $\mathcal{H}$ involve higher powers of θ.

The conclusion is that for fixed ν, k and A we can find $B > 0$ such that our estimate (F.1.3) holds in the region

$$R_{A,B} = \{z = (x, y)\ :\ x \geq B,\ y \leq A\}\,.$$

What is furthermore important about the region $R_{A,B}$ is that (again with ν, k fixed) for any $\delta > 0$ we can find A and r_0 large enough so that the co-factor of $\mathcal{H}$ in (F.1.1) $\varphi = \varphi(z) = r^{\nu-2}\theta^{k-2}$ satisfies

$$1 - \delta \leq \frac{\varphi(z_1)}{\varphi(z_2)} \leq 1 + \delta\,, \quad |z_1 - z_2| \leq 1\,, \qquad |z_1| \geq r_0,\ z_1 \notin R_{A,r_0}\,.$$

In other words that cofactor varies as slowly as we like outside the regions $R_{A,B}$ for large $|z|$.

We shall now procede with the proof of the proposition. Because of c) above, it remains to prove (F.1.3) in the region outside R_A, r_0. With the notations already introduced we have then

$$\mu\#\mathrm{Hess}\, F_{\nu,k}(z_0) = \int \psi(z_0; z)\, d\mu(z - z_0)\,,$$

where

$$\psi(z_0, z) = \varphi(z)\Big(\lambda_1(z)\Big\langle e_1(z), \frac{z - z_0}{|z - z_0|}\Big\rangle^2 + \lambda_2(z)\Big\langle e_2(z), \frac{z - z_0}{|z - z_0|}\Big\rangle^2\Big)$$

and where z_0 lies in the region $0 < \theta < \varepsilon_0$.

For any $\varepsilon \geq 0$ let us denote

$$B_\varepsilon(z_0) = \{z\ :\ |z - z_0| \leq 1,\ |\langle e_1(z_0), z - z_0\rangle| \geq \varepsilon|z - z_0|\}$$

so that B_0 is the unit ball centered at z_0. Let us now fix $\varepsilon > 0$, we can then find t_0, k_0 so that

$$\lambda_1(z)\,\Big\langle e_1(z_0), \frac{z - z_0}{|z - z_0|}\Big\rangle^2 \geq 10^{100}\,|\lambda_2(z)|\,\Big\langle e_2(z_0), \frac{z - z_0}{|z - z_0|}\Big\rangle^2,$$

for $\nu/k \geq t_0$, $k \geq k_0$, $z \in B_\varepsilon$. Observe that here the argument z has been frozen to $z = z_0$ in $e_i(\cdot)$. We shall fix $k = k_0$ and will not change it anymore. Using b) for every $\nu \geq \nu_0(k_0)$ we can find $r_0 = r_0(\nu)$ such that

$$\lambda_1(z)\,\Big\langle e_1(z), \frac{z - z_0}{|z - z_0|}\Big\rangle^2 \geq 10^{10}\,|\lambda_2(z')|\,\Big\langle e_2(z'), \frac{z' - z_0}{|z' - z_0|}\Big\rangle^2,$$

for $\nu \geq \nu_0$, $|z - z'| \leq 10$, $|z| \geq r_0(\nu)$, $z \in B_\varepsilon$. Using c) and the slow variation of φ outside R_A, r_0 we see that there exists $A > 0$ such that for all $\nu \geq \nu_0$ large enough there exists $r_1(\nu) > 0$ such that

$$\text{(F.1.5)}\qquad \begin{aligned}&\varphi(z)\,\lambda_1(z)\,\Big\langle e_1(z), \frac{z - z_0}{|z - z_0|}\Big\rangle^2 \\ &\qquad \geq 10^5\,\varphi(z')\,|\lambda_2(z')|\,\Big\langle e_2(z'), \frac{z' - z_0}{|z' - z_0|}\Big\rangle^2,\end{aligned}$$

for $|z-z'| \leq 10$, $z=(x,y)\in\mathbb{R}^2$, $|z|\geq r_1(\nu)$, $|y|>A$, $z\in B_\varepsilon$. At this point one should remember that we are working close to the edge of C_α *i.e.* in a range $0<\theta<\varepsilon_0$ for some small ε_0. The estimate (F.1.5) will now be used in conjunction with the fact that because of our hypothesis $E\in\mathcal{P}_b$, for $\varepsilon>0$ small enough, we have $\mu(B_0\backslash B_\varepsilon)\leq 10^{-10}$ (*cf.* Section D.1). If we integrate (F.1.5) against $d\mu(z-z_0)\,d\mu(z'-z_0)$ we obtain that for all $\nu\geq\nu_0$ there exists $r_0=r_0(\nu)$ such that

$$\text{(F.1.6)}\qquad \mu\#\text{Hess}\,F_{\nu,k}\geq 10^{-2}\varphi(z_1)\,\lambda_1(z_1)\left\langle e_1(z_1),\frac{z_1-z_0}{|z_1-z_0|}\right\rangle^2,$$

for $\nu\geq\nu_0$, $z_i=(x_i,y_i)\in\mathbb{R}^2$, $i=0,1$, $|z_0-z_1|\leq 1$, $z_1\in B_0$, $|z_i|\geq r_0(\nu)$, $y_i\geq 2A+10$. In particular (F.1.6) holds for $z_1=z_0$ which together with c) shows that for $k=k_0$ and every $\nu\geq\nu_0$ there exists $r_0=r_0(\nu)$ such that

$$\text{(F.1.7)}\quad \mu\#\text{Hess}\,F_{\nu,k_0}(z)\geq 0\,,\qquad \nu\geq\nu_0,\ 0<\theta<\varepsilon_0,\ |z|\geq r_0(\nu)\,.$$

To finish the proof of the proposition, since $k=k_0$ has been fixed, we shall complete the definition of $u(\theta)$ in $F_{\nu,k}=r^\nu u(\theta)$ and use the formula (F.1.2) for the Hessian. Using that formula and the same method (this method now applies much easier. Indeed we do not have the edge, where the co-factor θ^{k-2} vanishes, to worry about!) we finally see again that there exists $\nu_1\geq 0$ such that for all $\nu\geq\nu_1$ there exists $r_0=r_0(\nu)$ such that

$$\mu\#\text{Hess}\,F_{\nu,k_0}(z)\geq 0\,,\qquad \nu\geq\nu_1,\ \varepsilon_0\leq|\theta|\leq 2\theta_0-\varepsilon_0,\ |z|\geq r_1(\nu)\,.$$

If we combine this (F.1.7) we see that we have a proof of the proposition.

The rest of this section will be devoted to the proof of the proposition when the support of μ is not compact under some additional conditions. This is interesting on its own right but is not essential for the rest of the paper. We start by extracting as much as possible from our previous argument.

Let $\xi_i(z)$, $i=1,2$ be the two eigenvalues of Hess$F_{\nu,k}$ with $|\xi_1|\geq|\xi_2|$ (when $0<\theta<\varepsilon_0$ with our previous notations we have of course $\xi_i=r^{\nu-2}\theta^{k-2}\lambda_i$, $i=1,2$). By analizing our previous argument we see that if $\operatorname{supp}\mu\subset[|x|\leq\beta]$ for some fixed β, then we can find $k_0,\nu_0,c_0>0$ such that for all $\nu\geq\nu_0$ there exists $r_0=r_0(\nu)$ such that

$$\text{(F.1.8)}\qquad \begin{aligned}\mu\#F_{\nu,k_0}(z)&\geq c_0\int\xi_1(z+x)\,d\mu(x)\\ &=c_0\|\text{Hess}\,F_{\nu,k_0}\|*\mu(z)\,,\qquad |z|\geq r_0(\nu)\,.\end{aligned}$$

What is important is to analyze the dependence of c_0 on the parameters of the construction. Following the construction through and preserving the same notations we see that if β and b are kept fixed then we can set $c_0 \geq C_1(b,\beta)\varepsilon^2$. This $\varepsilon > 0$ (which is assumed small) is the ε that was used in the definition of $B_\varepsilon(z_0)$.

The important aspect of the estimate (F.1.8) is that it is "scale invariant". First of all it is clear that nothing changes if we replace $F_{\nu,k}$ by $CF_{\nu,k}$ some constant multiple of $F_{\nu,k}$. Because of the homogeneity of F it follows that we can replace μ by any μ_ρ where μ_ρ is the image of the measure μ by the dilatation $\rho : x \mapsto \rho x$ in $\mathbb{R}^2$. Clearly the dilatation ρ replaces β by $\rho\beta$ and b by $\max\{\rho^2, \rho^{-2}\}\, b$. It follows in particular that we cannot shrink a large β to 1 without at the same time having b go to ∞.

The dependence of c_0 on b, for fixed say $\beta = 1$, must therefore be examined. That dependence is of course picked up by the condition $\mu(B_0 \backslash B_\varepsilon) \leq 10^{-10}$. This gives $\varepsilon \sim b^{-1/2}$ and by the above dilatation argument we conclude that for fixed $b > 1$, $c_0 \sim 1/\beta$. More explicitely if b is fixed, we can choose k_0, ν_0, c such that for all $\nu \geq \nu_0$ and all $\beta > 1$ there exists $r_0 = r_0(\nu,\beta)$ such that

$$\text{(F.1.9)} \qquad \begin{aligned} \mu\#\mathrm{Hess}\, F_{\nu,k_0}(z) &\geq \frac{c}{\beta}\int \xi_1(z+x)\, d\mu(x) \\ &= \frac{c}{\beta}\int \|\mathrm{Hess} F_{\nu,k_0}(z+x)\|\, d\mu(x)\,, \end{aligned}$$

for $\mu \in \mathbb{P}[|x| \leq \beta]$, $E(\mu) \in \mathcal{P}_b$, $|z| \geq r_0(\nu,\beta)$. The only thing that really counts in (F.1.9) is that the dependence of the co-factor of the integrals is polynomial in β. A co-factor of the form c/β^{10} would have been just as good for our purposes.

With the help of (F.1.9) we shall generalise our proposition to measure that are not compactly supported. To do this we have to go back to Section D.1 and to start from some $\mu \in \mathbb{P}(\mathbb{R}^n)$ such that

$$\int |x|^N d\mu(x) < +\infty$$

and such that

$$\overline{\mu} = \int x\, d\mu(x) = 0\,, \qquad \Big(\int x_i x_j\, d\mu(x)\,;\ i,j = 1,\dots,n\Big) \in \mathcal{P}_c\,,$$

for some large N large enough and some $c > 1$. We shall next consider the measure λ that corresponds to μ as in Section D.1. It is for that measure λ that we shall need to generalize our proposition and prove that

$$(\lambda \# \operatorname{Hess} F_{\nu,k_0})(x) \geq 0 , \qquad x \in \mathbb{R}^n,\ |x| \geq r_0 . \tag{F.1.10}$$

This new measure λ also satisfies

$$\int |x|^N d\lambda(x) < +\infty ,$$

$$E = \Big(\int x_i x_j \, d\lambda(x) \, ; \ i,j = 1, \ldots, n \Big) \in \mathcal{P}_b ,$$

for some N as large as we like and some $b > 1$ (but does not necessarily satisfy $\overline{\lambda} = \int x \, d\lambda(x) = 0$).

The next step is to examine $\|\mathrm{Hess} F_{\nu,k}\|$ as obtained from the two formulas (F.1.1), (F.1.2). An easy calculation gives

$$C_2(k^2 r^{\nu-2}\theta^{k-2} + \nu^2 r^{\nu-2}\theta^k) \leq \|\mathrm{Hess} F_{\nu,k}\| \leq C_1(k^2 r^{\nu-2}\theta^{k-2} + \nu^2 r^{\nu-2}\theta^k)$$

valid in the $1/2$-C_α cone $(0 < \theta < \theta_0)$ that is closest to the edge $\theta = 0$. It follows that in that region if we use cartesian coordinates we have

$$\begin{aligned} \|\operatorname{Hess} F_{\nu,k}\| \sim k^2&(x^{\nu-k} y^{k-2} + x^{-k+2} y^{\nu+k-4}) \\ &+ \nu^2 (x^{\nu-k-2} y^k + x^{-k} y^{\nu+k-2}) . \end{aligned} \tag{F.1.11}$$

If we combine the two $1/2$ subregions of C_α we see that if we denote by $\xi = \xi(z) = \mathrm{dist}(z, \partial C_\alpha)$ we obtain the estimate

$$\begin{aligned} \|\operatorname{Hess} F_{\nu,k}\| \sim k^2&(x^{\nu-k} \xi^{k-2} + x^{-k+2} \xi^{\nu+k-4}) \\ &+ \nu^2 (x^{\nu-k-2} \xi^k + x^{-k} \xi^{\nu+k-2}) , \end{aligned} \tag{F.1.12}$$

valid in the whole C_α. Let us consider the functions $\psi_{A,B}(z) = x^A \xi^B$ $(z \in C_\alpha)$ and $\psi(z) \equiv 0$ $(z \notin C_\alpha)$ where (A, B) takes the four possible values that appear in the right hand side of (F.1.12). To prove (F.1.10) it will suffice to show that any of the above four functions $\psi_{A,B} = \psi_{\nu,k}$ has the following property: There exist $C_1, C_2 > 0$ that do not depend on ν (but may depend on k) such that for all $\nu \geq 1$ there exists $u_0(\nu) > 0$ such that

$$\int_{|z|>\beta} \psi_{\nu,k}(u+z) \, d\lambda(z) \leq \Big(\frac{C_2}{\beta}\Big)^{\nu/2} \int_{|z|\leq C_1} \psi_{\nu,k}(u+z) \, d\lambda(z) , \tag{F.1.13}$$

for $\nu \geq 1$, $|u| \geq u_0(\nu)$, $\beta \geq 1$, $u \in C_\alpha$. Indeed once we have (F.1.13) we shall truncate λ at $[|z| \leq \beta]$. If we use (F.1.9) and the same correcting argument as at the end of Section D.1 we see that (F.1.10) follows.

Let us fix $R, \varepsilon > 0$, let us assume that λ satisfies (D.1.2) and let us denote

$$m_{\nu,k}(u) = \sup_{x;|x-u|=R} \{\inf_z [\psi_{\nu,k}(z) \;:\; |x-z| \leq R-\varepsilon]\}.$$

It is clear from (D.1.2) that for an appropiate $C_1 > 0$ we have then

$$m_{\nu,k}(u) \leq C_1 \int_{|z|\leq C_1} \psi_{\nu,k}(u+z)\, d\lambda(z).$$

(F.1.13) will therefore follow as soon as we can show that

$$\int_{|z|>\beta} \psi_{\nu,k}(u+z)\, d\lambda(z) \leq \Big(\frac{C_2}{\beta}\Big)^{\nu/2} m_{\nu,k}(u)$$

with ν, u and β as in (F.1.13). The only thing, of course, that really has to be verified in the above estimate is that the constant C_2 is uniform in ν. By the structure of the above functions ψ it is clear also that we can fix $A \gg R + \varepsilon$ and distinguish the following two cases.

Case 1. distance$(u, \partial C_\alpha) > A$. One then simply has to verify that

$$\int_{|z|>\beta} \psi_{\nu,k}(u+z)\, d\lambda(z) \leq \Big(\frac{C_2}{\beta}\Big)^{\nu/2} \psi_{\nu,k}(u).$$

Case 2. distance$(u, \partial C_\alpha) \leq A$.

Observe finally that we are essentially dealing with two types of functions

$$\psi = r^n \xi^a,\; r^{-a}\xi^n,$$

where the notations and $a, n \geq 0$ are as in (F.1.12) and that the estimates obtained must be uniform in n. Clearly also because of the symmetry about the axis of C_α of the above functions we may suppose that u lies in the half of C_α that is closest to $\theta = 0$. We can then substitute in the integrand the following two functions $\tilde{\psi}$ (that up to a multiplicative constant, dominate ψ)

$$\tilde{\psi}(z) = y^a x^n,\; \tilde{\psi} = x^{-a} y^n, \qquad z = (x,y) \in C_\alpha = \{0 < \theta < 2\theta_0\}$$

and $\tilde{\psi}$ is assumed to be $\equiv 0$ outside C_α.

For these new functions, and an appropiate choice of R, ε and A as above, the verification that we have to make in Case 2 reduces to

$$\int_{|z|\geq\beta} \tilde{\psi}_{\nu,k}(u+z)\, d\lambda(z) \leq \Big(\frac{C_2}{\beta}\Big)^{\nu/2} \tilde{\psi}_{\nu,k}((x,A))\,, \tag{F.1.14}$$

for $u=(x,y)$, $|u|\gg 1$, $0<y<A$. Finally if $0<\operatorname{Arg} u<\theta_0$, *i.e.* if u lies in the half of C_α closest to the x-axis, it is easy to see that it suffices to make the above verifications with a modified $\tilde{\psi}$ given by

$$\tilde{\psi}(z) = \begin{cases} y^a x^n,\ x^{-a}y^n, & z=(x,y),\ y>0,\ x>1\,,\\ 0\,, & \text{otherwise}\,. \end{cases}$$

Four inequalities have to be verified (uniformly in n) and I can see no other way than to just compute. Or rather let the reader compute for himself. At this point life can be made considerably simpler if we impose the following stronger condition on λ

$$d\lambda(z) \leq C_N(1+|x|)^{-N}(1+|y|)^{-N}dz \qquad \text{for all } N\geq 1\,.$$

This condition if applied to (F.1.14) "splits" with respect to the two variables x and y and the calculations simplify since they now reduce to the calculation of 1 dimensional integrals. Given that for all our applications the above stronger condition on λ actually holds the verification under this stronger condition is "good enough". The details will be left to the interested reader.

F.2. An alternative approach and higher dimensions.

For the dimension $n=2$ the method that I developed in Section F.1 is unduly complicated. Indeed in the case $n=2$ it is much easier (and also throws additional light to the problem) to procede differently.

I shall briefly outline here this alternative method. We shall only examine what happens close to the boundary ∂C_α, because for θ away from $\pm\theta_0$ everything is much easier. We shall therefore use the formula (F.1.1). If we denote by $\operatorname{Hess} F = (a_{ij})_{i,j=1,2}$ the coefficients of that Hessian it is very easy to verify that for any $\varepsilon_0>0$ there exist k_0, ν_0, θ_0 (all depending on ε_0) such that

$$(1+\varepsilon_0)a_{11}a_{22} \geq a_{12}^2\,, \qquad \nu\geq\nu_0,\ k\geq k_0,\ 0\leq\theta\leq\theta_0\ .$$

The key to this alternative method is to show that under appropiate conditions on the measure $\mu(z)$, we can "make up" for the factor $(1+\varepsilon_0)$ and guarantee that the matrix $B = (b_{ij})_{i,j=1,2}$

$$
\begin{aligned}
b_{11} &= \int a_{11}(z)\cos^2\theta\, d\mu(z)\,,\\
b_{12} &= \int a_{12}(z)\cos\theta\sin\theta\, d\mu(z)\,,\\
b_{22} &= \int a_{22}(z)\sin^2\theta\, d\mu(z)\,,
\end{aligned}
$$

satisfies $b_{11}b_{22} \geq b_{12}^2$. The matrix B is therefore positive definite and our proposition follows.

The details of the above method are easy to carry out. At any rate they are much easier than what was done in Section F.1. The reason why I presented the proof for $n = 2$ in Section F.1 as I did was because the method of Section F.1 generalizes in a more or less obvious way (although the computations are somewhat tedious to carry out) to higher dimensions. I shall not write the proof down for $n \geq 3$ here. Indeed in a future publication the whole problem will be reexamined from a more general point of view.

F.3. A final remark.

The proofs given in this section of the appendix are very technical, to say the least. All this work seems to be incompressible if we wish to consider convolution operators with an arbitrary Gs-measure $\mu \in \mathbb{P}(G)$ as in Theorem B. If however we only wish to develop the necessary tools for the lower estimate of A_2) then a completely different approach (that is more sophisticated and deep but technically much easier) can be used.

This approach will be developed at great lenght elsewhere I shall give however here the basic principles. It relies on the following two facts:

1) There exists $u \geq 0$ some non zero function on X that is continuous, vanishes outside $C \times K$, and satisfies $Du = 0$ in $C \times K$.

The existence of such a positive "harmonic" function relies on non trivial ideas from potential theory (A. Ancona [32] and L. Carleson [33] are the key references) which we have to adapt in our context.

2) A function u that satisfies the above conditions is automatically unbounded and of polynomial growth.

The proof of 2) is "lighter" than that of 1) but does rely on a scaled Harnack principle which, for large balls, can only be obtained by the Moser iterative process (*cf.* [34]). At any rate all the details will eventually be presented in a separate paper.

References.

[1] Varopoulos, N. Th., Saloff-Coste, L., Coulhon, Th., *Analysis and geometry on groups.* Cambridge Univ. Press, 1992.

[2] Hörmander, L., Hypoelliptic second order equations. *Acta Math.* **119** (1967), 147-171.

[3] Varopoulos, N. Th., Thèorie de Hardy-Littlewood sur les groupes de Lie. *C. R. Acad. Sci. Paris, Serie I* **316** (1993), 999-1003.

[4] Bony, J. M., Principe du maximum, inégalité de Harnack et unicité du problème de Cauchy pour les opérateurs elliptiques dégénérés. *Ann. Inst. Fourier* **19** (1969), 277-304.

[5] Jacobson, N., *Lie algebras.* Interscience, 1962.

[6] Varadarajan, V. S., *Lie groups, Lie algebras and their representations.* Prentice-Hall, 1984.

[7] Chevalley, C., *Théorie des groupes de Lie*, tomes II & III. Hermann, 1955.

[8] Jacobson, N., *Basic algebra*, vol. I & II. W. H. Freeman & Comp. 1989.

[9] Helgason, S., *Differential geometry, Lie groups and symmetric spaces*, vol. **80**. Academic Press, 1978. *Groups and geometric analysis*, vol. **113**. Academic Press, 1984.

[10] Onischik, A. L., Vinberg, E. B., *Lie groups and algebraic groups.* Springer-Verlag, 1988.

[11] Guivarc'h, Y., Croissance polynomiale et périodes des fonctions harmoniques. *Bull. Soc. Math. France* **101** (1973), 333-379.

[12] Alexopoulos, G., An application of homogenization theory to harmonic analysis: Harnack inequalities and Riesz transforms on Lie groups of polynomial growth. *Can. J. Math.* **44** (1992), 691-727.

[13] Varopoulos, N. Th., Diffusion on Lie groups (I) and (II). *Can. J. Math.* **46** (1994), 438-448 and 1073-1093.

[14] Knapp, A., *Representation theory of semisimple groups.* Princeton Univ. Press, 1986.

[15] Varopoulos, N. Th., Hardy-Littlewood theory on unimodular groups. *Ann. Inst. H. Poincaré* **31** (1995), 669-688.

[16] Reiter, H., *Classical harmonic analysis and locally compact groups.* Oxford Math. Monograph, 1968.

[17] Moser, J., On the volume elements on a manifold. *Trans. Amer. Math. Soc.* **120** (1965), 286-294.

[18] Gromov, M., *Asymptotic invariance of infinite groups.* I.H.E.S.-M-92-8.

[19] Varopoulos, N. Th., Small time Gaussian estimate of Heat diffusion kernels (I) and (II). *Bull Sc. Math.* **113** (1989), 253-277. *J. Funct. Anal.* **93** (1990), 1-33.

[20] Hebisch, W., Estimates on the semigroups generated by left invariant operators on Lie groups. *J. für die reine und angew. Math.* **423** (1992), 1-45.

[21] Vauthier, J., *Bull. Sci. Math.,* 2e série **103** (1979), 129-177.

[22] Ibragimov, I. A., Linnik, Yu. V., *Independent and stationary sequences of random variables.* Wolters-Noordhoff, 1971.

[23] Hörmander, L., *The analysis of linear partial differential operators*, vol. III. Springer-Verlag, 1985.

[24] Bers, L., John, F., Schechter, M., *Partial differential equations.* Lectures in appl. Math., vol. III. Interscience, AMS, 1974.

[25] Bourbaki, N., *Eléments de mathématiques, Livre* VI*, Intégration.* Hermann, 1963.

[26] Warner, G., *Harmonic analysis on semisimple Lie groups*, vol. I. Springer-Verlag, 1972.

[27] Burkholder, D. L., Distribution function inequalities for martingales. *Ann. Prob.* **1** (1973), 19-42.

[28] Mc Kean, H. P., Jr., *Stochastic integrals.* Academic Press, 1969.

[29] Stroock, D. W., Varadhan, S. R. S., *Multidimensional diffusion processes.* Springer-Verlag, 1979.

[30] Varadhan, S. R. S., *Larges deviations and applications.* SIAM, 1984.

[31] Schaefer, H. H., *Banach lattices and positive operators.* Springer-Verlag, 1975.

[32] Ancona, A., Principe de Harnack à la frontière et théorème de Fatou pour un operateur elliptique dans un domaine lipchitzien. *Ann. Inst. Fourier* **28-4** (1978), 162-213.

[33] Carleson, L., On the existence of boundary values for harmonic functions of several variables. *Ark für Math.* **4**, (1962).

[34] Varopoulos, N. Th., Fonctions harmoniques sur les groupes de Lie. *C. R. Acad. Sci. Paris* **304(I)** (1987), 519-521.

[35] Wallach, N. R., *Real reductive groups*, vol. I. Academic Press, 1988.

[36] Feller, W., *An introduction to probability theory and its applications*, vol. I & II. Wiley, 1968.

[37] Varopoulos, N. Th., Wiener-Hopf theory and non-unimodular groups. *J. Funct. Anal.* **120(2)** (1994), 467-483.

[38] Varopoulos, N. Th., *Potential theory of non-unimodular groups: "Harmonic analisys and discrete potential theory"*. Plenum Press, 1992.

[39] Bensoussan, A., Lions, J. L., Papanicolaou, G., *Asymptotic analysis for periodic structures.* North-Holland, 1978.

[40] Varopoulos, N. Th., Mustapha, S., Forth coming book. Cambridge University Press.

[41] Bougerol, Ph., Théorème central limit local sur certain groupe de Lie. *Ann. Sci. Ec. Norm. Sup.* 4e *sér.* **14** (1981), 403-432.

[42] Steenrod, N., *The topology of fiber bundles.* Princeton University Press, 1951.

[43] Varopoulos, N. Th., Distance distortion on Lie Groups. *Institute Mittag-Leffler Report* **31**, 1995/96.

Recibido: 15 de junio de 1.995
Revisado: 25 de julio de 1.995

Nick Th. Varopoulos
I.U.F. and
Departement de Mathématiques
Université de Paris VI
75005 Paris, FRANCE