Revista Matemática Iberoamericana
Vol. 9, N.° 1, 1993

Semigroup Commutators under differences, II

N. Th. Varopoulos

0. The scope of the paper.

This is the second instalment of my previous paper with the same title, [1]. This paper consist of two different parts. The first part is devoted to improvements of the results developed in [1]. These improvements are explained in Section 0.1 below and developed in sections 1 to 5, and 9 to 10; they are in fact technically distinct from [1] and rely on a systematic use of "microlocalisation" in the context of Hörmander-Weyl calculus. These paragraphs can therefore be read quite independently from [1].

The second part studies a different problem and is, in its aim, fairly disjoint from [1]. This problem is explained in Section 0.2 below and developed in sections 6 to 8. The techniques used however in sections 6 to 8 (and also in Section 10 which in its scope is attached to the first part) are very close to the techniques of [1]. I feel that the reader would find it very difficult to follow these sections without being familiar with [1].

0.1. Pseudodifferential operators and the geometric problem.

The main technical estimate in [1] was the estimate (0.2) that asserted

that

$$(0.1) \qquad \|[\cdots[A^\sigma, S_1], S_2], \cdots], S_k]\, f\|_m \leq C\, \|A^{\sigma - k/2} f\|_{m+n_1+\cdots+n_k}\ ,$$

when $f \in C_0^\infty$.

Here $[x, y] = xy - yx$ are as usual the commutators of two operators, $\|\cdot\|_\alpha$ indicate the usual Sobolev norms in $H_\alpha = \{f : \Lambda^\alpha f \in L^2(\mathbb{R}^n)\}$ ($\Lambda = (1 - \sum \partial^2/\partial x_i^2)^{1/2}$), $\sigma \in \mathbb{C}$, and $A = a^\omega(x, D) + \lambda_0$ for some large $\lambda_0 > 0$ and $0 \leq a(x, \xi) \in S^2_{1,0}$ and finally $S_j = s_j^\omega(x, D)$ with $s_j \in S^{n_j}_{1,0}$. It will turn out that a systematic use of Weyl calculus [10] (rather than ordinary $S^m_{1,0}$ pseudodifferential calculus) will be convenient in several places and will therefore be used interchangeably with pseudodifferential calculus.

The estimate (0.1) was proved in [1] for sums of square (-Hörmander) operators: $A = \sum X_j^* X_j$ where X_j are C^∞ fields. This estimate was not even proved for a general second order self adjoined differential operator of positive characteristic (*cf.* [1], (0.1)). Indeed, as far as I can tell the problem is as hard in this case as for a general pseudodifferential. As a result the main geometric theorem in [1] (and all the rest for that matter) was established only for Hörmander operators.

In this paper I shall give a complete proof of (0.1) for $A = a^\omega(x, D) + \lambda_0$ in full generality but only for $k = 1$. This will be done in sections 1 to 4 in the context of Hörmander's-$S(m, g)$ calculus with $A \in S(1/h^2, g)$. I shall also show that (0.1) holds (and this is easy because of previous work of R. Beals, *cf.* also the appendix at the end of this paper) for arbitrary k but with an A that is polyhomogeneous and subelliptic with a loss of one derivative (*cf.* Section 9 for the appropriate definitions).

Using the above results we shall show in Section 10 that in the main geometric theorem of [1] we can relax the sum of squares condition for the "top operator" L_1 (the set-up was $\|L_2^\beta f\| \leq C\,(\|L_1^\alpha f\| + \|f\|)$), which can therefore be an arbitrary self adjoint differential operator

$$L_1 \equiv a_{ij} \frac{\partial^2}{\partial x_i \partial x_j} + \cdots$$

The above estimate (0.1) for $k = 1$ has a number of other more "esoteric" consequences, *e.g.* the boundedness of the operators

$$A^{i\sigma},\ e^{i\sigma A^{1/2}} : H_\alpha \to H_\alpha; \qquad \alpha, \sigma \in \mathbb{R}$$

i.e. the imaginary powers of A and the corresponding wave operators. These facts will be proved in sections 5 and 6.

0.2. The Beals characterisation and the $S^m_{\rho,\delta}$.

In Section 8, I will give the following characterisation of pseudodifferential operators (which is but a variant of the characterisations given by R. Beals [3]).

Criterion. *Let T be an arbitrary linear operator $T : C_0^\infty(\mathbb{R}^d) \to \mathcal{E}'(\mathbb{R}^d)$ and let $1/2 \leq \rho \leq 1$ and $m \in \mathbb{R}$ be such that*

$$\| [[\cdots [T, E_1], E_2] \cdots, E_k] \|_{\alpha \to \alpha + k\rho - m} \leq C, \quad k \geq 0, \alpha \in \mathbb{R} \tag{0.2}$$

where $E_j \in S^0_{1,0}$ are arbritary. Then $\varphi T \varphi \in OPS^m_{\nu,\nu}$ for all $\nu \in [1-\rho, \rho]$ and all $\varphi \in C_0^\infty$.

Here we use of course the standard Hörmander notations for $S^m_{\rho,\delta}$ (*cf.* [2]) and $\| \cdot \|_{\alpha \to \beta}$ indicates the operator norm between the corresponding Sobolev spaces $H_\alpha(\mathbb{R}^d)$. The C in (0.2) depends, of course, on α and k as well as on the E_j's.

The Beals theorem that we refer to appeared for the first time in [3] (*cf.* also [2]). Essentially the same proof was given later in [5]. In [5] the authors work in the context of classical pseudodifferential operators and their assumption is

$$\| [T, X_1, \cdots, X_k] \|_{(1-\rho)k + m \to 0} \leq C \tag{0.3}$$

with $\rho = 1$, or $1/2$ and where X_j are C^∞ fields on $\mathbb{R}^d$. The proofs in [3], [4] and [5] easily generalise and give (under the hypothesis (0.3)) the same conclusion

$$T \in \bigcap_{1-\rho \leq \nu \leq \rho} OPS^m_{\nu,\nu} = \mathcal{B}^m_\rho \, .$$

Incidentally, standard pseudodifferential calculus can be used and it follows that conversely every $T \in \mathcal{B}^m_\rho$ satisfies the commutator estimates (0.2) and (0.3). This implies in particular that $\mathcal{B}^m_\rho$ can be defined in a coordinate free way (*i.e.* on a manifold). The reference [5] is perhaps the easiest for the reader who is not familiar with (φ, Φ) calculus.

Using the above criterion we shall prove in Section 8 the following

Theorem. *Let $A = a^{\omega}(x, D) \in OPS^2_{1,0}$ with symbol $a(x, \xi) \geq 0$, let $\sigma \in \mathbb{C}$, $\operatorname{Re} \sigma \leq 0$, and let us assume that A is subelliptic:*

$$\|f\|^2_{1-\delta} \leq C\,(Af, f) + C_1 \|f\|^2$$

with $0 \leq \delta \leq 1/2$. Then for all $\varphi \in C^{\infty}_0$ and $\lambda > 0$ large enough the operator $\tilde{A}^{\sigma} = \varphi(A + \lambda)^{\sigma}\varphi$, (this is just a banal modification to reduce the problem to compact supports) satisfies

$$\tilde{A}^{\sigma} \in \mathcal{B}^{2\operatorname{Re}(\sigma(1-\delta))}_{1-\delta} .$$

The proof of this theorem will be given in sections 7 and 8. It is interesting to compare the above result with the final theorem in Beals [3]. Beals theorem (if the proof is pushed to its limit) will give a better conclusion since it will show that the correspondig parametrix belongs to $S^{2\operatorname{Re}(\sigma(1-\delta))}_{1-\delta,\delta}$. Beals theorem is also better in so far that it can deal with operators of higher order $S^m_{1,0}$, $m \geq 2$ and does not require that the symbol is positive (but only that the principal symbol takes values in an appropriate sector).

Our theorem above has however some advantages, the most significant of which is that it can deal with general symbols (and not only polyhomogeneus ones as seems to be the case in Beals). The other advantage is an advantage of the method of the proof (which is different from Beals' method) rather than of the result. Indeed, in our considerations, we can replace the Sobolev norms $H_{\alpha}(\mathbb{R}^d)$ by the corresponding L^p-Sobolev norms

$$H^p_{\alpha} = \{f : \quad \Lambda^{\alpha} f \in L^p\}\,, \qquad 1 < p < \infty\,,$$

and the estimates are relatively insensitive to that change, provided, that the original operator is a *differential* operator with positive characteristic. In view of the fact that the Hörmander classes $S^0_{\rho,\delta}$ do not in general stabilise L^p, results of this kind are perharps of some interest.

Finally other functions than the complex powers (with non positive real part) of A can be treated with our methods. It easily follows, for instance, that under the same conditions, and with the same notations as in our Theorem, we have

$$\varphi e^{-zA}\varphi \in \mathcal{B}^0_{1-\delta}\,, \qquad |z| \leq L,\ |\operatorname{Arg} z| \leq \frac{\pi}{2} - \varepsilon_0 \tag{0.4}$$

uniformly in z and any fixed $\varepsilon_0 > 0$. And also that

$$\tilde{A}^\sigma \in \mathcal{B}_{1-\delta}^{2\mathrm{Re}\,\sigma}\,, \qquad \mathrm{Re}\,\sigma \geq 0\,. \tag{0.5}$$

1. The Hörmander metrics.

This section is purely technical and contains nothing new. I simply collect a number of comments and elaborations on the $S(m,g)$ calculus as presented in L. Hörmander book [6]. All the notations will be (unless otherwise stated) identical to those of [6]. The facts that I shall need will be enumerated below. The proofs are just cross references in [6] and will be briefly explained after each fact.

(A) In [6], Lemma 18.4.4 can be improved to (with the same notations): given ν then the number of balls B_μ that intersects B_ν is bounded by N_ε.

This slightly stronger local finiteness property will simplify several of our arguments. When we examine the proof of the above lemma in [6], which is to be found in [6], Lemma 1.4.9, we see that this stronger property is in fact implicit in that proof.

(B) The choice of the balls U_ν, U'_ν defined in [6] just after relation (18.4.13) can be refined in the following way: we can choose (for $k = 1, 2, \ldots$ given in advance)

$$U_\nu = U_\nu^{(0)} \subset U_\nu^{(1)} \subset \ldots U_\nu^{(k)}\,, \quad U_\nu^{(j)} = \{x : g_{x_\nu}(x - x_\nu) < c_j\}$$

for $j = 0, 1, \ldots, k$, in such a way that the balls $U_\nu^{(k)}$ have the local finiteness property of (A). Furthermore the ε in Lemma 18.4.4 and $c_0 > 0$ the radius of U_ν can be chosen small enough, so as to guarantee that

$$U_\mu \cap U_\nu^{(j)} \neq \emptyset \text{ implies } U_\mu \subset U_\nu^{(j+1)}\,, \quad \forall \nu, \mu\,, \quad j = 1, 2, \ldots, k-1\,.$$

Observe that in the elaborations and proofs of [6], sections 18.4 and 18.5 the two balls $U_\nu^{(j)} \subset U_\nu^{(j+1)}$, for any $j = 0, 1, \ldots, k-1$, could be used in the place of the paire $U_\nu \subset U'_\nu$ of [6]. The point to watch, and which is vital for us, is what lies between relations (18.4.19) and (18.4.21) in [6].

All this is fairly automatic from [6] and the proof will be left to the reader. Let me simply say that the reader whether he likes it or not will have to really understand [6] sections 18.4 and 18.5 if he wishes to follow what is happening. This applies especially here and in the next few pages.

(C) Let $a_i \in S(m_i, g)$, $(i = 1, 2)$, be such that supp $a_1 \cap$ supp $a_2 = \emptyset$, let $b \in S(m_1 m_2, g)$ be such that $b^\omega = a_1^\omega a_2^\omega$, then we actually have $b \in S(m_1 m_2 h^N, g)$, $N \geq 0$.

This is contained in [6], Theorem 18.5.4.

(D) Let U_ν, U'_ν be as in [6] just after (18.4.13) and let $a \in S(m, g)$. We shall say that a is strongly concentrated "at ν in $S(m, g)$" if it satisfies the condition

$$|a|_s^g(\omega) \leq C_{k,s}\, m(\omega)\,(1 + d_\nu(\omega))^{-k}\,, \quad \forall k, s, \omega\,. \tag{1.1}$$

We shall say that a is concentrated (without the adjective strongly) "at ν in $S(m, g)$" if the same estimate (1.1) holds but *only* for $\omega \notin U'_\nu$ (Definition in [6] just after relation (18.4.13)).

The "subtlety" of the above notion lies in the fact that the balls U_ν are defined by the metric g while the distance d_ν is defined by the metric g^A (or g^σ in our case). In [4] Beals introduced an analogous notion which he then exploited in the special case when $g = g^\sigma$.

Observe that the above definition depends on the particular choice of U_ν, U'_ν. The conclusions that this property of concentration will allow us to draw will, on the other hand, be independent of that choice (*cf.* especially property (E) and Section 2 below). So, therefore, at the end, it will be irrelevant with respect which particular U_ν, U'_ν we are making the definition. The above notion will prove itself to be useful in the following properties.

(E) Let $a_\nu \in S(m, g)$ $(\nu \in \mathbb{N})$ be a family of operators so that a_ν is concentrated at ν in $S(m, g)$ for each ν, and that furthermore these conditions are verified *uniformly* in ν. Then the family $\sum a_\nu$ is "absolutely summable" in the sense that we have

$$\sum |a_\nu|_s^g(\omega) \leq C_s\, m(\omega)\,, \qquad \forall \omega\,. \tag{1.2}$$

If we demand that (a_ν) should be *strongly* concentrated and consider the case $m \equiv 1$, then the above statement is an immediate consequence of [6], Lemma 18.4.8. The modifications needed for the proof when m is arbitrary are obvious. If we only impose the weaker property of concentration (rather than strong concentration), then we have to split the sum in (1.2) as follows

$$\sum_{\omega \in U'_\nu} + \sum_{\omega \notin U'_\nu}$$

(This is essentially the argument of [6] between the relations (18.4.19)-(18.4.21)). The first of the two sums is bounded by $C\, m(\omega) \sup_\nu \|a_\nu\|$ because of the local finiteness of our partition. To controle the second sum we apply the same argument (*cf.* [6], Lemma 18.4.8) as before.

(F) Let $a^{(i)} \in S(m_i, g)$, $i = 1, 2$, and let $a \in S(m_1 m_2, g)$ be such that $a^\omega = a^{(1)\omega} a^{(2)\omega}$. Let us suppose further that for some fixed ν and *either* $i = 1$, *or* $i = 2$ (or both) we have supp $a_\nu^{(i)} \subset U_\nu$. Then a is strongly concentrated at ν in $S(m_1 m_2, g)$.

(F′) We impose the same conditions on $a^{(1)}$, $a^{(2)}$, a as in (F) (with say supp $a^{(1)} \subset U_\nu$). In additions we demand that $U'_\nu \cap$ supp $a_\nu^{(2)} = \emptyset$. Then a is concentrated at ν in $S(m_1 m_2 h^N, g)$ for all $N \geq 0$. (Here $U_\nu \subset U'_\nu$ are as in [6] just after relation (18.4.13)).

In other words we are in an "arbitrary small class" (with respect to h) provided that we are prepared to sacrifice the property of *strong* concentration. This property should be thought as an elaboration of both (F) and (C).

For the proof of (F) the relevant passage in Hörmander is what lies a dozen lines after relation (18.6.6) and goes on until Theorem 18.6.6. In fact in that passage one essentially finds the proof of our statement for $m_1 \equiv m_2 \equiv 1$. Indeed let $m_1 \equiv m_2 \equiv 1$ and supp $a^{(1)} \subset U_\nu$ and let us proceed as in Hörmander and decompose $a^{(2)} = \sum a_\mu$ so that (in Hörmander's notations):

$$a = \sum_\mu a_{\nu\mu}\,, \quad a_{\nu\mu}^\omega = a^{(1)\omega} a_\mu^\omega\,.$$

(We ignore the complex conjugation of Hörmander here). The estimate

$$|a_{\nu\mu}(\omega)| \leq c_k\, |1 + M(\omega)|^{-k} = c_k\, |1 + d_\nu(\omega) + d_\mu(\omega)|^{-k} \tag{1.3}$$

(when $m_1 \equiv m_2 \equiv 1$ and where M is as in [6] bottom of p. 167, vol. III, 1985), then holds and if we use the uniform polynomial growth of M_μ (notations of proof of [6], Lemma 18.4.8) established in [6], Lemma 18.4.8, we obtain that

$$\sum_\mu |a_{\nu\mu}(\omega)| \leq C_k \, |1 + d_\nu(\omega)|^{-k} .$$

This is the required estimate (1.1) for $s = 0$. The modification for arbitrary m_1, m_2 is clear, we just have to insert the factor

$$|a_{\nu\mu}(\omega)| \leq c_k \, m_1(\omega) \, m_2(\omega) \, |1 + M(\omega)|^{-k}$$

in the estimate (1.3). This can clearly be done by the definition of $a_{\nu\mu}$ (*cf.* the same passage of Hörmander: bottom of p. 167, vol. III, 1985 edition, and also the estimate [6], (18.4.12)). To pass to the estimate for the more general seminorms $|a_{\nu\mu}|_s^g$, $(s \geq 1)$, we have to improve (1.3) exactly the way Hörmander does in (18.6.7) and (18.6.8). We obtain

$$|a_{\nu\mu}(\omega)| \leq C_k \, (1 + d_{\nu\mu})^{-k} \, |1 + M(\omega)|^{-N}$$

which is essentially [6], (18.6.9) except that we retain all the information given by [6], (18.6.7) and [6], (18.6.8). We then reason exactly as in [6] (the passage that follows relation [6], (18.6.8)) and we are done in the case $m_1 \equiv m_2 \equiv 1$. The general m_1, m_2 are treated similarly.

To obtain the refinement that is presented in (F′) we have to combine the above argument with the passage in [6] between relations (18.4.19) and (18.4.21) (*i.e.* p. 148-149, vol. III, 1985 edition) what is shown there is that we can improve by an arbitrary power of h^N provided that we are away from the "support". More specifically in the relation that defines $a_{\nu\mu}(x, \xi)$ (bottom of [6], p. 167) if we know that $(x, \xi) \notin U'_\nu$ we can obtain the following improvement to the estimate (1.3)

$$(1.4) \qquad |a_{\nu\mu}(\omega)| \leq C_k \, h^N(\omega) \, (1 + M(\omega))^{-k} , \qquad \omega \notin U'_\nu .$$

This is explained in [6] in the passage between relations (18.4.19)-(18.4.21). In that passage we set $V = W \oplus W$, the metric is $G = g \oplus g$ (*i.e.* $g_1 = g_2 = g$) and $A = 2\,\sigma(\hat{x}, \hat{\xi}, \hat{y}, \hat{\eta})$ as in [6], p. 152. Observe that since $g_1 = g_2$ the metric G is now temperate everywhere on $W \oplus W$

and not only on the diagonal. This makes the reasonning easier for it implies that with $\omega = (x, \xi)$, $\omega' = (y, \eta)$, we have

$$|e^{i\sigma(D_x, D_\xi; D_y, D_\eta)/2} a^{(1)}(\omega) a_\mu(\omega')| \leq C_k \, H^N (1 + d_\nu(\omega) + d_\mu(\omega'))^{-k} \, ,$$

$\forall (\omega, \omega') \notin U'_\nu \times U'_\mu$. From this (1.4) follows immediately by setting $\omega = \omega'$ (since then $H(\omega, \omega') \sim h(\omega) \sim h(\omega')$). This outlines the proof when $m_1 \equiv m_2 \equiv 1$ and $s = 0$. The proof of the estimates in the general case follows by the same modifications as before. Once the estimate (1.4) has been proved (F') follows since (C) guarantees that $a \in S(m_1 m_2 h^N, g)$, $\forall N \geq 0$.

2. The localisation of the commutator estimate.

Let $U_\nu \subset U_\nu^{(1)} \subset \ldots \subset U_\nu^{(k)}$ be as in (B) ($k = 5$ will in fact suffice). Assume that $A = a^\omega(x, D)$, $B_\nu^{(j)} = b_\nu^{(j)\omega}(x, D)$ with $a, b_\nu^{(j)} \in S(m, g)$, $(j = 1, 2, \ldots, s)$. Let us also make the hypothesis that supp $b_\nu^{(j)} \subset U_\nu^{(p)}$ for some $1 \leq p < k$ and let us denote

$$\begin{aligned} A'_\nu &= \sum_\lambda \left\{ (a\varphi_\lambda)^\omega : \quad U_\lambda \cap U_\nu^{(p+1)} \neq \emptyset \right\} \\ \tilde{A}_\nu &= \sum_\lambda \left\{ (a\varphi_\lambda)^\omega : \quad U_\lambda \cap U_\nu^{(p+1)} = \emptyset \right\} \end{aligned} \tag{2.1}$$

where $\sum \varphi_\lambda \equiv 1$ is a Hörmander partition of unity subordinated to the covering $\{U_\nu\}$ as in [6], Lemma 18.4.4.

Then by (F'), $\tilde{A}_\nu B_\nu^{(1)}$ is concentrated at ν in $S(m^2 h^N; g)$ with an $N \geq 1$ that can be given in advance; but then, by a succesive application of (F), $\tilde{A}_\nu B_\nu^{(1)} B_\nu^{(2)}$ (respectively, $\tilde{A}_\nu B_\nu^{(1)} \ldots B_\nu^{(j)}$) is strongly concentrated at ν in $S(m^3 h^N; g)$ (respectively, $S(m^{j+1} h^N; g)$) with the same N.

On the other hand the following two sums are absolutely summable in the sense of (E):

$$\begin{aligned} A' &= \sum A'_\nu B_\nu^{(1)} \ldots B_\nu^{(s)} \, , \\ \tilde{A} &= \sum \tilde{A}_\nu B_\nu^{(1)} \ldots B_\nu^{(s)} \, . \end{aligned}$$

This is because supp $b_\nu^{(s)} \subset U_\nu^{(p)}$ and thus each term of the above summation is concentrated (even strongly) at ν in $S(m^{s+1}; g)$, and so (E)

applies. But, by what we have said just above, $\tilde{A} = \sum \tilde{A}_\nu B_\nu^{(1)} \dots B_\nu^{(s)}$ is in fact absolutely summable in $S(m^{s+1}h^N; g)$ for any $N \geq 1$ arbitrary large. It follows in particular that

$$A \sum B_\nu^{(1)} \dots B_\nu^{(s)} \equiv A' \mod S(m^{s+1}h^N; g), \quad N \geq 0. \tag{2.2}$$

We shall apply these facts to localise a specific expression involving commutators. Let $A = a^\omega$ with $a \in S(m; g)$ and let us follow our notational convention of [1] and denote by $E = e^\omega$ where $e \in S(1, g)$ is arbitrary. The various e's and E's that appear below are not necessarily all the same. Let φ_ν be a partition of unity as in (2.1) and let $E \in OPS(1; g)$ and $A_\nu = (a\,\varphi_\nu)^\omega$. Let E'_ν be defined from E the way A'_ν was defined from A in (2.1) with $p = 0$. We obtain therefore from (2.2) that

$$EA = \sum_\nu EA_\nu \equiv \sum_\nu E'_\nu A_\nu \mod S(h^N m; g), \qquad N \geq 0.$$

But then it follows that

$$\begin{aligned} EAE = \sum EA_\nu E &\equiv (\sum E'_\nu A_\nu)E \\ &= \sum E'_\nu A_\nu E \mod S(h^N m; g) \end{aligned} \tag{2.3}$$

because multiplication is distributive over absolute summation. A simple application of (C) allows to conclude on the other hand that

$$A_\nu \tilde{E}_\nu \in S(mh^N; g) \qquad (N \geq 1, \text{ uniformly in } \nu).$$

This together with (F) (we do not need to use (F') anymore !) implies that $E_\nu A_\nu \tilde{E}_\nu$ is (strongly) concentrated at ν in $S(mh^N; g)$ and we can therefore sum these terms in $S(mh^N; g)$ by Section 1. The conclusion is that

$$\begin{aligned} \sum_\nu E'_\nu A_\nu E &= \sum_\nu E'_\nu A_\nu E'_\nu + \sum_\nu E'_\nu A_\nu \tilde{E}_\nu \\ &\equiv \sum E'_\nu A_\nu E'_\nu, \mod S(mh^N; g), \qquad (N \geq 1). \end{aligned}$$

Combining this with (2.3) we conclude that

$$EAE = \sum E'_\nu A_\nu E'_\nu \mod S(h^N m; g), \quad (N \geq 0).$$

We shall use this idea again to the two products of

$$\sum_\nu [E'_\nu\, A_\nu\, E'_\nu, E] = \sum_\nu (E'_\nu\, A_\nu\, E'_\nu\, E - E E'_\nu\, A_\nu E'_\nu)$$

and we obtain

$$\begin{aligned}[EAE, E] &\equiv \sum [E'_\nu\, A_\nu E'_\nu, E] \\ &\equiv \sum [E'_\nu\, A_\nu\, E'_\nu\,, E''_\nu] \quad \text{mod-}S(mh^N; g) \qquad (N \geq 0)\end{aligned}$$

where E''_ν is constructed from E (the same way A'_ν was constructed from A) as in (2.1) with $p = 3$.

We shall carry this process one step further and finally deduce that

$$[EAE, E] = \sum I_\nu [E'_\nu\, A_\nu\, E'_\nu\,, E''_\nu] I_\nu \quad \text{mod-}S(mh^N; g)$$

where $I_\nu = i_\nu^\omega$: $i_\nu = \left\{ \sum \varphi_\lambda :\ U_\lambda \cap U_\lambda^{(5)} \neq \emptyset \right\}$.

Let us now examine more closely the operators I_ν and A_ν. The first observation is that

$$\sum \|I_\nu f\|^2 \leq C \|f\|^2\,, \qquad \forall f \in C_0^\infty\,. \tag{2.4}$$

This is best seen by considering vector valued symbols (*cf.* [6] relation 18.6.24). Alternatively (and equivalently) the estimate (2.4) can be obtained by taking the expectation on $\| \sum \pm I_\nu f \|$ (as is done in Section 8). Consider next the operator

$$\sum_\nu I_\nu A_\nu I_\nu$$

where again, by that we have said, the summation is absolute. To examine this operator we shall impose for the first time on $a(x, \xi)$, the symbol of A, our basic conditions. We shall assume that

$$a(x, \xi) \geq 0\,, \qquad a(x, \xi) \in S(1/h^2; g)\,. \tag{2.5}$$

Let $\theta = \sum i_\nu^2$. Under the conditions (2.5), we then have

$$a - (a\theta)^\omega = a_1^\omega + a_2^\omega \tag{2.6}$$

where $a_1(x,\xi)$ takes pure imaginary values and $a_2(x,\xi) \in S(1,g)$. Similarly if we denote by $\Theta_{\pm 1} = (\theta^{\pm 1/2})^\omega$, then we have

$$\Theta_1 A \Theta_1 - (a\theta)^\omega = a_1^\omega + a_2^\omega \tag{2.7}$$

with similar conditions on a_1, a_2. Furthermore we have

$$\Theta_1 \cdot \Theta_{-1} \; ; \; \Theta_{-1} \cdot \Theta_1 \equiv I + ib^\omega(x,D) \mod\text{-}S(h^2;g), \tag{2.8}$$

where $b \in S(h,g)$ is real. The last two relations (2.7) and (2.8) are obtained by $S(m,g)$ calculus (in (2.8) we in fact have $b \equiv 0$! One can compare this with the argument in [6], p. 171 just before Theorem 16.6.8). The relation (2.6) is also obtained by symbolic calculus but the presence of the infinite sum that is involved in the definition of a makes life slightly more complicated. There are many ways to deal with that infinite sum, the most elegant is, in my opinion, the one presented in [6] just after relation (18.6.24) where the author uses operator valued symbols.

I shall now show how the above considerations can be used to localise commutators. Our problem [*cf.* Section 0.1] is to show that for operators $A = a^\omega(x,D)$ with symbols that satisfy (2.5) we have

$$\|[A,E]f\|^2 \leq C\,(Af,f) + C_1 \|f\|^2 \,, \qquad f \in C_0^\infty$$

or better still that

$$\|[EAE,E]f\|^2 \leq C\,(Af,f) + C_1 \|f\|^2 \,, \qquad f \in C_0^\infty \,. \tag{2.9}$$

$(\cdot,\cdot)$ indicates of course the scalar product in L^2 and $\|\cdot\|$ the corresponding norm. The reason why we did all the work in this section was because we wanted to show that the estimate (2.9) is "localisable". Let me be more specific and let us assume that we can find some Hörmander covering (U_ν) of the (x,ξ) space as above for which the estimate (2.9) holds "for each ν separately", *i.e.* such that

$$\|[E_\nu\, A_\nu\, E_\nu\,,\, E_\nu]f\|^2 \leq C\,(Af,f) + C_1 \|f\|^2 \tag{2.10}$$

provided that $A_\nu = (ae_\nu)^\omega$, $E_\nu = e_\nu^\omega$ with $e_\nu \in S(1,g)$ uniformly in ν and supp $e_\nu \subset U'_\nu$. Then we shall deduce that the estimate (2.9) itself also holds.

The first thing to observe is that we have

$$\|[EAE, E]f\| \leq \|\sum_{\nu} I_\nu [E'_\nu A_\nu E'_\nu , E''_\nu] I_\nu f\| + C\|f\| . \tag{2.11}$$

We shall need the following

Lemma. *Let $J_\nu = j^w_\nu(x, D)$ with real valued $j_\nu \in S(1; g)$ (uniformly in ν) and* supp $j_\nu \subset U'_\nu$ *then there exists a constant C such that*

$$\|\sum J_\nu f_\nu\|^2 \leq C \sum \|f_\nu\|^2 , \quad f_\nu \in C_0^\infty .$$

The proof of the lemma will be given presently. Let us draw the conclusions: From (2.11), the lemma, and the local hypothesis (2.10), we deduce inmmediately that $\|[EAE, E]f\|^2$ can be estimated by

$$\begin{aligned} &\sum_{\nu} \|[E'_\nu A_\nu E'_\nu , E''_\nu] I_\nu f\|^2 + \|f\|^2 \\ &\qquad \leq \sum_{\nu} (I_\nu A I_\nu f, f) + C\|f\|^2 + C \sum \|I_\nu f\|^2 . \end{aligned} \tag{2.11$'$}$$

(The local hypothesis (2.10) is applied to each $I_\nu f$ separately). And this by (2.4), (2.6) and (2.7) can be estimated by

$$(A\Theta_1 f, \Theta_1 f) + C\|f\|^2 .$$

The upshot is that we have

$$\|[EAE, E]f\|^2 \leq C_1 (A\Theta_1 f, \Theta_1 f) + C_2 \|f\|^2 , \quad f \in C_0^\infty , \tag{2.12}$$

for some $C_1, C_2 > 0$. If we apply (2.12) to $f = \Theta_{-1}\varphi$ we obtain

$$\|[EAE, E]\Theta_{-1}\varphi\|^2 \leq C_1 (A\varphi, \varphi) + C_2 \|\varphi\|^2 \tag{2.13}$$

since $\Theta_{-1} \in OPS(1; g)$. But now we almost have our required estimate. Indeed set $\psi = [EAE, E]\varphi$, we have

$$\|\psi\| \leq \|\Theta_1 \Theta_{-1} \psi\| + \|R[EAE, E]\varphi\| \tag{2.14}$$

with $R \in S(h, g)$ by (2.8). On the other hand

$$\|\Theta_{-1}\psi\| \leq C (\|[EAE, E]\Theta_{-1}\varphi\| + \|\varphi\|) \tag{2.15}$$

since $[\Theta_{-1}, [EAE, E]] \in OPS(1; g)$.

So putting (2.13), (2.14) and (2.15) together we obtain the required estimate

$$\|[EAE, E]\varphi\|^2 \leq C\,(\,\|\Theta_{-1}\psi\| + \|\varphi\|\,)^2 \leq C_1\,(A\varphi, \varphi) + C_2\,\|\varphi\|^2\,.$$

It remains to give the proof of the lemma which is standard. Indeed

$$\|\sum J_\nu f_\nu\|^2 = \sum_{\nu,\mu}(K_{\mu\nu} f_\nu, f_\mu) \leq \|K\| \sum \|f_\nu\|^2$$

where $K_{\nu,\mu} = J_\mu J_\nu$ and $\|K\|$ is the operator norm of the "Hilbert space matrix" $(K_{\nu,\mu})_{\nu,\mu}$ acting on $L^2 \otimes \ell^2$. The boundedness of that norm is a consequence of the estimate

$$\|K_{\nu,\mu}\| \leq C\,(1 + d_{\nu\mu})^{-N}\,. \tag{2.16}$$

The proof of (2.16) can be found in [6], p. 168, just before relation (18.6.10) and in the few lines that follow. Observe that any of the standard proof that (2.16) implies the boundedness of the scalar valued matrix operator $(k_{\nu\mu})$ also works in the present vector valued case.

3. The Fefferman-Phong reduction of variables.

In this section, I shall give a proof of

$$\|[E(ea)^\omega E, E]f\|^2 \leq C_1\,(a^\omega f, f) + C_2\,\|f\|^2\,, \quad f \in C_0^\infty\,, \tag{3.1}$$

where $C_1, C_2 \geq 0$, E, e are as in Section 2 and $a(x, \xi) \in S(1/h^2; g)$ with $a \geq 0$. The main step of the proof is an inductive procedure (due to Fefferman-Phong) on the "essential" number of variables $x_1, \ldots, x_n$, $\xi_1, \ldots, \xi_n$ that appear in $a(\cdot, \cdot)$. To make this precise I shall say that $a(\cdot)$ depends only on k variables $0 \leq k \leq 2n$ if there exists $2n - k$ linear independent vectors $l_1, l_2, \ldots, l_{2n-k}$ in the (x, ξ) space $T(\mathbb{R}^{2n})$ such that $\mathrm{d}a(l_j) = 0$, $j = 1, 2, \ldots, 2n - k$ (*i.e.* a is constant along these vectors).

When a depends on 0 variables our estimate holds (the constants C_1, C_2 depend on a since (3.1) is not homogeneous in a !). Observe incidentally that the e inside $(ea)^\omega$ is imposed by technical reasons due to the above inductive procedure. In reality e can be absorbed in the E's outside since $E(ea)^\omega E \equiv Ea^\omega E$ mod-$S(1/h; g)$ and the $S(1/h; g)$ disappears after the commutator is taken.

Observe also that one situation in which our estimate (3.1) holds trivially is when $a = c^2$ with $c \in S(1/h; g)$ (real valued). Indeed, banal symbolic calculus shows in that case that

$$[E(ea)^\omega E, E] = E\, c^\omega + E\,.$$

It follows therefore that the left hand side of (3.1) can be estimated by $\|c^\omega f\|^2 + \|f\|^2$. On the other hand we have:

$$(c^\omega)^2 - (c^2)^\omega = a_1^\omega + a_2^\omega$$

with a_1 purely imaginary (in fact here we have $a_1 \equiv 0$!) and $a_2 \in S(1; g)$ (again by symbolic calculus). We can therefore estimate $\|c^\omega f\|^2 = ((c^\omega)^2 f, f)$ by $((c^2)^\omega f, f) + O(\|f\|^2)$ which gives our assertion.

The next observation is that in proving (3.1) we can reduce everything to the case when $g = g_0$ is a *constant* metric, *i.e.* a positive definite quadratic form on the $2n$ variables (x, ξ) for which $g/g^\sigma \leq \lambda^2$ where $0 < \lambda (= h) \leq 1$. This of course is the whole point of the localisation explained in [6, Lemma 18.4.4]. Indeed by what we did in the last section we see that our estimate (3.1) is "localisable" to each U_ν where the metric can be considered as constant.

More can in fact be demanded form the constant metric $g = g_0$. We can even ensure that $g = \lambda e$ where e is the euclidian metric $\sum(dx_i^2 + d\xi_i^2)$, and $0 < \lambda \leq 1$ as before. To see this we argue as in [6] in the first few lines of the proof of Lemma 18.6.10. Indeed we can, by a linear symplectic transformation T, reduce $g(x, \xi) = \sum \lambda_\nu (x_\nu^2 + \xi_\nu^2)$ with $\lambda = \sup \lambda_j$ and since our hypothesis on a, e is that $|a|_k^g \leq c_k \lambda^{-2}$, $|e|_k^g \leq c_k$, we can replace all the λ_j's by λ in the hypothesis. The estimate we wish to prove is

$$\|[e^\omega (ea)^\omega e^\omega, e^\omega] f\|^2 \leq C_1\, (a^\omega f, f) + C_2\, \|f\|^2\,, \quad f \in C_0^\infty\,.$$

The linear symplectic transformation has the following effect on the symbols and the corresponding operators

$$(e \circ T)^\omega = U^{-1} e^\omega U \quad , \quad (a \circ T)^\omega = U^{-1} a^\omega U$$

(*cf.* [6], Theorem 15.5.9) where $U : L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$ is a unitary transformation. It is then clear that the above conjugation operation commutes with all the "elements" of our estimate and we are done.

From everything that we have done up to now we see that the proof of (3.1) is reduced to the proof of the following inductive step for $k = 1, 2, \ldots, 2n$.

(I_k): *All the metrics are of the form* λe *for some* $0 < \lambda \leq 1$. *We shall assume that the estimate* (3.1) *holds if* a *depends on* $k-1$ *variables and we shall conclude that it holds for any* a *that depends on* k *variables.*

The proof is but a variant of the Fefferman-Phong argument. I shall follow closely the presentation given in [6] in the proof of Lemma 18.6.10.

Our original metric is $g = \lambda e$ and our conditions on a and e are (with apologies for the confusing notation!)

$$|e|_k^e \leq C_k \lambda^{k/2} , \quad |a|_k^e \leq C_k \lambda^{(k-4)/2}$$

(observe that all our estimates below have to be uniform in λ).

The metric g will be replaced by

$$G_{x,\xi} = mg = H(x,\xi)e$$

where

$$\frac{1}{H(\omega)} = \max\{1, a(\omega)^{1/2}, |a|_2^e(\omega)\} .$$

Since clearly by our hypothesis $H \geq \lambda$ we have $m \geq 1$ and clearly also $G/G^\sigma \leq H^2 \leq 1$. It follows therefore that to show that G is σ-temperate it suffices to show that it is slowly varying and invoque [6], Proposition 18.5.6 (or one can even give a direct proof, *cf.* [6]). The fact that G is slowly varying is proved in Hörmander (although I feel that the corresponding passage of the proof of Lemma 18.6.10 in [6] is unclear. Indeed I had to work somewhat to convince myself that it works ! Maybe the reader can do better). Be it as it may, we now have a σ-temperate metric G, and since $G \geq g$ we have $S(1;g) \subset S(1;G)$, but we also have $a \in S(1/H^2;G)$. To see this, since $m \geq 1$, it is enough to check that

$$|a|_k^e \leq C H^{(k-4)/2} , \quad k = 0,1,2,3$$

(compare with [6, relations (18.6.13)-(18.6.13)$''$]). For $k = 0,2$ this follows by the definition of H and for $k = 1,3$ by the standard log-convexity of the $\|\cdot\|_\infty$ norm of the derivatives ($\|F'\|_\infty^2 \leq C \|F\|_\infty \|F''\|_\infty$). What really has been done up to now is simply to transform all the data to a new metric $G = He$.

$$e \in S(1,G) , \quad a \in S(1/H^2, G) \tag{3.2}$$

and the λ has disappeared.

Our next step consists in a new localisation that will allow us to suppose that H is constant and so be able to use the inductive hypothesis. We consider a convering of the phase space by balls $U_\nu \subseteq U'_\nu$ as in (A) for the metric G. These balls are in fact *euclidean* balls centered around the points ω_ν of radius $c\,H_\nu^{-1/2}$ where $H_\nu = H(\omega_\nu)$ (for appropriate constants c).

Our strategy now is simple. We shall prove the estimate (3.1) when the e's ($e \in S(1;G)$) are such that supp $e \in U'_\nu$ for *some fixed* ν and when a only depends on k variables. This will be done under the inductive hypothesis that I_{k-1} holds. It is of course impossible (since incompatible with its constancy along certain directions) to assume that a also has supp $a \in U_\nu$ (hence the factor e in the $(ea)^\omega$ of our estimate).

There is one case that can be delt with immediately, this case is when $H_\nu = 1$. Indeed we then have $(ea) \in S(1; H_\nu e)$ and our estimate follows. It suffices therefore to analyse the case

$$1 \leq \max\{H_\nu^2 a(\omega_\nu), H_\nu |a|_2^e(\omega_\nu)\} \leq C\,. \tag{3.3}$$

The upper bound follows from (3.2). We consider then the function

$$f(z) = H_\nu^2 a(\omega_\nu + z/H_\nu^{1/2})$$

that satisfies $\max\{|f(0)|, |f|_2^e(0)\} \sim 1$ (in the sense of (3.3)). We shall apply to that function a slight variant (in the sense that the constants in (18.6.14) and (18.6.15) are different) of [6], Lemma 18.6.9. This allows us to decompose

$$f(x) = f_1(x) + g^2(x)\,, \quad |x| \leq C \tag{3.4}$$

with the appropriate bounds on the derivatives of f_1 and g, and f_1 depending only on $k-1$ variables. To see that we actually gain one extra direction of constancy, we have to apply, in fact, the proof of the lemma and not just the lemma itself.

Finally we shall go back to the original symbol a and cut it off by a function $\varphi_k \in S(1; H_\nu e)$ constant along the same directions as a. This allows us to define $\tilde{a} = a\varphi_k \in S(1/H_\nu^2; H_\nu e)$ globaly. The localised estimate that we wish to show refers to $(ea)^\omega$ and not to a^ω itself, so by properly choosing φ_k, we can replace a by $\tilde{a}$ on the left hand side of (3.1) without changing anything. Therefore, since

$$e \in S(1; H_\nu e)\,, \quad \tilde{a} \in S(H_\nu^{-2}; H_\nu e)\,, \tag{3.5}$$

we have succeeded in reducing everything a constant conformal metric again.

The decomposition of f in (3.4) induces then, by scalling back, a decomposition (*cf.* [6], p. 175)

$$\tilde{a} = b + c^2 \tag{3.6}$$

such that $b \in S(H_\nu^{-2}; H_\nu e)$, $c \in S(H_\nu^{-1}; H_\nu e)$ with b only depending on $k-1$ variables, we obtain therefore by the inductive hypothesis that

$$\begin{aligned} \|[e^\omega (ea)^\omega e^\omega, e^\omega] f\|^2 &= \|[e^\omega (e\tilde{a})^\omega e^\omega, e^\omega] f\|^2 \\ &\leq C\,(\tilde{a}^\omega f, f) + C\,\|f\|^2\,, \quad f \in C_0^\infty\,. \end{aligned} \tag{3.7}$$

To see (3.7), together with the induction hypothesis, we have to use the case $a = c^2$, that has already been delt with, and the Fefferman-Phong theorem (*cf.* [6], Theorem 18.6.8) that guarantees that

$$((c^2)^\omega f, f) \leq C\,(\tilde{a}^\omega f, f) + C\,\|f\|^2\,.$$

(The other estimate: $(b^\omega f, f) \leq C\,(\tilde{a}^\omega f, f) + C\,\|f\|^2$ is clear).

The estimate (3.7) is unfortunately not quite the wanted localised estimate. Indeed, we had to cut off the symbol a, and so we end up with $\tilde{a}$ and not a on the right hand side. (That cutting off was necessary to make (3.5), (3.6) work). It may well be that with a cleverer way of building up the induction I could have avoided that "misfiring". I propose to save the day differently. Indeed observe that the localised estimate is *only* used in Section 2 (*cf.* (2.11) and (2.11)′) for the special functions $f = I_\nu f$. To obtain our original symbol a on the right hand side of our estimate (3.1) it suffices therefore to be able to prove

$$|\sum_\nu (i_\nu^\omega (a - \tilde{a}_\nu)^\omega i_\nu^\omega f, f)| \leq C\,\|f\|^2\,, \qquad f \in C_0^\infty \tag{3.8}$$

where $\tilde{a}_\nu$ is the function $\varphi_k a = \tilde{a}$ for the index ν that was fixed just above. To see (3.8) we choose for each ν the corresponding "cutting off" function φ_k to be equal to 1 on some neighbourhood of supp i_ν. This choice makes the estimate (3.8) evident. Indeed by (C), (F), $i_\nu^\omega (a - \tilde{a}_\nu) i_\nu^\omega$ is then concentrated at ν in $S(1/h^2 h^N; g)$ for all ν and $N \geq 1$ (*cf.* also the considerations at the beginning of Section 2), and if we apply (E) we obtain (3.8). Observe incidentally that we do not have to prove

$\sum_\nu |(i_\nu^\omega(a-\tilde{a}_\nu)^\omega i_\nu^\omega f, f)| = O(\|f\|^2)$ and that we *can* put the modulus sign outside the summation. This is just as well because this stronger estimate would need a different proof.

4. The conjugation operators.

In this section, I shall only consider $0 \leq a(x,\xi) \in S^2_{1,0}$ a nonnegative "classical symbol" and $A = a^\omega(x,D)$. I shall show that

$$\|\Lambda^\alpha[\Lambda^{-\alpha}, A]f\|^2 \leq C_1\,(Af,f) + C_2\,\|f\|^2\,, \qquad f \in C_0^\infty \tag{4.1}$$

for appropriate $C_1, C_2 \geq 0$ and $\Lambda = (1+\Delta)^{1/2}$. We shall see that this follows easily from the results of sections 2 and 3.

The first step consists in a localisation of A at $\xi \sim 2^k$, $k = 1, 2, \ldots$ This is done as usual by a partition of unity of the form

$$1 \equiv \psi_0^3(\xi) + \sum_{j\geq 1} \psi^3(2^{-j}\xi)\,, \qquad \xi \in \mathbb{R}^n$$

where $\psi, \psi_0 \in C_0^\infty$ and

$$\text{supp } \psi \subset \left\{\xi : \quad \frac{1}{10} < |\xi| < 10\right\} \tag{4.2}$$

If we denote by $\varphi_0 = \psi_0$, $\varphi_k(\cdot) = \psi(2^{-k}\cdot)$, $A_k = (A\varphi_k)^\omega$ we have

$$A - \sum \varphi_k^\omega\, A_k\, \varphi_k^\omega = a_1^\omega + a_2^\omega \tag{4.3}$$

where $a_1 \in S^1_{1,0}$ takes pure imaginary values and $a_2 \in S^0_{1,0}$ (*cf.* [6], bottom of p. 174). Inserting (4.3) in our commutators we obtain

$$\begin{aligned}\|\Lambda^\alpha[\Lambda^{-\alpha}, A]f\|^2 &\leq \sum_k \|\Lambda^\alpha[\Lambda^{-\alpha}, \varphi_k^\omega A_k \varphi_k^\omega]f\|^2 + C\,\|f\|^2 \\ &= \sum_k \|\varphi_k^\omega \Lambda^\alpha[\Lambda^{-\alpha}, A_k]\varphi_k^\omega f\|^2 + C_1\|f\|^2\,, \quad f \in C_0^\infty\,.\end{aligned}$$

It follows therefore that it suffices to prove (4.1) for the localised symbols A_k. Indeed if that localised estimate is known to hold we obtain that

$$\|\Lambda^\alpha[\Lambda^{-\alpha}, A]f\|^2 \leq C\sum (A_k\varphi_k^\omega f, \varphi_k^\omega f) + C\sum \|\varphi_k^\omega f\|_2^2 + C\|f\|_2^2$$

when $f \in C_0^\infty$, which because of (4.3) gives the global result.

We shall suppose from now onwards that a is localised as above at $\xi \sim 2^{k_0} = \xi_0$ for some fixed $k_0 = 1, 2, \ldots$ Let us decompose then

$$\Lambda^\alpha = \check{\Lambda}^\alpha + \check{\Lambda}_\alpha\,, \quad \check{\Lambda}^\alpha = \left(\psi_1(\xi)(1+|\xi|^2)^{\alpha/2}\right)^\omega$$

where $\psi_1(\xi) = \psi(2^{-k_0}\xi)$ and where ψ satisfies (4.2) and is such that ψ_1 is equal to 1 on some neighbourhood of supp a (this last statement should be clear but it is abusive since a depends on ξ *and* on x).

We shall insert this decomposition in $\Lambda^\alpha[\Lambda^{-\alpha}, A]$. This will give rise to four different terms that have to be delt separately. For the first term we observe that (in terms of $S(m,g)$ notations) we have

$$\check{\Lambda}^{\pm\alpha} \in OPS(\xi_0^{\pm\alpha}; g_0)\,, \quad a \in S(\xi_0^2; g_0)$$

(uniformly in k_0) where $g_0 = dx^2 + \xi_0^{-2}d\xi^2$, so it follows that

$$\check{\Lambda}^\alpha[\check{\Lambda}^{-\alpha}, A] = (\xi_0^{-\alpha}\check{\Lambda}^\alpha)[(\xi_0^\alpha \check{\Lambda}^{-\alpha}), A]$$

where $\xi_0^{\pm\alpha}\check{\Lambda}^{\mp\alpha} \in S(1, g_0)$. This reduces the estimate to the corresponding result on $[E, A]$ examined in Section 3.

The other terms are very easy to estimate. Indeed

$$\text{supp } a \cap \text{supp (symb } \check{\Lambda}_\alpha) = \emptyset\,,$$

it follows therefore that the operators $\check{\Lambda}_\alpha A, A\check{\Lambda}_\alpha \in S_{1,0}^{-n}$, $(n \geq 0)$ for arbitrary n (by (C) among other things). Our estimate (4.1) is thus established.

5. The imaginary powers and the holomorphicity.

The operator A that we shall consider in this section is $A = a^\omega(x, D) + \lambda_0$ with $0 \leq a(x,\xi) \in S_{1,0}^2$ and λ_0 some appropriately large constant for A to be a positive Hilbert space operator. It follows from Section 4 that

$$\|(A - \Lambda^{-\alpha}A\Lambda^\alpha)f\|_{L^2} \leq C\,\|A^{1/2}f\|_{L^2} \tag{5.1}$$

provided that λ_0 is large enough to ensure that $\|f\|_{L^2} \leq C\,\|A^{1/2}f\|_{L^2}$. In fact in what follows I shall maintain the convention that was adopted

in [1] and drop the $\lambda_0 \geq 0$ altogether from all the formulas. (The convention is that this λ_0 is tacitly always there, and that it is large enough without necessarily appearing explicitely). An immediate consequence of the above estimate is that

$$\|e^{-tA} - \Lambda^{-\alpha} e^{-tA} \Lambda^{\alpha}\|_{L^2 \to L^2} = O(t^{1/2}) \qquad (\text{as } t \to 0)\,. \tag{5.2}$$

This was shown in [1] Section 3 and I shall not repeat the argument here since anyway these type of estimates will be examined in details in sections 6 and 7 below. I also wish to stress that from here onwards all the $O(t^a)$ notations that will appear refer to $t \to 0$ and that the "$t \to 0$" will usually be dropped. From (5.2) the boundedness of A^{is} on H_α $(s, \alpha \in \mathbb{R})$ follows as in [1, Section 3]. When A is a *differential* operator we can deduce from this the boundedness of

$$A^{is} : H^p_\alpha \longrightarrow H^p_\alpha\,; \qquad s \in \mathbb{R}, \quad 1 < p < \infty\,, \tag{5.3}$$

$H^p_\alpha = \{f :\ \Lambda^\alpha f \in L^p\}$. This is proved by interpolating the information between $(\alpha = 0,\ p = p_0)$ and $(\alpha = \alpha_0,\ p = 2)$. Once we have (5.1) all these facts extend to general A and the proofs of [1], Section 3 work in this general setting. In [1], Section 3 I also gave two distinct proofs of the holomorphicity of the action of e^{-tA} on the spaces H_α. The first works under very general conditions and does not use our basic estimate (5.1). The second used the action (5.3) of A^{is} on H_α. It turns out that if we make essential use of (5.1) we can give a direct proof of the fact that the operator $(1 + i\xi)A$, $(\xi \in \mathbb{R})$ is semibounded on each Hilbert space H_α, or equivalently that

$$\text{Re}\,(1 + i\xi)(\Lambda^{-\alpha} A \Lambda^{\alpha} f, f) \geq -C\,\|f\|^2_{L^2}\,. \tag{5.4}$$

Indeed the left hand side of (5.4) can be rewritten

$$(Af, f) + \text{Re}\,(1 + i\xi)((\Lambda^{-\alpha} A \Lambda^{\alpha} - A)f, f)$$

and because of (5.1) we can bound the second term by

$$C_1\,(1 + |\xi|)\,\|A^{1/2} f\|\,\|f\|\,.$$

It is therefore only a matter of choosing the C on the right hand side of (5.4) large enough. In fact the first proof of the holomorphicity of e^{-tA} (the one that does not use at all our main estimate (5.1)) will also

give the above semiboundedness. Indeed if we do that proof with care it will show that

$$\|e^{-zA}\|_{H_\alpha \to H_\alpha} \leq e^{\lambda|z|}\,, \quad |\text{Arg } z| \leq \theta \tag{5.5}$$

with any $0 \leq \theta < \pi/2$ and $\lambda > 0$ (depending on α and θ) and this is equivalent to (5.4), (*cf.* [7]). The fact that we have (5.5) rather that the coarser estimate $Me^{\lambda|z|}$ is of course not of great consequence. For *differential* operators the corresponding estimate (*i.e.* $M \equiv 1$) for the $H^p_\alpha \to H^p_\alpha$ norm also holds. This is seen by standard interpolation since it is well known that the semigroup e^{-tA} is symmetric submarkovian and therefore $\|e^{-tA}\|_{L^p \to L^p} \leq 1$, $(1 \leq p \leq \infty)$.

6. The square root $A^{1/2}$.

In this section I shall draw the first consequences of the two estimates that have been established in sections 1 to 4:

$$\|[A, E]f\|_X \leq C\,\|A^{1/2}f\|_X\,, \quad \|(c_\lambda - c_\mu)Af\|_X \leq C\,\|A^{1/2}f\|_X \tag{6.1}$$

when $f \in C_0^\infty$. Here $A \in S^2_{1,0}$ is as in Section 5, $c_\lambda(T) = \Lambda^\lambda T \Lambda^{-\lambda}$ is the conjugation operator applied to any operator T and $X = L^2$. We shall also denote $\|\cdot\| = \|\cdot\|_X$. This section relies very heavily on the methods, ideas and even notations of [1] and it would be unrealistic for the reader to try to read it without being familiar with [1].

The first consequence of (6.1) that I shall draw is

$$\|[A^{1/2}, E]f\|_X \leq C\,\|f\|_X\,, \quad \|(c_\lambda - c_\mu)A^{1/2}f\|_X \leq C\,\|f\|_X \tag{6.2}$$

when $f \in C_0^\infty$. For the proof I shall use the scale $X_\alpha = \{f :\ A^{\alpha/2}f \in X\}$, $(\alpha \in \mathbb{R})$. The norms $\|\cdot\|_\alpha$ and $\|\cdot\|_{\alpha\to\beta}$ will refer to that scale. To prove (6.2) I shall start by proving

$$\|[e^{-tA}, E]\|_{\alpha\to\beta}, \quad \|(c_\lambda - c_\mu)e^{-tA}\|_{\alpha\to\beta} = O\left(t^{1/2+(\alpha-\beta)/2}\right) \tag{6.3}$$

where $\alpha \in]-1, 1]$, $\beta \in [0, 2[$. Once we have established (6.3) the estimate (6.2) will follow automatically by the machinery of [1]. Indeed it would then follow from (6.3), by the real interpolation method of [1] Section 1, that

$$\|[A^\sigma, E]\|_{\alpha\to\beta} \leq C\,, \tag{6.4}$$

for $\alpha \in]-1,1[$, $\beta \in]0,2[$, $-\mathrm{Re}\,\sigma + \frac{1}{2} + \frac{\alpha-\beta}{2} = 0$ (the analogous estimate for $(c_\lambda - c_\mu)A^\sigma$ would also follow), so in particular we obtain

$$(6.4)' \quad \|[A^\sigma, E]\|_{\alpha\to\alpha}, \|(c_\lambda - c_\mu)A^\sigma\|_{\alpha\to\alpha} \le C\,, \quad \mathrm{Re}\,\sigma = \frac{1}{2},\ \alpha \in]0,1[\,.$$

By duality it follows that $(6.4)'$ also holds for $\alpha \in]-1,0[$ and thus, by interpolation, $(6.4)'$ also holds for $\alpha \in]-1,1[$. Our required estimate (6.2) is the case $\alpha = 0$ of (6.4). What follows from the above is, in fact, the stronger estimate where in (6.2) we replace $A^{1/2}$ by $A^{1/2+is}$ $(s \in \mathbb{R})$.

The proof of the basic estimates (6.3) can in fact be found in sections 4 and 5 of [1]. One simply has to run through the proof and observe that in the range $\alpha \in]-1,1]$, $\beta \in [0,2[$ the proof given there works under the assumption (6.1). For the convenience of the reader I shall recall the main points of the proof and I shall start with the easiest of the two estimates (which already contains the main idea). For $\alpha, \beta, \gamma \in \mathbb{R}$ we have

$$(6.5) \qquad \begin{aligned} &\|[e^{-tA}, E]\|_{\alpha\to\beta} \\ &\qquad \le \int_0^t \|e^{-(t-s)A}\|_{\gamma-1\to\beta}\|[E,A]\|_{\gamma\to\gamma-1}\|e^{-sA}\|_{\alpha\to\gamma}\,ds\,. \end{aligned}$$

To be able to exploit the above factorization we must have

$$\gamma \in [\alpha, \alpha+2[\,, \quad \beta \in [\gamma-1, \gamma+1[\,.$$

For the middle term we will use the following estimate

$$(6.6) \quad \|[E,A]\|_{\gamma\to\gamma-1} \le C\,, \quad \|(1-c_\lambda)A\|_{\gamma\to\gamma-1} \le C\,, \quad \gamma \in [0,1]\,.$$

Indeed the case $\gamma = 1$ of (6.6) is our hypothesis, the case $\gamma = 0$ is the dual statement and the values in between are obtained by interpolation. In fact in (6.5) we can just set $\gamma = 1$ (which is our hypothesis) and $\alpha \in]-1,1]$, $\beta \in [0,2[$ the integral in (6.4) gives then the required estimate as in Section 4 of [1]. The proof for $c_\lambda - c_\mu$ is more involved and is essentially contained in [1], Section 5. First for all by taking differences it is enough to consider the case $\mu = 0$. Let us then use the notations of [1], Section 5 and set

$$\varphi(t) = (1-c_\lambda)e^{-tA} = e^{-tA} - R_t\,.$$

I shall ignore the refinements of Section 5 in [1] and simply show that

$$\|\varphi(t)\|_{\alpha\to\alpha} = O(t^{1/2}), \quad \alpha \in [0,1]. \tag{6.7}$$

This will suffice to give us the estimate

$$\|R_t\|_{\alpha\to\alpha} = \|c_\lambda(e^{-tA})\|_{\alpha\to\alpha} = O(1), \quad \alpha \in [0,1]. \tag{6.8}$$

Once we have (6.8), we shall use the formula

$$\varphi(t) = \int_0^t e^{-(t-s)A}((1-c_\lambda)A)R_s\,ds \tag{6.9}$$

and the factorisation

$$\|\varphi(t)\|_{\alpha\to\beta} \le \int_0^t \|e^{-(t-s)A}\|_{\alpha-1\to\beta}\|(1-c_\lambda)A\|_{\alpha\to\alpha-1}\|R_s\|_{\alpha\to\alpha}\,ds$$

which together with (6.8) and (6.6) establishes (6.3) for $\alpha \in [0,1]$, $\beta \in [0,1[$. From this by the same method as before we finish the proof of (6.2). To establish (6.7) we use the same integral inequality as in Section 5 of [1]. We rewrite (6.9)

$$\varphi(t) = \int_0^t e^{-(t-s)A}(1-c_\lambda)A\varphi(s)ds + I(t),$$
$$I(t) = \int_0^t e^{-(t-s)A}(1-c_\lambda)Ae^{-sA}ds$$

and start by estimating

$$\|I(t)\|_{\alpha\to\alpha} = O(t^{1/2}), \quad \alpha \in [0,1].$$

This is done *exactly* as for $[E, e^{-tA}]$ (with $\alpha = \beta$). The next step is to fix some $f \in C_0^\infty$ in the unit ball of X_α and set $\psi(t) = \|\varphi(t)f\|_\alpha$. The function $\psi(t) \ge 0$ is such that $\psi(t) \to 0$ (as $t \to 0$) and satisfies the integral inequality

$$\psi(t) \le C\,t^{1/2} + C\int_0^t \|e^{-(t-s)A}\|_{\alpha-1\to\alpha}\|(1-c_\lambda)A\|_{\alpha\to\alpha-1}\psi(s)\,ds$$

so that we have

$$\psi(t) \le C\,t^{1/2} + C\int_0^t \psi(t-s)s^{1/2}\,ds, \quad \alpha \in [0,1].$$

This clearly implies the required estimate (6.7) just as in Section 5 of [1].

At this stage the scale X_α will be *abandoned* for good and the only information that will ve retained is at the level $\alpha = 0$, *i.e.* on the Hilbert space $X = L^2$ itself. This is exactly what was done in Section 6 of [1]. The scale we shall use from now onwards is the classical Sobolev scale

$$H_\alpha = \{f : \quad \Lambda^\alpha f \in L^2\}$$

and *from here onwards* right through the next section the norms $\|\cdot\|_\alpha$ and $\|\cdot\|_{\alpha\to\beta}$ will refer to that scale. Let us use the same notation as in Section 6 of [1] and set $B = \Lambda^s = (1+\Delta)^{s/2}$, $(s \in \mathbb{R})$. Then the estimate (6.1) of [1]

$$C\,\|A^\sigma f\|_X \le \|BA^\sigma B^{-1} f\|_X \le C\,\|A^\sigma f\|_X\,, \quad f \in C_0^\infty\,, \tag{6.10}$$

holds for $\sigma = 1$ and $\operatorname{Re}\sigma = 1/2$. The proof of this fact that we gave in Section 6 of [1] works because of (6.1) and (6.2). From (6.10) we deduce just as in Section 6 of [1] that

$$\begin{aligned} \|[A,E]f\|_\alpha &\le C\,\|A^{1/2} f\|_\alpha \\ \|(c_\lambda - c_\mu)Af\|_\alpha &\le C\,\|A^{1/2} f\|_\alpha \end{aligned} \tag{6.11}$$

when $f \in C_0^\infty$. We have thus generalised the estimate (6.1) to all the classical Sobolev norms. More generally just as in Section 6 of [1] we can deduce from (6.2), (6.10) and (6.11) that

$$\begin{aligned} \|S^{n_1} A^\sigma S^{n_2} f\|_m &\le C\,\|A^\sigma f\|_p \\ \|S^{n_1}[A^\sigma, S^{n_2}]S^{n_3} f\|_m &\le C\,\|A^{\sigma-1/2} f\|_p \end{aligned} \tag{6.12}$$

with $\sigma = 1$ or $\operatorname{Re}\sigma = 1/2$ and $p = m + \sum n_i$ and with $S^n \in OPS^n_{1,0}$ arbitrary pseudodifferential operators.

To illustrate (6.12) let us denote by $Q_t = e^{itA^{1/2}}$, $(t \in \mathbb{R}$ which is a group). We have then

$$Q_t - c_\lambda(Q_t) = c\int_0^t Q_{t-s}(1-c_\lambda)A^{1/2} c_\lambda(Q_s)\,dt$$

using then the same argument as in the beginning of this section and the fact $\|(1-c_\lambda)A^{1/2}\|_{\alpha\to\alpha} \le C$ (which is but a special case of (6.12)) we obtain that

$$\|Q_t - c_\lambda(Q_t)\|_{L^2\to L^2} = O(|t|)\,.$$

This in particular proves our last assertion in Section 0.1.

7. Commutators with E.

I strongly urge the reader (to help him get the idea) to read first the part of this paragraph that starts soon after relation (7.2) where two special cases are considered.

All the norms $\|\cdot\|_\alpha$ and $\|\cdot\|_{\alpha\to\beta}$ refer to the classical Sobolev scale $H_\alpha = \{f : \Lambda^\alpha f \in L^2\}$. The operator A is as in Section 6 and will be assumed subelliptic so that there exists $\delta \in [0,1[$ such that

$$\|f\|_{1-\delta} \leq C\,\|A^{1/2} f\|\,, \quad f \in C_0^\infty\,.$$

The letter δ will be reserved throughout to indicate that parameter.

We shall indicate multiple commutators throughout with the usual notation

$$[X, E_1, \dots, E_k] = [\dots [X, E_1], E_2, \dots], E_k]$$

with $E, E_1, \dots \in OPS^0_{1,0}$ and, as before, the same letter E will be reserved to indicate arbitrary $OPS^0_{1,0}$ which are not necessarily identical in different places. I shall also need the specific notation

$$\varphi(0) = 0\,, \quad \varphi(1) = \frac{1}{2}\,, \quad \varphi(2) = \varphi(3) = \dots = 1\,.$$

The following assertions (P_k) will be proved in this section inductively on $k = 0, 1, 2, \dots$.

Assertions (P_k):

$$(P'_k) \qquad \|[e^{-zA}, E_1, E_2, \dots, E_k]\|_{\alpha\to\beta} = O\left(|z|^{k/2+(\alpha-\beta)/2(1-\delta)}\right)$$

$$\text{where} \quad |\operatorname{Arg} z| < \frac{\pi}{4}\,, \quad \alpha, \beta \in \mathbb{R}, \quad \frac{k}{2} + \frac{\alpha-\beta}{2(1-\delta)} \leq \varphi(k)\,.$$

$$(P''_0) \qquad \|A^{1/2}\|_{\alpha\to\alpha-1} \leq C\,, \quad \alpha \in \mathbb{R}\,.$$

$$(P''_k) \quad \|[A^{1/2}, E_1, E_2, \dots, E_k]\|_{\alpha\to\alpha+(k-1)(1-\delta)} \leq C\,, \quad k \geq 1, \alpha \in \mathbb{R}\,.$$

Two more conditions will be considered in the induction (z always lies in the sector $|\mathrm{Arg}\ z| < \pi/4$)

$$(P_k''') \qquad \begin{aligned} &\|[A^{1/2}e^{-zA}, E_1, E_2, \dots, E_k]\|_{\alpha\to\beta} \\ &\qquad = O\left(|z|^{k/2+(\alpha-\beta)/2(1-\delta)-1/2}\right) \end{aligned}$$

$$\text{where} \quad \frac{k}{2} + \frac{\alpha-\beta}{2(1-\delta)} < \frac{1}{2}, \quad \frac{k}{2} + \frac{\alpha-\beta}{2(1-\delta)} \le \varphi(k).$$

$$(P_k^{(\mathrm{iv})}) \quad \|A^{1/2}[e^{-zA}, E_1, \dots, E_k]\|_{\alpha\to\beta} = O\left(|z|^{k/2+(\alpha-\beta)/2(1-\delta)-1/2}\right)$$

$$\text{where} \quad \frac{k}{2} + \frac{\alpha-\beta}{2(1-\delta)} < \frac{1}{2}, \quad \frac{k}{2} + \frac{\alpha-\beta}{2(1-\delta)} \le \varphi(k).$$

A few more obvious remarks are in order. (P_0'), (P_0'') are contained in [1] and [8]. (P_1'') is just our estimate (6.2). Also for $k = 0$ the two statements (P_0''') and $(P_0^{(\mathrm{iv})})$ are identical and are automatic consequences of the holomorphicity of the action of e^{-tA} on the Sobolev spaces H_α. Furthermore observe that for any $k \ge 1$ the statements (P_j'') in conjunction with the statements (P_j'''), $j \le k$, implies $(P_k^{(\mathrm{iv})})$. Graphically

$$(P_j'') \oplus (P_j'''), \quad 0 \le j \le k \quad \text{implies } (P_k^{(\mathrm{iv})}).$$

This is only a matter of developping out the commutator in (P''').

The next observation is less obvious and says that for $k = 2, 3, \dots$

$$(P_k') \quad \text{implies} \quad (P_k'').$$

In fact something more general holds: for $k \ge 2$ under the assumption that (P_k') holds (only needed for $z = t \ge 0$) we have

$$(7.1) \qquad \|[A^\sigma, E_1, E_2, \dots, E_k]\|_{\alpha\to\alpha+k(1-\delta)-2\mathrm{Re}\ \sigma(1-\delta)} \le C$$

for all $\alpha \in \mathbb{R}$ and $\mathrm{Re}\ \sigma < 1$. This follows from our interpolation theorem of Section 1 in [1] applied to the scale H_α and

$$\Phi(t) = [e^{-tA}, E_1, E_2, \dots, E_k] = [e^{-tA} - I, E_1, E_2, \dots, E_k].$$

So that for Re $\sigma < 1$, Re $\sigma \neq 0$, we have

$$[A^{\sigma}, E_1, \ldots, E_k] = c \int_0^{\infty} t^{-\sigma-1} \Phi(t) dt .$$

Indeed for $\beta = \alpha + k(1-\delta) - 2\text{Re}\,\sigma(1-\delta)$ we have $k/2 + (\alpha - \beta)/2(1-\delta) = \text{Re}\,\sigma < 1$, and that last strict inequality gives us the "room" that we need to play, for the interpolation of Section 1 in [1]. The case Re $\sigma = 0$ of (7.1) has to be delt separately but it can be deduced from Re $\sigma \neq 0$ by complex interpolation (applied to an analytic family of operators).

I shall finally show that for $k = 1, 2, \ldots$

$$(P'_k) \quad \text{implies} \quad (P'''_k) .$$

The proof relies on the fact that the function

$$F(z) = [e^{-zA}, E_1, E_2, \ldots, E_k] , \quad |\text{Arg } z| < \frac{\pi}{4}$$

is an operator valued holomorphic fuction. It follows therefore from Cauchy's Theorem and (P'_k) that

$$\left\| \left[\frac{d}{dz} e^{-zA}, E_1, \ldots, E_k \right] \right\|_{\alpha \to \beta} = O\left(|z|^{k/2 + (\alpha-\beta)/2(1-\delta) - 1} \right)$$

when $k/2 + (\alpha - \beta)/2(1-\delta) \leq \varphi(k)$. On the other hand we have

$$[A^{1/2} e^{-zA}, E_1, \ldots, E_k] = \int_0^{\infty} s^{-1/2} \frac{d}{ds} [e^{-(s+z)A}, E_1, \ldots, E_k] ds$$

and therefore

$$\begin{aligned} \|[A^{1/2} e^{-zA}, E_1, \ldots, E_k]\|_{\alpha \to \beta} & \\ & \leq C \int_0^{\infty} s^{-1/2} |s+z|^{k/2 + (\alpha-\beta)/2(1-\delta) - 1} ds \\ & = O\left(|z|^{k/2 + (\alpha-\beta)/2(1-\delta) - 1/2} \right) \end{aligned}$$

Provided that we have $k/2 + (\alpha - \beta)/2(1-\delta) < 1/2$ (the *strict* inequality (less than $1/2$) is needed to give uniform bounds at infinity. I do not know if this inequality has to be strict or whether it can be relaxed

to less than or equal to 1/2. But on the other hand this will be of no consequence at this point).

The upshot of all the above considerations is that in the proof of the inductive step of P_k it suffices simply to prove the step

$$(P'_{k-1}) \quad \text{implies} \quad (P'_k) \tag{7.2}$$

and in the proof of that step (7.2) I am allowed to use all the information contained in $(P_j)_{0\leq j\leq k-1}$.

To simplify notations from here onwards I shall drop the complex variable z ($|\text{Arg } z| < \pi/2 - \varepsilon_0$) and consider only $z = t > 0$. The proofs are identical for complex z. I also urge the reader at this point to study Section 4 of [1] since otherwise he will find it difficult to understand the considerations that follow. Let us first consider simple commutators. We have

$$\begin{aligned}
&\|[e^{-2tA}, E]\|_{\alpha\to\beta} \\
&\quad \leq \int_0^t \|e^{-(2t-s)A}A^{1/2}\|_{\alpha\to\beta}\|A^{-1/2}[E,A]\|_{\alpha\to\alpha}\|e^{-sA}\|_{\alpha\to\alpha}ds \\
&\qquad + \int_0^t \|e^{-(t-s)A}\|_{\beta\to\beta}\|[E,A]A^{-1/2}\|_{\beta\to\beta}\|A^{1/2}e^{-(t+s)A}\|_{\alpha\to\beta}ds\,.
\end{aligned}$$

The two middle "factors" inside the integrals are adjoined of each other and are bounded (*cf.* (6.11)). The other terms can be estimated by the results of [8] and the holomorphicity of the semigroup e^{-tA} on H_α. Putting everything together we obtain

$$\|[e^{-tA}, E]\|_{\alpha\to\beta} = O\left(t^{1/2+(\alpha-\beta)/2(1-\delta)}\right)\,, \quad \beta \geq \alpha\,.$$

At this stage I could embark in the proof of the general inductive step (7.2). What is involved there however hides the main idea of the proof. To help the reader understand what is going on, I propose to prove "ad hoc" (P'_2) (*i.e.* $k = 2$), and then perform the general inductive step. In fact if the reader is a "believer" he could skip in a fist reading the proof of that general inductive step. Let us examine commutators of order 2 where we shall expand $[e^{-2tA}, E, E]$ into six integrals as in Section 4 of [1]

$$\int_0^t e^{\cdot\cdot A}[A,E,E,]e^{\cdot\cdot A}\,, \qquad \int [e^{\cdot\cdot A}, E][A,E]e^{\cdot\cdot A}\,, \tag{7.3}$$

$$\int e^{\cdot\cdot A}[A,E][e^{\cdot\cdot A}, E]\,.$$

The ".." on $e^{..A}$ at the two ends of the integrals indicate the two combinations $-(2t-s)$ and $-s$ or $-(t-s)$ and $-(s+t)$ and they will be needed to perform the jump $\alpha \to \beta$ at one end or the other. The above integrals give corresponding "factorisations" of the $\|[e^{-tA}, E, E]\|_{\alpha\to\beta}$ norm and will be delt one at a time. For the first we proceed as follows

$$\int_0^t \|e^{-(2t-s)A}\|_{\alpha\to\beta}\| \cdot \|_{\alpha\to\alpha}\|e^{-sA}\|_{\alpha\to\alpha}$$
$$\int_0^t \|e^{-(t-s)A}\|_{\beta\to\beta}\| \cdot \|_{\beta\to\beta}\|e^{-(t+s)A}\|_{\alpha\to\beta}$$

and since $[A, E, E,] \in OPS^0_{1,0}$, its $\|\cdot\|_{\gamma\to\gamma}$ norm is bounded and the contribution of both of these two integrals is $O\left(t^{1+(\alpha-\beta)/2(1-\delta)}\right)$ for $\beta \geq \alpha$ (the results of [8] have to be used again).

The second integral in (7.3) gives rise to the factorisation

$$\int \|[e^{..A}, E]\|_{\gamma\to\beta}\|[A, E]A^{-1/2}\|_{\gamma\to\gamma}\|A^{1/2}e^{..A}\|_{\alpha\to\gamma}$$

where the γ is either α or β depending on the combination that we have adopted, $\{-(2t-s)A; -sA\}$ or $\{-(t-s)A; -(t+s)A\}$ of $\{..;..\}$. The norm $\|[A, E]A^{-1/2}\|_{\gamma\to\gamma}$ is bounded by (6.11), and using our previous result on the simple commutator $[e^{-tA}, E]$, we obtain again the contribution $O\left(t^{1+(\alpha-\beta)/2(1-\delta)}\right)$ (observe that there is a "loss" of $1/2$ at one end, $A^{1/2}e^{..A}$, but a "gain" of $1/2$ at the other end, $[e^{..A}, E]$.

The final integral in (7.3) is of course dual to the one we just considered. We put everything together and we conclude therefore that

$$\|[e^{-tA}, E, E]\|_{\alpha\to\beta} = O\left(t^{1+(\alpha-\beta)/2(1-\delta)}\right)$$

for $\beta \geq \alpha$. This is our statement (P_2'').

To work out the proof of the general inductive step (7.2) we shall introduce the following

$$K_k(t) = [e^{-tA}, E_1, \ldots, E_k], \quad k = 0, 1, \ldots, \quad t > 0$$

and use the analogue of our previous formulas to decompose $K_k(t)$ into a number of integrals of the form

$$I_{p,q,r} = \int_0^t K_p(..)[A, E_1, \ldots, E_q]K_r(..)ds \tag{7.4}$$

where $q \geq 1$, $p+q+r=k$ and where the combination "..,.." at the two ends of the integral is as before either $2t-s, s$ or $t-s, t+s$. We shall assume that $(P_j)_{j\leq k-1}$ holds and we shall distinguishes two cases in (7.4).

Case (i): $q \geq 2$. Let

$$a = \frac{k}{2} + \frac{\alpha-\beta}{2(1-\delta)} \leq \varphi(k) = 1 \tag{7.5}$$

since we may suppose that $k \geq 2$. I shall introduce $\alpha', \beta' \in \mathbb{R}$ such that

$$x = \frac{r}{2} + \frac{\alpha-\alpha'}{2(1-\delta)} \leq \varphi(r)\,, \quad y = \frac{p}{2} + \frac{\beta'-\beta}{2(1-\delta)} \leq \varphi(p)\,, \tag{7.6}$$

$$z = \frac{q}{2} + \frac{\alpha'-\beta'}{2(1-\delta)} \leq \varphi(q) = 1\,,$$

$$\beta' \leq \alpha' + q - 2 \quad \text{or equivalently} \quad \frac{2-q\delta}{2-2\delta} \leq z\,. \tag{7.7}$$

The compatibility of the above conditions will be examined shortly. It is then possible to estimate

$$\|I_{p,q,r}\|_{\alpha\to\beta} \leq \int_0^t \|K_p(..)\|_{\beta'\to\beta}\| \cdot \|_{\alpha'\to\beta'}\|K_r(..)\|_{\alpha\to\alpha'} ds$$

where the middle term is bounded (because of (7.7)) and for the other two terms we can use the inductive hypothesis. It is necessary in the above to make sure that the integral converges at the two ends. Only one of the two ends will be a problem, and which one of the two ends will give trouble depends on the choice of combination "..;..". To ensure the convergence of the integral we must impose therefore the additional condition

$$x > -1 \quad \text{(respectively, } y > -1\text{)} \tag{7.8}$$

(the "respectively" refers of course to the choice of "..;..").

Assuming that the conditions (7.6), (7.7) and (7.8) are verified, we obtain then that (*cf.* (7.5))

$$\|I_{p,q,r}\|_{\alpha\to\beta} = O\left(t^{1+x+y}\right) = O\left(t^{x+y+z}\right) = O\left(t^a\right)$$

provided that $z \leq 1$, which proves the statement (P'_k). To prove the compatibility of our conditions observe that it suffices to find $x, y \in \mathbb{R}$ such that

$$x + y + z = a \leq 1\,, \quad \frac{2 - q\delta}{2 - 2\delta} \leq z \leq 1\,,$$
$$x \leq \varphi(r)\,, \quad y \leq \varphi(p)\,,$$
$$-1 < x\,, \quad (\text{respectively, } -1 < y)\,.$$

For indeed α', β' can then be determined to satisfy the *three* equations (7.6) (since $p + q + r = k$).

The above conditions on (x, y, z) are clearly compatible. It suffices to set

$$z = 1 \quad \text{and} \quad (x, y) = (0, a - 1) \quad (\text{respectively } (a - 1, 0))\,.$$

Case (ii): $q = 1$. We shall also assume without loss of generality that $p \geq 2$. Indeed one of the two p or r is larger than or equal to 2, since $k \geq 3$, and we can pass from one to the other by considering the adjoint operator. We proceed then as follows

$$\|I_{p,q,r}\|_{\alpha \to \beta} \leq \int_0^t \|K_p(..)\|_{\gamma \to \beta} \|[A, E]A^{-1/2}\|_{\gamma \to \gamma} \|A^{1/2} K_r(..)\|_{\alpha \to \gamma} ds$$

with the same meaning to the notation "..;..". We shall choose the $\gamma \in \mathbb{R}$ so that

$$x = \frac{r}{2} + \frac{\gamma - \alpha}{2(1 - \delta)} \leq \varphi(r)\,, \quad x < \frac{1}{2}\,, \tag{7.9}$$
$$y = \frac{p}{2} + \frac{\beta - \gamma}{2(1 - \delta)} \leq \varphi(p) = 1\,,$$

and

$$-1 < x \quad (\text{respectively, } -1 < y)\,, \tag{7.10}$$

with the same meaning as before for the "respectively" (it depends on the choice of "..;.." which is necessary to make the integral converge). When (7.9) and (7.10) are verified we can integrate and we obtain the required inductive step

$$\|I_{p,q,r}\|_{\alpha \to \beta} = O\left(t^{x+y-1/2+1}\right) = O\left(t^{k/2 + (\alpha - \beta)/2(1-\delta)}\right)\,.$$

To check the compatibility set

$$x + y = \frac{k}{2} + \frac{\alpha - \beta}{2(1-\delta)} - \frac{1}{2} = a \leq \varphi(k) - \frac{1}{2} = \frac{1}{2}\,.$$

It is enough to choose $x, y \in \mathbb{R}$ so that

$$x < \frac{1}{2}, \quad x \leq \varphi(r), \quad y = a - x \leq 1$$

and also

$$-1 < x \quad \text{(respectively, } -1 < y\text{)}$$

for then $\gamma \in \mathbb{R}$ can be determined to satisfy (7.9).

The compatibility of the above x, y conditions is clear, indeed it suffices to set $x = 0$ (respectively $x = 0$ if $a \in [0, 1/2]$ or $x = a$ if $a < 0$).

In the following proposition we collect together some important information obtained up to now.

Proposition. *Let $\sigma \in \mathbb{C}$ with $\operatorname{Re}\sigma \leq 0$, $\alpha \in \mathbb{R}$, $k = 0, 1, \ldots$ Then for commutators of length k we have*

$$\|[A^\sigma, E, E, \ldots, E]\|_{\alpha \to \alpha + k(1-\delta) - 2\operatorname{Re}\sigma(1-\delta)} \leq C\,.$$

The proof was given in (7.1) for $k \geq 2$ and $\operatorname{Re}\sigma < 1$. It is very easy to see that the same proof works for $k = 1$, $\operatorname{Re}\sigma < 1/2$ and $\operatorname{Re}\sigma < 0$, $k = 0$. For $\operatorname{Re}\sigma = 0$, $k = 0$ the Proposition holds by (5.3). Observe finally that the above estimate also holds for $\operatorname{Re}\sigma = 1/2$, $k = 1$ (*cf.* (6.12), and the remark a couple of lines after (6.4)) but we shall have no use of the cases $\operatorname{Re}\sigma > 0$ in what follows.

8. General commutators and the classes $S^m_{\rho,\delta}$.

In this section A will denote a general linear operator $A : C_0^\infty \to \mathcal{D}'$ (to avoid necessary complications I shall also suppose, when necessary, that A is "compactly supported" in the sense that there exists some compact set K such that $A(\varphi) \equiv 0$ if $\operatorname{supp}\varphi \cap K = \emptyset$ and $A(\varphi) \equiv 0$ outside K, $\forall \varphi \in C_0^\infty$). I shall denote as usual by $E \in S^0_{1,0}$ and also by

$$\mathfrak{a}_k = \mathfrak{a} = [A, E, E, \ldots, E] = [\ldots[[A, E], E], \ldots]$$

where k is the length of the commutators. The E's are as usual $E \in OPS^0_{1,0}$ not necessarily all the same.

Our standing hypothesis in this section will be that

$$\|\mathfrak{a}_k\|_{\alpha\to\alpha+\delta k-m} \leq C\,, \quad \alpha \in \mathbb{R}, \quad k = 0, 1, \ldots \tag{8.1}$$

for some fixed $0 < \delta \leq 1$, $m \in \mathbb{R}$. $\|\cdot\|_\alpha$ and $\|\cdot\|_{\alpha\to\beta}$ refer to the standard Sobolev norms.

To present the arguments in this section it is necessary to establish a good set of notations. The basis of our reasoning is the classical decomposition of unity

$$1 = \sum_{j=1}^{\infty} \psi_j(\xi)\,, \quad \psi_0 \in C_0^\infty\,, \quad \psi_j(\xi) = \psi^N(2^{-j}\xi)\,, \quad \xi \in \mathbb{R}^m \tag{8.2}$$

for some $\psi \in C_0^\infty$ with supp $\psi \subset \{\xi :\ 1/10 < |\xi| < 10\}$ and where the power N will be important because it will allow us to decompose the corresponding components into arbitrarily many factors. The $N \geq 1$ will be chosen at the beginning and appropriately large. The partition (8.2) will be used to decompose

$$\Lambda^\alpha = \sum_{j\geq 1} 2^{\alpha j} E_j + E_0\ . \tag{8.3}$$

In (8.3) and in what follows, I shall reserve throughout the notation $E_j \in S^0_{1,0}$ for operators indexed by $j \geq 1$ that will satisfy several properties which will be enumerated below. It is important to understand that although all the E_j's have these properties they are not in general identical when they appear in different places. This notational convention gives us great flexibility in the arguments. Observe finally that E_0 that comes from ψ_0 is special, and will often enough be ignored since it never causes any trouble. All the properties below will be satisfied uniformly in the indices when the case arises.

(i) $E_j = e_j(D)\,, \quad \text{supp } e_j \subset \{\xi :\ 2^j/K < |\xi| < K\,2^j\}$

(the $K \gg 1$ can vary from place to place but does not depend on j).

(ii) $2^{\alpha j} E_j \in S^\alpha_{1,0}$,

I shall also adopte the notation $\lambda_j = 2^j$, $j \geq 0$.

(iii) $\sum \sigma_j \lambda_j^\alpha E_j \in S^\alpha_{1,0}$ for arbitrary $(\sigma_j)_{j\geq 0} \in l^\infty$.

This is an automatic consequence of (i) and (ii) and it implies that

$$\|\sum \sigma_j \lambda_j^\alpha E_j\|_{\beta+\alpha\to\beta} \le C\,, \quad \alpha,\beta\in\mathbb{R} \tag{8.4}$$

(iv) Each e_j and therefore each E_j can be factored into as many e_j's (respectively E_j's) as we need $E_j = E_jE_j\dots E_j$.

This simply come from the large exponent N in (8.2).

(v) For arbitrary $(\sigma_j)\in l^\infty$ and $\alpha,\beta\in\mathbb{R}$ we have

$$\|\sum \sigma_j\lambda_j^\alpha E_j f_j\|_\beta^2 \le C\sum \|f_j\|_{\alpha+\beta}^2\,, \quad f_j\in C_0^\infty\,. \tag{8.5}$$

Indeed let $F=\sum\sigma_j\lambda_j^\alpha E_jf_j$, $\varphi_j=\Lambda^{\alpha+\beta}f_j$ we have $\Lambda^\beta F=\sum\sigma_jE_j\varphi_j$ (the new E_j is of course different!) It suffices therefore to prove our assertion for $\alpha=\beta=0$. But then $\|F\|^2=\sum_{j,k}(E_jE_k\varphi_k,\varphi_j)$ and by (i) we can estimate this by $\sum\|E_j\varphi_j\|^2$. This gives our assertion.

(vi) Using the fact that $E_j=E_j^2$ (for a different E_j !) we can deduce from (8.5) (simply set $f_j=E_jf$) that

$$\|\sum\sigma_j\lambda_j^\alpha E_jf\|_\beta^2 \le C\sum\|E_jf\|_{\alpha+\beta}^2\,. \tag{8.6}$$

We also have

$$\sum\lambda_j^\alpha\|E_jf\|_\beta^2 \le C\,\|f\|_{\alpha+\beta}^2\,. \tag{8.7}$$

To see this, we set $F_\sigma=\sum\sigma_j\lambda_j^\alpha E_jf$ so that $\|F_\sigma\|_\beta\le C\,\|f\|_{\alpha+\beta}$, uniformly in σ, by (8.4). If we take expectations over $\sigma_j=\pm1$, (8.7) follows.

At this point let me recall that for ±1 centered independent random variables $\zeta_j,\xi_j,\eta_j,\dots$ and $h_{i,j,k,\dots}\in X$ (=some Hilbert space) we have

$$E\|\sum(\zeta_i\xi_j\eta_k\dots)h_{i,j,k,\dots}\|_X^2 \sim \sum\|h_{i,j,k,\dots}\|_X^2\,. \tag{8.8}$$

This is standard. (What is slightly less standard is that we have the one sided inequalities for $X=L^p(\Omega)$, $1\le p\le 2$. We have to make essential use of this refinement if we want to develop the L^p-theory of these operators).

The following terminology will now be used. We shall say that $T : C_0^\infty \to \mathcal{D}'$ is of smoothing order $\leq n \in \mathbb{R}$ (or simply "is of order n" if no confusion arises) if

$$\|Tf\|_\alpha \leq C \|f\|_{\alpha+n} , \quad \alpha \in \mathbb{R} . \tag{8.9}$$

In this terminology our operators $\mathfrak{a}_k = [A, E, E, \ldots, E]$, $k \geq 0$ are of order $\operatorname{ord}(\mathfrak{a}_k) = m - k\delta$. Clearly when T is of order n so is its adjoined T^*. We have then

(vii) Let $\mathfrak{a}_p = [A, E, E, \ldots, E]$ be as above. Then for every fixed $q \in \mathbb{R}$ the following two operators (adjoined of each other)

$$\sum_j \lambda_j^q E_j [E_j, \ldots, E_j, \mathfrak{a}] , \qquad \sum_j \lambda_j^q [E_j, \ldots, E_j, \mathfrak{a}] E_j \tag{8.10}$$

are of order $m + q - (p + n)\delta$. Here n is the number of E_j's inside the brackets of (8.10).

Indeed from the above observations it follows that it suffices to consider the first operator and from (8.6) it follows that it suffices to prove that for all n and α we have

$$\sum_j \|[E_j, E_j, \ldots, E_j, \mathfrak{a}_p] f\|_\alpha^2 \leq C \|f\|_{\alpha+m-(n+p)\delta}^2 \tag{8.11}$$

when $f \in C_0^\infty$, $\alpha \in \mathbb{R}$, $p = 0, 1, \ldots$. Here n indicates, as before, the number of E_j's inside the bracket. To prove this estimate we consider

$$E_\zeta = \sum \zeta_i E_i , \quad E_\xi = \sum \xi_i E_i , \ldots \in S_{1,0}^0$$

where $\zeta_i, \xi_i, \ldots$ are independent ± 1 random variables as above. Taking expectations and using (8.8) we obtain

$$\begin{aligned} \sum_{j_1, \ldots, j_n} \|[E_{j_1}, E_{j_2}, \ldots, E_{j_n}, \mathfrak{a}_p] f\|_\alpha^2 &\leq C \sup_{\zeta, \xi, \ldots} \|[E_\zeta, E_\xi, \ldots, \mathfrak{a}_p] f\|_\alpha^2 \\ &\leq C \|f\|_{\alpha-(p+n)\delta+m}^2 \end{aligned}$$

where for the second inequality we use our hypothesis (8.1). The above estimate contains (8.11).

We now come to the main estimate of this section:

Let $\mathfrak{a} = \mathfrak{a}_p = [A, E, E, \ldots, E]$, $(p \geq 0)$ as before and let $\alpha_1, \ldots, \alpha_k \in \mathbb{R}$ $(k \geq 0)$. We shall prove that the operator

$$[\Lambda^{\alpha_1}, \Lambda^{\alpha_2}, \ldots, \Lambda^{\alpha_k}, \mathfrak{a}] = B \tag{8.12}$$

is of order $\alpha_1 + \cdots + \alpha_k - (p+k)\delta + m$.

To prove this fact, I shall partition each Λ^α as in (8.3) (and also $I = \sum E_i$) and I shall write

$$B = \sum_{i, j_1, \ldots, j_k} \lambda_{j_1}^{\alpha_1} \ldots \lambda_{j_k}^{\alpha_k} E_i [E_{j_1}, \ldots, E_{j_k}, \mathfrak{a}] .$$

I shall decompose the above summation into two parts. The first comes from terms for which $i \sim j_1 \sim \ldots \sim j_k$, *i.e.* equal up to a fixed constant. The contribution we obtain then is

$$B' = \sum_i \lambda_i^{\sum \alpha_i} E_i [E_i, \ldots, E_i, \mathfrak{a}]$$

and using (vii) we see that B' has the correct order. In the second summation, since the j_r's are interchangeable (all the E_k's commute !) we may suppose that $|j_1 - i| \geq C \gg 1$. This gives the following contribution (In the argument below I make essential use of the fact that the E_k's commute. On the other hand even if the E_k's did not commute we could make this argument work by considering higher commutators)

$$\begin{aligned} B'' &= \sum_{j_2, \ldots, j_k} \lambda_{j_2}^{\alpha_2} \ldots \lambda_{j_k}^{\alpha_k} [E_{j_2}, \ldots, E_{j_k}, \sum_i E_i [\sum_{|j-i| \geq C} \lambda_j^{\alpha_1} E_j, \mathfrak{a}]] \\ &= [S_2, S_3, \ldots, S_k, M] \end{aligned} \tag{8.13}$$

where $S_r \in S_{1,0}^{\alpha_r}$, $r = 2, \ldots, k$ and

$$M = \sum_{|j-i| \geq C} \lambda_j^{\alpha_1} E_i [E_j, \mathfrak{a}] . \tag{8.14}$$

Using the fact that each E_j can be written as E_j^N and also the fact that $E_i E_j = 0$ for $|j - i| \geq C$ we deduce that the general term in the summation (8.14) can be replaced by

$$\lambda_j^{\alpha_1} E_i [E_j, E_j, \ldots, E_j, \mathfrak{a}] E_j$$

with as many E_j's as we need inside the bracket. We conclude therefore that

$$M = \sum_{|i-j|\geq C} \lambda_j^{\alpha_1} E_i [E_j, \ldots, E_j, \mathfrak{a}] E_j \,.$$

Summing first over i and observing that $\sum_{|i-j|\geq C} E_i = E - E_j$ we deduce that

$$M = E(\sum_j \lambda_j^{\alpha_1} [E_j, \ldots, E_j, \mathfrak{a}] E_j) + \sum_j \lambda_j^{\alpha_1} E_j [E_j, \ldots, E_j, \mathfrak{a}] E_j \tag{8.15}$$

but in the second summation we can absorb the E_j on the right by introducing an *extra commutator.*

Putting together (8.13), (8.14) and (8.15) we finally see that it is a consequence of (vii) that M and thus B'' have as low an order as we like (*i.e.* they are infinitely regularising. Indeed it is only a matter of taking the length of the brackets in (8.15) high enough). This proves our assertion.

Let us now consider arbitrary $S_j \in OPS^{n_j}_{1,0}$. The final claim that I will be made in this section concerns (always under the hypothesis (8.1)) the smoothing order of the following commutator

$$C_p = [A, S_1, \ldots, S_p] = [\ldots [A, S_1], S_2] \ldots S_p] \,, \tag{8.16}$$

$$\text{Smoothing order } C_p \leq \sum n_j + m - \delta k \,. \tag{8.17}$$

Here of course the smoothing order is defined as in (8.9). This statement will be proved by induction on the length p. It clearly holds for $p = 0$. I shall assume it to hold up to $p-1$, and proceed to prove the inductive step.

Towards that I start by factorising each $S_j = E\Lambda^{n_j} = \Lambda^{n_j} E$ and expand the commutators (8.16). What is obtained by that expansion is a linear combination of terms

$$P[A, T_1, T_2, \ldots, T_k] Q + \ldots \tag{8.18}$$

where $T_1, T_2, \ldots, T_j = E$, I shall then say that ord $T_i = t_i = 0$ $(1 \leq i \leq j)$, and $T_r = \Lambda^{t_r}$ and say ord $T_r = t_r$, $(j+1 \leq r \leq k)$. Furthermore $P \in OPS^p_{1,0}$, $Q \in OPS^q_{1,0}$.

The important point is that $p + q + \sum t_i \leq \sum n_i$. This is obvious because in the various monomials that appear in the decomposition of C_p there is *no way* at all that we can increase the total order of the

pseudodifferentials. The first term in (8.18) has therefore the required smoothing order by the corresponding statement on (8.12) (the reader has to make here the distinction between the smoothing order and the order of a pseudodifferential).

It remains to examine the remainder terms "..." in (8.18). These are the terms for which the "principal commutator" has length $p' < p$, and they look like

$$P[A, S'_1, \dots, S'_{p'}]Q$$

with $p' < p$. It should be clear what is ment by "principal commutator": it is the commutator that contain A. All the other commutators contract to P and Q which are ordinary pseudodifferential operators. We have again $\sum \text{ord } S'_j + \text{ord } P + \text{ord } Q \leq \sum n_i$. But more can in fact be asserted, we have

$$\sum \text{ord } S'_j + \text{ord } P + \text{ord } Q + (p - p') \leq \sum n_i \,. \tag{8.19}$$

After a moment reflexion the reason for this should be clear. Indeed if we have decreased the length of the "principal commutators", say, be one unit, this is because somewhere in the product we have bracketed two S's, $[S_j, S_k]$. But this bracket makes us gain one unit in the total order (I mean here the order in the sense of pseudodifferential calculus) and so on.

From (8.19) it follows that the inductive step applies. Indeed in the conclusion (8.17) we gain $p - p'$ and lose $-\delta(p - p')$ and since $\delta \leq 1$ the inductive hypothesis gives us, if anything, a stroger estimate. This completes the proof.

If we put together everything that was done in this section we see that we can reduce our criterion at the beginning of Section 0.2 to the Beals criterion [3]. Since A is assumed to be "compactly supported" we can, in fact, use the form of the Beals criterion given in [5], Chapter III.

The assertion (0.4) follows by the same criterion and the estimates (P_k) of Section 7; the proof is therefore, if anything, easier. The assertion (0.5) also follows from the criterion of Section 0.2. To see this let us call $\mathcal{C}^m_\rho$, $(m, \rho \in \mathbb{R})$ the class of operators T as in (0.2) that satisfy the condition (0.2). It is then a formal verification to see that $T_i \in \mathcal{C}^{m_i}_{\rho_i}$, $(i = 1, 2)$ implies that $T_1 T_2 \in \mathcal{C}^{m_1+m_2}_{\min\{\rho_1,\rho_2\}}$. If we combine therefore our result $A^{is} \in \mathcal{C}^0_{1-\delta}$, $(s \in \mathbb{R})$ of Section 6, together with the fact $A^n \in S^{2n}_{1,0} \subset \mathcal{C}^{2n}_1$ $(n = 1, 2 \dots)$, we deduce that $A^\sigma \in \mathcal{C}^0_{1-\delta}$ for $\text{Re } \sigma = 0, 1, 2, \dots$. Complex interpolation gives then that $A^\sigma \in \mathcal{C}^{2\text{Re }\sigma}_{1-\delta}$, $(\text{Re } \sigma \geq 0,\ 0 \leq \delta < 1)$. Our criterion does the rest.

9. An application of a theorem of R. Beals.

We shall place ourselves here in the context of Theorem 5.4 of [3] (*cf.* also [4], [9] for the general setup). We set $P = p^{\omega}(x, D)$ with $p \in S^2_{phg}$ a polyhomogeneous symbol (*cf.* Definition 18.1.5 in Hörmander, vol. III). More general symbols in S^m_{phg}, $m = 2, 3 \ldots$ can also be treated but we shall restrict ourselves to $m = 2$ for simplicity.

Following Beals we must impose on the principal symbol p_2 the same conditions as in [3] (*e.g.* $p_m(x, \xi)$ belongs to the sector $|\mathrm{Arg}\ z| \leq \pi/2 - \varepsilon_0$ or the even less restrictive condition [3]) and also that P is subelliptic with a loss of 1 derivative, *i.e.* that

$$\|u\|_1 \leq C\left(\|Pu\| + \|u\|\right), \quad u \in C_0^{\infty} \tag{9.1}$$

for the usual Sobolev norms $\|\cdot\|_{\alpha}$ and $\|\cdot\| = \|\cdot\|_0$. We shall suppose also that the complex powers P^{σ}, ($\sigma \in \mathbb{C}$) can be defined by say, a ray of minimal growth (*cf.* [9], p. 153). For simplicity we shall in fact assume here that the symbol of P is $p(x, \xi) + \lambda_0$ with some large λ_0 and $p(x, \xi) \geq 0$ and then all the above conditions are verified.

I shall show in this section how the results of R. Beals in [3], [4] and [9] imply our basic estimate (0.1) very easily and in full generality, for $A = P$ as above.

To do this we introduce the (φ, Φ) functions of p. 56 in [3] (with $m = 1$) and consider the corresponding metric

$$g_{x,\xi}(y, \eta) = \frac{|y|^2}{\varphi^2(x, \xi)} + \frac{|\eta|^2}{\Phi^2(x, \xi)} = m\left(|y|^2 + \frac{|\eta|^2}{1 + |\xi|^2}\right) = m g_0(y, \eta)$$

(*cf.* [10], Example 3, p. 378) with $m = \langle \xi \rangle^2 \Phi^{-2}(x, \xi) \geq 1$ and the uncertainty parameter $h = (\varphi\Phi)^{-1} \approx \langle \xi \rangle (|p_2| + \langle \xi \rangle)^{-1} \leq 1$. What counts of course is that the symbol of P lies in the class $S(\Phi^2; g)$ (This is proved in [3] and here I swich freely from Beals to Hörmander's notations).

The additional observation that we need is the fact that $\langle \xi \rangle^m \in O(\Phi, \varphi)$ (with the notations of [3] and [9]) since $R = \Phi/\varphi = \langle \xi \rangle$, in other words $\langle \xi \rangle^m$ is an admissible weight function (in Hörmander's terminology [6], sections 18.4 and 18.5 for the classes $S(m; g)$) for the metric g. This will allow us to exploit the "mixed symbolic" calculus

$$S(m_1; g_1) \times S(m_2; g_2) \quad \longrightarrow \quad S(m_1 m_2; g_1 + g_2)$$

of Theorem [6], 18.5.5 and make a gain on the "order" of the commutators. Le us be more explicit. We shall apply this Theorem 18.5.5 with

$$g_1 = g_0\,, \quad g_2 = g\,, \quad \frac{g_1+g_2}{2} \approx g$$

$m_1 = \langle\xi\rangle^m$, $(m \in \mathbb{R})$ and m_2 any weight function of g. As we just saw both m_1, m_2 are then continuous $\sigma-(g_1+g_2)/2$ temperate weight functions. As for the condition [6], (18.5.13) on g_1, g_2 it is guarantied here by [6], Proposition 18.5.7, which also gives us that

$$H = (h_1 h_2)^{1/2} = (|p_2| + \langle\xi\rangle)^{-1/2}\,.$$

The application of Beals theory ([3], [4], *cf.* Appendix at the end of the paper) gives then that for all $\sigma \in \mathbb{C}$ we have

$$P^\sigma = q^\omega_\sigma(x,D)\,, \quad q_\sigma \in S((|p_2|+\langle\xi\rangle)^{\mathrm{Re}\ \sigma}; g)\,.$$

From this and [6], Theorem 18.5.5 we deduce that

$$[P^\sigma, S] = a^\omega(x,D)\,, \quad a \in S(\langle\xi\rangle^n(|p_2|+\langle\xi\rangle)^{\mathrm{Re}\ \sigma-1/2}; g)$$

for any $S \in OPS^n_{1,0}$. The application of [6], Theorem 18.5.5 can clearly be iterated and we obtain

$$\begin{aligned} &S_0[\dots[P^\sigma, S_1]\dots]S_k]S_{k+1} = b^\omega(x,D)\,,\\ &b \in S(\langle\xi\rangle^{\sum n_j}(|p_2|+\langle\xi\rangle)^{\mathrm{Re}\ \sigma-k/2}; g) \end{aligned} \tag{9.2}$$

for arbitrary pseudodifferentials $S_j \in OPS^{n_j}_{1,0}$, $(0 \le j \le k+1)$.

To obtain our basic estimate (0.1) from (9.2) we must find a way to prove that

$$\|f\| = \|f\|_{L^2} + \|\Lambda^n P^m f\|_{L^2}\,, \quad f \in C^\infty_0 \tag{9.3}$$

is an "admissible norm" (in the sense of [4]) for the space $H(\langle\xi\rangle^n(|p_2|+\langle\xi\rangle)^m; g)$, $(n, m \in \mathbb{R})$, (with Beals notations in [4]). That this is the case for $n = 0$ is proved in Beals [3], [4] and the key to that is Theorem 3.7 of [4] (one easily sees that the same argument gives $n \in \mathbb{R}$, $m = 1, 2, \dots$).

No doubt one can generalise Beals theory to obtain the above more general result for arbitrary $n, m \in \mathbb{R}$. This will not be necessary here however. Indeed from (9.2) and the above results of Beals we certainly have the special case (since then $n = \sum n_i = 0$)

$$\|\mathcal{L}_1\mathcal{L}_2\dots\mathcal{L}_k(P^\sigma)f\|_X \le C\,\|P^{\mathrm{Re}\ \sigma-k/2}f\|_X\,, \tag{9.4}$$

$k = 0, 1, \ldots,\ f \in C_0^\infty$, where I denote by $\mathcal{L}_j$, $(j = 1, \ldots, k)$

$$\mathcal{L}_j(T) = [E, T], \quad \text{or } (c_\lambda - c_\mu)(T)$$

with $E \in S^0_{1,0}$ as usual, and $c_\lambda(T) = \Lambda^{-\lambda} T \Lambda^\lambda$, $X = L^2$ (T indicates an arbitrary operator). At this stage we have to go back to Section 6 of [1] where it was shown that (9.4) implies our estimate (0.1). (This was done in Section 6 of [1] only for $k = 2$ but the proof is clearly general. Observe also that this is essentially the same argument that is used at the end of Section 8 to deal with the general commutator (8.16). The reader should have no difficulty to adapt the argument here). Our proof is complete.

10. The generalisation of the geometric theorem.

This section relies very heavily on the methods, ideas and notations of sections 7,8 and 9 of [1]. What I shall do is to use the results of the previous section to give the generalisation of the main geometric Theorem of [1], Section 0 as was promised in Section 0.1.

Let $M = a^\omega(x, D) + \lambda_0$ with $0 \leq a(x, \xi) \in S^2_{phg}$ and large $\lambda_0 > 0$ satisfying the conditions of Section 9, and let $L = \sum X_j^* X_j$ a subelliptic Hörmander operator or more generally an operator of the form $L = \sum X_j^* X_j + \Lambda^\alpha$, $(0 \leq \alpha \leq 2)$ where $\sum X_j^* X_j$ is again assumed to be subelliptic. These operators were denoted by $\tilde{L} = \sum Y_j^* Y_j$ in [1], Section 9, and the Y's that we shall consider below are the Y's defined there. We shall further assume that the two operators L, M satisfy the subellipticity estimates of [1], Section 7,

$$(10.1)\quad \|f\|_{1-\delta} \leq C \,\|(I + L)^{1/2} f\|_X\,, \quad \|f\|_{1-d} \leq C \,\|(I + M)^{1/2} f\|_X\,.$$

Our conditions on M imply that $d \leq 1/2$. We shall also assume that $d + \delta \leq 1$ and we shall extend our Proposition in Section 7 of [1] in the present setting. More specifically we shall prove [1], equality (7.2)

$$\|(I + L)^{j/2} e^{-tM} (I + L)^{-j/2}\|_{\alpha \to \beta} = O\left(t^{(\alpha-\beta)/2(1-d)}\right), \quad \beta \geq \alpha\,,$$

for the above operators L and M. The norms $\|\cdot\|_\alpha$, $\|\cdot\|_{\alpha\to\beta}$ refer throughout to the classical Sobolev norms H_α. The proof of [1], (7.2), that I shall give below is very close in spirit to the proof given in [1],

Section 7. Indeed it is in some sense *dual* to the proof there. Once the [1], (7.2) has been generalised for our present operators we can obtain the generalisation of the geometric theorem that was announced in Section 0.1 exactly as in [1].

Before we start the proof of the estimate, we shall need to note an easy *algebraic* identity

$$\begin{aligned}&[m, y_1 y_2 \dots y_k]\\ &= \sum_{\sigma,j} p_{\sigma,j}[m, y_{\sigma(1)}, \dots, y_{\sigma(j)}] y_{\sigma(j+1)} \cdots y_{\sigma(k)}\,, \quad k \geq 1\end{aligned} \tag{10.2}$$

for arbitrary indeterminates m; $y_1, \dots, y_k$ and $p_{\sigma,j} \in \mathbb{Z}$, where σ runs through the permutations of $1, 2, \dots, k$. This is easily proved by induction on k.

We shall also need to introduce the following notation: for $Y_1, Y_2, \dots \in S^1_{1,0}$ determined by the operator L (or rather $\tilde{L}$ as in Section 9 of [1]) and $k = 0, 1, \dots$ I shall denote by

$$R_k(t) = Y_{i_1} Y_{i_2} \dots Y_{i_k} e^{-tM} (I + L)^{-k/2}\,.$$

There are of course several R_k's for a fixed k and they depend on the choice of $i_1, \dots, i_k$.

Our first step is to prove by induction on k that for all $\beta \geq \alpha$ and $k = 0, 1, \dots$, we have

$$\|R_k(t)\|_{\alpha \to \beta} = O\left(t^{(\alpha-\beta)/2(1-d)}\right)\,. \tag{10.3}$$

This statement for $k = 0$ is contained in [8] (*cf.* also Section 3 of [1]).

Our aim is therefore to assume that (10.3) holds for $0, 1, \dots, k$ and prove it for $k + 1$. Towards that we fix $(i_1, \dots, i_{k+1})$ which to simplify notations we shall *rename* $1, 2, \dots, k+1$. We then develop in our usual way

$$\begin{aligned}[e^{-tM}, Y_1 \dots Y_{k+1}] &= I_1 + I_2\\ &= \int_0^t e^{-(2t-s)M} [M, Y_1 \dots Y_{k+1}] e^{-sM} ds\\ &\quad + \int_0^t e^{-(t-s)M} [M, Y_1 \dots Y_{k+1}] e^{-(t+s)M} ds\,.\end{aligned}$$

This together with our identity (10.2) gives us a decomposition

$$[e^{-tM}, Y_1 \dots Y_{k+1}](I+L)^{-(k+1)/2} = \sum_j p_j (I_j^{(1)} + I_j^{(2)})$$

$$I_j^{(1)} = \int_0^t e^{-(2t-s)M} [M, Y_1 \dots Y_j] R_{k+1-j}(s)\, ds\, (I+L)^{-j/2}$$

and the analogous expression with the usual switch $(2t-s) \to t-s$, $s \to t+s$ for $I_j^{(2)}$. The Y_i's in the above formula have, of course, undergone one more renaming (they really are $Y_{\sigma(i)}$'s for an appropriate permutation σ).

We shall factor $\|I_j^{(1)}\|_{\alpha \to \beta}$, $(\beta \geq \alpha)$ and estimate it by

$$\begin{aligned} &\int_0^t \|e^{-(2t-s)M} M^{1-j/2}\|\, \|M^{-1+j/2}[M, Y_1, \dots, Y_j]\| \\ &\qquad \cdot \|R_{k+1-j}(s)\|\, ds\, \|(I+L)^{-j/2}\| \end{aligned} \tag{10.4}$$

which is an in Section 7 of [1] an appropriate cascade of $\|\cdot\|_{r \to s}$ norms, that unfolds as follows

$$\begin{aligned} \|(I+L)^{-j/2}\|_{\alpha \to \alpha + j(1-\delta)} &\leq C\,, \\ \|R_{k+1-j}(s)\|_{\alpha + j(1-\delta) \to \alpha + j(1-\delta)} &= O(1)\,, \\ \|M^{-1+j/2}[M, Y_1, \dots, Y_j]\|_{\alpha + j(1-\delta) \to \alpha - j\delta} &\leq C \end{aligned}$$

for the first estimate (*cf.* [1], [8]). The second follows from our inductive hypothesis and to see the third we use the result of Section 9 together with the fact that each $Y_j \in OPS^1_{1,0}$. To estimate the first term in the integral (10.4) we recall that for $\lambda \geq 0$ we have for $\beta \geq \alpha$

$$\begin{aligned} \|M^\lambda e^{-tM}\|_{\alpha \to \beta} &\leq \|M^\lambda e^{-t/2M}\|_{\beta \to \beta} \|e^{-t/2M}\|_{\alpha \to \beta} \\ &= O\left(t^{-\lambda + (\alpha-\beta)/2(1-d)}\right) \end{aligned} \tag{10.5}$$

and

$$\begin{aligned} \|e^{-tM} M^{-\lambda}\|_{\alpha \to \beta} &\leq \|e^{-tM}\|_{\alpha + 2\lambda(1-d) \to \beta} \|M^{-\lambda}\|_{\alpha \to \alpha + 2\lambda(1-d)} \\ &= O\left(t^{\lambda + (\alpha-\beta)/2(1-d)}\right) \end{aligned} \tag{10.6}$$

provided that $\beta \geq \alpha + 2\lambda(1-d)$ (since the factor $\|M^{-\lambda}\|$ is bounded (*cf.* [1], [8])). We apply this to the first factor inside the integral of

(10.4) and distinguish two cases $j = 1, 2$ and $j > 2$. In the first case (10.5) gives us

$$\begin{aligned} &\|e^{-(2t-s)M} M^{1-j/2}\|_{\alpha-j\delta\to\beta} \\ &\qquad = O\left((2t-s)^{-1+j/2+(\alpha-j\delta-\beta)/2(1-d)}\right). \end{aligned} \tag{10.7}$$

We multiply out and integrate and obtain

$$\|L_j^{(1)}\|_{\alpha\to\beta} = O\left(t^{(j(1-d-\delta))/2(1-d)+(\alpha-\beta)/2(1-d)}\right). \tag{10.8}$$

If $j > 2$ we use the estimate (10.6) to obtain (10.7) again and we obtain also exactly the same estimate (10.8) for $I_j^{(1)}$ as long as the *exponent* of t in $O(t^{\text{exponent}})$ of (10.7) is ≤ 0. This however is always te case since $\beta \geq \alpha$ and $0 \leq d < 1$.

The integrals $I_j^{(2)}$ are estimated by the analogue of the integral (10.4) where we replace $(2t - s)$ by $(t - s)$ and s by $(t + s)$ on the exponentials and $R_{k+1-j}(\cdot)$. The cascade of $\|\cdot\|_{r\to s}$ norms runs now as follows

$$\begin{aligned} \|(I+L)^{-j/2}\|_{\alpha\to\alpha+j(1-\delta)} &\leq C\,, \\ \|R_{k+1-j}(t+s)\|_{\alpha+j(1-\delta)\to\gamma+j(1-\delta)} &= O(1)\,, \end{aligned}$$

by the induction hypothesis provided that $\gamma = \alpha + \varepsilon \geq \alpha$. We also have, just as before,

$$\|M^{-1+j/2}[M, Y_1, \ldots, Y_j]\|_{\gamma+j(1-\delta)\to\gamma-j\delta} \leq C\,.$$

To estimate the first term we have to distinguish again the two cases $j = 1, 2$ and $j > 2$. In the first case we have

$$\begin{aligned} &\|e^{-(t-s)M} M^{1-j/2}\|_{\gamma-j\delta\to\beta} \\ &\qquad = O\left((t-s)^{-1+j/2+(\gamma-\beta-j\delta)/2(1-d)}\right) \end{aligned} \tag{10.9}$$

as long as $\beta \geq \gamma - j\delta$. Then since $\beta \geq \alpha$ we can choose $\gamma = \beta$ and after integration we obtain

$$\|I_j^{(2)}\|_{\alpha\to\beta} = O\left(t^{(j(1-d-\delta))/2(1-d)+(\alpha-\beta)/2(1-d)}\right). \tag{10.10}$$

In the second case $j > 2$ we obtain the same estimates (10.9) and (10.10) provided that (the left inequality below is to make the integral converge)

$$(10.11)\qquad \begin{aligned} -1 &< \frac{j}{2} - 1 + \frac{\gamma - \beta - j\delta}{2(1-d)} \\ &= -1 + \frac{j(1-d-\delta)}{2(1-d)} + \frac{\alpha - \beta}{2(1-d)} + \frac{\varepsilon}{2(1-d)} \leq 0 . \end{aligned}$$

At first sight it looks as if here we are in trouble. Indeed for $1-d-\delta > 0$ and $j \gg 0$, (10.11) is incompatible for $\varepsilon \geq 0$. But of course we can get round that difficulty simply by *assuming* that $1 - d - \delta = 0$. This is no loss of generality since we can allways increase the d and δ in the definition of subellipticity of L and M without altering the validity of the conditions (10.1). In that case $d + \delta = 1$ the inequalities are then always compatible for some $\varepsilon \geq 0$ since by our hypothesis $\beta \geq \alpha$.

All in all we have therefore established that, under the inductive hypothesis, we have

$$(10.12)\qquad \|[e^{-tM}, Y_1 \dots Y_{k+1}](I+L)^{-(k+1)/2}\|_{\alpha\to\beta} = O\left(t^{(\alpha-\beta)/2(1-d)}\right)$$

(provided that $d + \delta \leq 1$).

At this stage we shall invoque the estimate (9.1) of [1], (where the subellipticity of $\sum X_j^* X_j$ is apparently needed). This together with (10.12) and the (standard by now, I hope) fact that

$$\|e^{-tM}\|_{\alpha\to\beta} = O\left(t^{(\alpha-\beta)/2(1-d)}\right)$$

establishes the inductive step and complete the proof of (10.3) in all generality.

We shall now finish the proof of [1], (7.2). Assume that $j = 2k$ is an even integer, then

$$(I+L)^{j/2} = \sum_{p\leq k} \lambda_i Y_{i_1} Y_{i_2} \dots Y_{i_{2p}}, \quad \lambda_j \in \mathbb{Z}$$

and our estimate [1], (7.2) for $\alpha = \beta$ follows from (10.3). Equivalently what we have proved is

$$e^{-tA} : X_{2j} \longrightarrow X_{2j}, \quad j = 1, 2, \dots$$

with our old notation $X_\alpha = \{f : (1+L)^{\alpha/2} f \in L^2\}$. Duality and interpolation completes the proof of [1], (7.2) for $\alpha = \beta$.

This is good enough for our purposes and proves the analogue of the proposition in Section 7 of [1]. We can however also prove [1], (7.2) in full generality $\beta \geq \alpha$ by a slightly more sophisticated variant of complex interpolation. This was explained in Section 7 of [1].

REMARK. One of the facts that was used in [1], Section 8, is that $e^{-t\widetilde{L}}$ acts on the spaces X_α ($\alpha \in \mathbb{R}$). This fact when $\alpha = 2n$ ($n = 1, 2, \dots$) is a consequence of the semiboundedness of Δ on X_{2n}, and this was proved in [1], Section 10. The general fact follows then by duality and interpolation.

Contrary to what was asserted in [1], Section 10, on the other hand, this actual semiboundedness of Δ on each X_α (for some appropriate scalar product) does *not* seem to follow by interpolation. This semiboundedness is however never used anywhere else so we do not need to prove it.

Appendix on the Beals theory.

In this appendix, using the Beals theory [3], [4] and [9], I shall outline a proof of the fact that the norm $\|f\|$ in (9.3) with $n = 0$ and $m = N/2$, $N = 1, 2, \dots$) a half integer is an admissible norm for the space $H((|p_2| + \langle\xi\rangle)^m; g)$. This fact is explicitely proved in the papers of Beals. The point is however, that the direct proof that I give here, only uses the basic definitions of the Beals theory and none of the more sophisticated machinery developed by Beals. On the other hand this special case ($n = 0$, $m = N/2$) is all that is needed for the proof of our basic estimate (0.1). In other words we only need (9.4) for $\sigma = N/2$ (a half integer) and then if we inject that information in Section 6 of [1] we can make everything work.

The first thing to observe towards that goal is that our basic hypothesis (9.1) implies that

$$C\,\|(P+I)^\alpha f\| \geq \|(P+a\Lambda)^\alpha f\|\,, \quad f \in C_0^\infty \tag{A.1}$$

for all a, $\alpha \geq 0$. Indeed it suffices to prove (A.1) for $\alpha = 1, 2, \dots$ we can develop then $(P+a\Lambda)^\alpha$ and we reduce the problem to proving that

$$\|L_1 L_2 \dots L_k f\| \leq C\,\|(P+I)^\alpha f\| \tag{A.2}$$

where L_j is either $A = P+I$ or $L_j \in S^{n_j}_{1,0}$ and where if s is a the number of A's then $\sum n_j + s \leq \alpha$. For $s = 0$ (A.2) is a consequence of (9.1) (*cf.* [8]). We can thus use induction on s. The inductive hypothesis and the fact that $[A, S^n_{1,0}] \subset S^{n+1}_{1,0}$ allow us then to commute and bring all the A's at the beginning of the product. (A.2) is thus reduced to

$$\|TA^s f\| \leq C \|A^\alpha f\|, \quad f \in C_0^\infty$$

with $T \in S^{\alpha-s}_{1,0}$, $0 \leq s \leq \alpha$. This is clearly a consequence of [8] (set $\varphi = A^s f$).

Having proved (A.1) let us denote $P_N = (P + a\Lambda)^N$ (for some large $a \geq 0$, $N = 1, 2, \dots$). Our problem is to show that $\|f\|_N = (P_N f, f)^{1/2}$ is a norm for the space $H(m^{N/2}; g)$ where we denote by

$$m = p + C\langle\xi\rangle \approx p_2 + C\langle\xi\rangle.$$

For simplicity we shall suppose here that the symbol of P is nonnegative, $p(x, \xi) \geq 0$.

The proof of this fact is an easy consequence of the existence of the following two "parametrices"

$$\text{(A.3)} \qquad \begin{cases} Q_\pm = q^\omega_\pm(x, D), & q_\pm \in S(m^{\pm N/2}; g) \\ P_N \equiv Q^*_+ Q_+ \ \text{mod-}OPS(m^N h^s; g), \\ Q_- Q_+ \equiv I \ \text{mod-}OPS(h^s; g) \end{cases}$$

where $s \geq 1$ can be chosen in advance and arbitrarily large. To construct these parametrices let us denote by $R_N = (m^N)^\omega(x, D)$, $(N \in \mathbb{R})$ and let us observe that by standard symbolic calculus we have

$$R_{-N/2} P_N R_{-N/2} \equiv 1 \quad \text{mod-}OPS(h; g).$$

This allows us to use the binomial $(1 + z)^{1/2} = 1 + z/2 + \dots$ and write

$$R_{-N/2} P_N R_{-N/2} \equiv Y^2 \quad \text{mod-}OPS(h^s; g), \ Y = Y^* \in OPS(1; g)$$

with arbitrary high $s \geq 1$. Similarly we have

$$R_{-N/2} R_{N/2} \equiv 1 \quad \text{mod-}OPS(h; g)$$

and the Neumann series $1 - z + z^2 - \dots$ allows us to construct a parametrix $\tilde{R}_{N/2} \in OPS(m^{N/2}; g)$

$$R_{-N/2} \tilde{R}_{N/2} \equiv 1 \quad \text{mod-}OPS(h^s; g)$$

with arbitrarily high $s \geq 1$. Combining these two facts we obtain

$$P_N \equiv \tilde{R}_{N/2}^* Y^2 \tilde{R}_{N/2} \quad \text{mod-}OPS(m^N h^s; g).$$

It follows thus that we can set $Q_+ = Y\tilde{R}_{N/2} = q_+^\omega(x, D)$ in (A.3).

Observe now that $Y \equiv 1$ mod-$OPS(h; g)$ and so $R_{-N/2}Q_+ = R_{-N/2}Y\tilde{R}_{N/2} \equiv 1$ mod-$OPS(h; g)$. The same Neumann series $1 - z + z^2 - \ldots$ allows us therefore to construct in (A.3) the required parametrix Q_- of Q_+.

Once we have (A.3) we can write (with $s \geq N/2$)

$$(P_N f, f) - \|Q_+ f\|^2 = (Tf, f),$$
$$T \in OPS(m^{N/2} h^s; g) \subseteq OPS(\langle \xi \rangle^{N/2}; g).$$

It follows that

$$|(Tf, f)| \leq \|R\Lambda^{N/4} f\| \, \|\Lambda^{N/4} f\|, \quad R = \Lambda^{-N/4} T \Lambda^{-N/4} \in OPS(1; g).$$

On the other hand (with obvious notations !) we have

$$\|f\|_{\tilde{m}} \leq C \, \|Q_- Q_+ f\|_{\tilde{m}} + C \, \|f\|_{\tilde{m}h^s}, \quad f \in C_0^\infty$$

(for any arbitrary weight function $\tilde{m}$). If we set $\tilde{m} = m^{N/2}$ and $s \geq 1$ large enough we obtain

$$\|f\|_{m^{N/2}} \leq C \left(\|Q_+ f\| + \|f\|_{\langle \xi \rangle^{N/2}} \right). \tag{A.5}$$

But clearly also

$$\|f\|_{\langle \xi \rangle^{N/2}} \leq \|\Lambda^{-N/2} \Lambda^{N/2} f\|_{\langle \xi \rangle^{N/2}} \leq C \, \|\Lambda^{N/2} f\| \tag{A.6}$$

since $\Lambda^{-N/2} \in OPS(\langle \xi \rangle^{N/2}; g)$. Putting together (A.4), (A.5) and (A.6) we deduce that

$$\|f\|_{m^{N/2}} \leq C \, \|f\|_N, \quad f \in C_0^\infty,$$

which is the desired estimate.

References.

[1] Varopoulos, N. Th. Semigroup commutators under differences. *Revista Mat. Iberoamericana*, **8** (1992), 1-43.

[2] Hörmander, L., Pseudo-differential operators and hypoellitptic equations. *Proc. Sump. Pure Math.* **X** (1967), 138-183.

[3] Beals, R., Characterization of pseudodifferential operators and applications. *Duke J. Math.* **44** (1977), 45-57.

[4] Beals, R., Weighted distributions spaces and pseudodifferential operators. *Journal d'Analyse Math.* **39** (1981), 131-187.

[5] Coiffman, R.R., Meyer, Y., Au delà des operateurs pseudo-différentiels. *Astérisque* **57** (1978), 1-185.

[6] Hörmander, L. *The analysis of linear partial differential operators.* Springer-Verlag, vol. I, 1983; vol. III, 1985.

[7] Davies, E. B., *One parameter semigroups.* Academic Press, 1980.

[8] Varopoulos, N.Th., Puissances des opérateurs pseudo-différentiels. *C.R. Acad. Sci. Paris* **310** (1990), 769-774.

[9] Beals, R., A general calculus of pseudodifferential operators. *Duke Math. J.* **42** (1975), 1-42.

[10] Hörmander, L. The Weyl calculus of pseudodifferential operators. *Comm. Pure Appl. Math.* **32** (1979), 359-443.

Recibido: 5 de diciembre de 1.991

N. Th. Varopoulos
Departement de Mathématiques
Université de Paris VI
75005 Paris, FRANCE

REVISTA MATEMÁTICA IBEROAMERICANA
VOL. 9, N.° 1, 1993

Non-separable bidimensional wavelet bases

Albert Cohen and Ingrid Daubechies

Abstract. We build orthonormal and biorthogonal wavelet bases of $L^2(\mathbb{R}^2)$ with dilation matrices of determinant 2. As for the one dimensional case, our construction uses a scaling function which solves a two-scale difference equation associated to a FIR filter. Our wavelets are generated from a single compactly supported mother function. However, the regularity of these functions cannot be derived by the same approach as in the one dimensional case. We review existing techniques to evaluate the regularity of wavelets, and we introduce new methods which allow to estimate the smoothness of non-separable wavelets and scaling functions in the most general situations. We illustrate these with several examples.

I. Introduction.

In the most general sense, wavelet bases are discrete families of functions obtained by dilations and translations of a finite number of well chosen mother functions. The most well known are certainly dyadic orthonormal bases of $L^2(\mathbb{R})$, of the type

$$\psi_k^j(x) = 2^{-j/2}\psi(2^{-j}x - k), \quad j,k \in \mathbb{Z}\,. \tag{1.1}$$

These constructions have found many interesting applications, both in mathematics because they form Riesz bases for many functional spaces

and in signal processing because wavelet expansions are more appropriate than Fourier series to represent the abrupt changes in non-stationary signals.

Several examples have been given by Meyer [Me1], Lemarié [Le] and Daubechies [Dau1], generalizing the classic Haar basis in which the mother wavelet $\psi = \chi_{[0,1/2]} - \chi_{[1/2,1]}$ suffers from a lack of regularity since it is not even continuous. All are based on the concept of multiscale analysis, *i.e.* a ladder of closed subspaces $\{V_j\}_{j\in\mathbb{Z}}$ which approximates $L^2(\mathbb{R})$,

$$\{0\} \to \dots V_1 \subset V_0 \subset V_{-1} \dots \to L^2(\mathbb{R}) \,, \tag{1.2}$$

(note that in some papers and in Meyer's book, the converse convention is used, *i.e.* $V_j \subset V_{j+1}$) and satisfies the following properties,

$$f(x) \in V_j \Longleftrightarrow f(2x) \in V_{j-1} \Longleftrightarrow f(2^j x) \in V_0 \,, \tag{1.3}$$

(1.4) there exists a function $\varphi(x)$ in V_0 such that the set $\{\varphi(x-k)\}_{k\in\mathbb{Z}}$ is an orthonormal basis for V_0 .

Since $V_0 \subset V_{-1}$, the scaling function $\varphi(x)$ has to be the solution of a two scale difference equation,

$$\varphi(x) = 2\sum_{n\in\mathbb{Z}} c_n \, \varphi(2x-n) \,. \tag{1.5}$$

The associated wavelet is then derived from the scaling function by the formula

$$\psi(x) = 2\sum_{n\in\mathbb{Z}} (-1)^n \, \overline{c}_{1-n} \, \varphi(2x-n) \,. \tag{1.6}$$

In the standard interpretation of a multiresolution analysis, the projections of a function f on the spaces V_j are viewed as successive approximations to f, with finer and finer resolution as j decreases. The wavelets can then be used to express the additional details needed to go from one resolution to the next finer level, since the $\{\psi(x-k)\}_{k\in\mathbb{Z}}$ constitute an orthonormal basis for W_0, the orthogonal complement of V_0 in V_{-1}. The whole set $\{\psi_k^j(x)\}_{j,k\in\mathbb{Z}}$ forms then an orthonormal basis of $L^2(\mathbb{R})$.

We are here interested in similar constructions adapted to functions or signals of more than one variable.

The most commonly used method to build a multiresolution analysis and wavelet bases in $L^2(\mathbb{R}^n)$ is the tensor product of a multiresolution analyses of $L^2(\mathbb{R})$. In $L^2(\mathbb{R}^2)$ it leads to a ladder of spaces $\mathcal{V}_j = V_j \otimes V_j \subset \mathcal{V}_{j-1}$ generated by the families,

$$\Phi^j_{k\ell}(x,y) = 2^{-j}\varphi(2^{-j}x-k)\,\varphi(2^{-j}y-\ell), \quad k,\ell \in \mathbb{Z}\,. \tag{1.7}$$

Three wavelets are then necessary to construct the orthogonal complement of $\mathcal{V}_0$ in $\mathcal{V}_{-1}$, namely,

$$\Psi_a(x,y) = \varphi(x)\psi(y)\,, \tag{1.8}$$

$$\Psi_b(x,y) = \psi(x)\varphi(y)\,, \tag{1.9}$$

$$\Psi_c(x,y) = \psi(x)\psi(y)\,. \tag{1.10}$$

Actually, the theory of multiresolution analysis, as it was introduced by S. Mallat and Y. Meyer (see [Ma1] and [Me1]) was first motivated by the possibility of building these separable wavelets for the analysis of digital picture.

It is clear, however, that this choice is restrictive and that it gives a particular importance to the x and y directions, since Ψ_a and Ψ_b match respectively the horizontal and vertical details.

A more general way of extending multiresolution analysis to n dimensions consists in replacing the axioma (1.3) and (1.4) by

$$f(x) \in V_j \iff f(Dx) \in V_{j-1} \tag{1.11}$$

(1.12) There exists a function ϕ in V_0 such that the set $\{\phi(x-k)\}_{k\in\mathbb{Z}^n}$ is an orthonormal basis for V_0 ,

where D is a $n \times n$ dilation matrix.

All the singular values $\lambda_1, \ldots, \lambda_n$ of D must satisfy

$$|\lambda_m| > 1\,, \tag{1.13}$$

to ensure that the approximation gets finer in every direction as j goes to $-\infty$. Furthermore, we require D to have integer entries. This condition means that the action of D on the translation grid $\mathbb{Z}^n$ leads to a sublattice $\Gamma \subset \mathbb{Z}^n$.

The number of basic wavelets required to characterize the orthogonal complement of V_0 in V_{-1} is in that case trivially given by the following heuristic argument. This complement should be generated

by the action of $\mathbb{Z}^n$ on the basic wavelets, in the same way that V_0 is generated by the action of $\mathbb{Z}^n$ on ϕ, whereas V_{-1} is generated by the action of $D^{-1}\mathbb{Z}^n$. Consequently, each of the generating functions can be associated with an elementary coset of $D^{-1}\mathbb{Z}^n/\mathbb{Z}^n \sim \mathbb{Z}^n/D\mathbb{Z}^n$ except one which corresponds to the scaling function (see figure 1). Therefore, $d = |\det D| - 1$ different wavelets are needed. Note that it is not strictly necessary that the entries of D be integer to build wavelet bases using D as the elementary dilation.

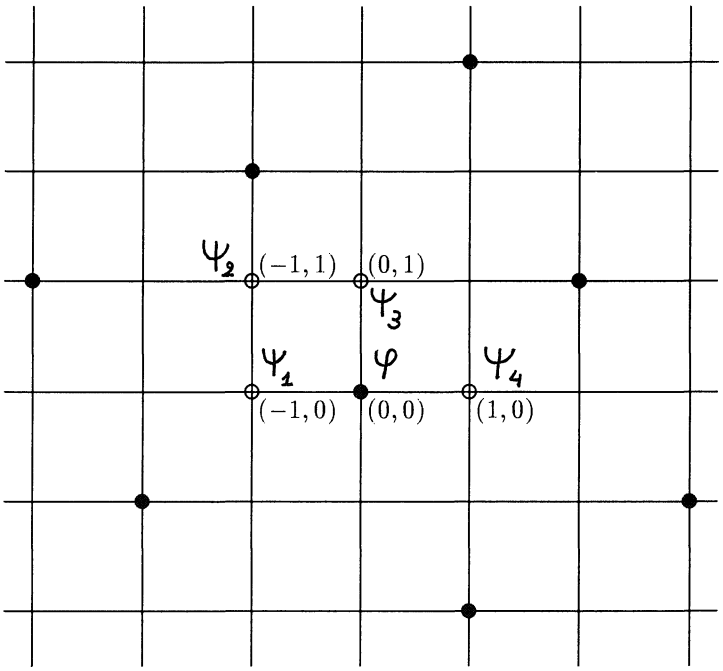


Figure 1

$\mathbb{Z}^2$ and $D\mathbb{Z}^2$ in the case where $D = \begin{pmatrix} 2 & -1 \\ 1 & 2 \end{pmatrix}$.
The scaling function and the four basic wavelets
are indexed by an element of $\mathbb{Z}^2/D\mathbb{Z}^2$.

However, the condition seems to be necessary for the existence of a multiresolution analysis based on a single, real valued, compactly supported scaling function.

In this work we shall indeed focus on real valued, compactly supported scaling functions and wavelets. They have the advantage that the sequence $\{c_n\}_{n\in\mathbb{Z}}$ introduced in the two scale difference equation (1.5) is real and finite. These coefficients play an important part in the numerical applications because they are used directly in the Fast Wavelet Transform algorithm as decomposition and reconstruction filters. They constitute in that case an FIR (finite impulse response) filter which can be implemented very easily. Furthermore, this finite set of coefficients contains all the information about the multiresolution analysis since the functions φ and ψ can be constructed as solutions of (1.5) and (1.6). Our starting point to build wavelet bases will thus be a finite set of coefficients and the associate two-scale difference equation, rather than the approximation spaces V_j themselves.

The main difficulty in this approach is the design of the FIR filter $\{c_n\}_{n=0,\dots,N}$ in such a way that φ and ψ are smooth and have orthonormal translates.

In the one dimensional case, it is shown in [Dau1] that orthonormal wavelets can be constructed by choosing a filter which corresponds to a particular case of exact reconstruction subband coding schemes, and which can be made arbitrarily regular by increasing the number of taps in a proper way. Several contributions have followed, giving supplementary information on the type of filter which has to be used (see [Me2], [DL], [Co1], [Dau2], [Co2], [Dau3]).

In the present bidimensional case, the design of filters associated to "nice wavelet bases" turns out to be more difficult because some of the one-dimensional techniques do not generalize trivially (or do not generalize at all!) to higher dimensions and new methods have to be introduced. This article concentrates on the situation where D is a 2×2 matrix with $|\det D| = 2$.

We deliberately restrict ourselves to this set of matrices for two reasons:

- These dilations have already been considered by electrical engineers and seem to have interesting applications in signal analysis and image processing. For example, since only one basic wavelet is required, one may hope for a more isotropic analysis than with the separable construction. Subband coding schemes with decimation on the quincunx sublattice have been studied in the works

of J. C. Feauveau [Fea] and M. Vetterli and J. Kovacevic [KV]. Our work is complementary to their signal processing approach since we investigate here the mathematical properties, such as the Hölder regularity of the wavelet bases associated to these schemes. This regularity is important when one asks that the reconstruction of the signal from the coarse scales has a smooth aspect (see Section II.2).

- These dilations are simple and our study will be reduced to the case of two basic matrices. However, the difficulties which appear in the evaluation of the regularity of the corresponding wavelets are common to all the non-separable constructions, and the techniques that we develop to solve this problem can be used for other types of dilations. We believe that the set of integer matrices with $|\det D| = 2$ constitutes an interesting "laboratory case" in the general framework of multidimensional wavelets.

In the next section of this paper, we shall give an overview of different techniques which can be used in the construction of one dimensional compactly supported wavelets. Some new tools will be introduced specifically to be generalized and used in the multidimensional situation.

The third section examines the possible subband coding schemes with decimation on the quincunx sublattice and their general relations with non-separable wavelet bases.

In the fourth section, orthonormal bases of wavelets are constructed from such coding schemes. We show that for the same filters, different bases with widely differing regularity can be obtained, depending on the choice of the dilation matrix. Finally, we use a biorthogonal approach, in Section V, to construct more symmetrical wavelet bases corresponding to linear phase filters and allowing a more isotropic analysis. We show that arbitrarily high regularity can be attained and we give some asymptotical results.

II. The construction of compactly supported wavelets in one dimension: A complete toolbox.

The purpose of this section is to review, in the one dimensional case, many different techniques that can be used to build regular wavelets from subband coding schemes, theoretically and numerically. Some of these techniques, like the Littlewood-Paley estimation of smoothness,

are not frequently used in the one dimensional case, but they turn out to be very useful for the non-separable bidimensional wavelets. For more details, the reader can also consult [Dau1], [Me1], [Ma1], [Ve1], [Dau2], [Me2], [Co2].

Wavelet bases and subband coding schemes.

II.1.a. The orthonormal case.

Let $\{V_j\}_{j\in\mathbb{Z}}$ be a multiresolution analysis of $L^2(\mathbb{R})$. We can use the discrete Fourier transform of the finite sequence $\{c_n\}_{n=N_1}^{N_2}$, *i.e.* the transfer function

$$m_0(\omega) = \sum_{n\in\mathbb{Z}} c_n e^{-in\omega} = \sum_{n=N_1}^{N_2} c_n e^{-in\omega} , \tag{2.1}$$

to rewrite the two scale difference equation (1.5) that characterizes $\varphi(x)$. We suppose that the c_n are real. Taking the Fourier transform of (1.5) and (1.6) we obtain

$$\hat{\varphi}(2\omega) = m_0(\omega)\hat{\varphi}(\omega) \tag{2.2}$$

$$\hat{\psi}(2\omega) = e^{-i\omega}\, \overline{m_0(\omega+\pi)}\, \hat{\varphi}(\omega) = m_1(\omega)\hat{\varphi}(\omega) . \tag{2.3}$$

Two fundamental properties of $m_0(\omega)$ can be derived from the multiresolution analysis properties

- Since $\{\varphi(x-k)\}_{k\in\mathbb{Z}}$ is an orthonormal basis of V_0, the Fourier transform $\hat{\varphi}(\omega)$ satisfies a Poisson identity

$$\sum_{n\in\mathbb{Z}} |\hat{\varphi}(\omega + 2n\pi)|^2 = 1 . \tag{2.4}$$

Combined with (2.2) this leads to

$$|m_0(\omega)|^2 + |m_0(\omega+\pi)|^2 = 1 \tag{2.5}$$

which may also be written as

$$2\sum_{n\in\mathbb{Z}} c_n c_{n+2k} = \delta_{k,0} \ (= 1 \text{ if } k = 0,\ 0 \text{ otherwise}) . \tag{2.6}$$

- The denseness of $\{V_j\}_{j\in\mathbb{Z}}$ in $L^2(\mathbb{R})$ is equivalent to

$$\hat{\varphi}(0) \;=\; \int \varphi(x)dx \;=\; 1\,,$$

(see [Me1], [Ma1] or [Co1]).

Consequently, we have

$$m_0(0) \;=\; 1 \quad \text{and} \quad m_0(\pi) \;=\; 0\,, \tag{2.7}$$

which may also be written as

$$\sum_{n=N_1}^{N_2} c_n \;=\; 1 \quad \text{and} \quad \sum_{n=N_1}^{N_2} (-1)^n c_n \;=\; 0\,. \tag{2.8}$$

The subband coding scheme associated to our multiresolution analysis appears clearly in the Fast Wavelet Transform Algorithm of S. Mallat [Ma2]. Let us recall how it works. The initial data are considered as the approximation of a continuous function at the scale $j = 0$,

$$S_k^0 \;=\; \langle f, \varphi(x-k)\rangle, \quad k\in\mathbb{Z}\,. \tag{2.9}$$

This allows the computation of the approximations and the details at coarser scales, *i.e.*

$$S_k^j \;=\; 2^{-j/2}\,\langle f, \varphi_k^j\rangle \quad \text{and} \quad D_k^j \;=\; 2^{-j/2}\,\langle f, \psi_k^j\rangle, \quad j>0\,. \tag{2.10}$$

(The coefficients are normalized in such way that if $f\equiv 1$ locally, then $S_k^j = 1$ in that area). The sequence $\{S_k^j\}_{k\in\mathbb{Z}}$ (respectively $\{D_k^j\}_{k\in\mathbb{Z}}$) is then derived from $\{S_k^{j-1}\}_{k\in\mathbb{Z}}$ by a convolution with the filter $\overline{m_0(\omega)}$ (respectively, $\overline{m_1(\omega)}$) followed by a decimation of one sample out of two to keep the same total amount of information, *i.e.*

$$S_k^j \;=\; \sum_n c_{n-2k}\, S_n^{j-1}\,, \quad D_k^j \;=\; \sum_n (-1)^{n-1}\, c_{2k+1-n}\, S_n^{j-1}\,.$$

The algorithm then iterates on $\{S_k^j\}_{k\in\mathbb{Z}}$. Conversely, the sequence $\{S_k^{j-1}\}_{k\in\mathbb{Z}}$ can be recovered by applying the same filters $m_0(\omega)$ and $m_1(\omega)$ on $\{S_k^j\}_{k\in\mathbb{Z}}$ and $\{D_k^j\}_{k\in\mathbb{Z}}$ after inserting a zero between every

pair of consecutive samples, and summing the two components (multiplied by two for normalization purposes), *i.e.*

$$S_n^{j-1} = 2\sum_k c_{n-2k}\, S_k^j + (-1)^{n-1}\, c_{2k+1-n}\, D_k^j .$$

All these operations, decomposition - decimation - interpolation - reconstruction, constitute a complete subband coding scheme as shown on figure 2. The property of exact reconstruction can now be derived in two ways. It is a natural consequence of the multiresolution approach, since $V_j = V_{j+1} \overset{\perp}{\oplus} W_{j+1}$ but it can also be viewed as a consequence of formula (2.5) for the filter m_0. This type of filter pair (m_0, m_1) is known as a pair of "conjugate quadrature filters" (CQF); they were first discovered by Smith and Barnwell in 1983, *cf.* [SB1]. The design of FIR pairs, with real coefficients and perfect reconstruction, has been generalized in [Dau1]. It also appears in [ASH], [SB2], [Ve1].

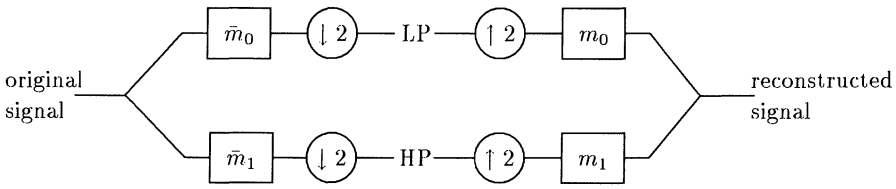


Figure 2

Subband coding scheme corresponding to the FWT algorithm. The sign $2\downarrow$ stands for "decimation of one sample out of two" and $2\uparrow$ for the insertion of zeros at the intermediate values.

Since $m_0(\omega)$ is regular (it is a trigonometric polynomial) and since $m_0(0) = 1$, we can iterate (2.2) to obtain

$$\hat{\varphi}(\omega) = \prod_{k=1}^{+\infty} m_0(2^{-k}\omega) . \tag{2.11}$$

Given a conjugate quadrature filter $m_0(\omega)$ (*i.e.* a trigonometric polynomial satisfying (2.5) and (2.7)), it is thus possible to define the scaling

function, either as a solution of the two scale difference equation (1.5), or explicitly with the above infinite product. However, this does not always lead to a multiresolution analysis: the function $\varphi(x) = \frac{1}{3}\chi_{[0,3]}$ generated by the CQF $m_0(\omega) = (1 + e^{3i\omega})/2$, for example, does not satisfy the orthonormality of the translates. Orthonormality of the $\varphi(x-k)$ turns out to be equivalent to the L^2 convergence of the truncated products $\hat{\varphi}_n(\omega) = \prod_{k=1}^{n} m_0(2^{-k}\omega)\chi_{[-2^n\pi, 2^n\pi]}(\omega)$ to $\hat{\varphi}(\omega)$ (because $\{\varphi_n(x-k)\}_{k\in\mathbb{Z}}$ is an orthonormal set as soon as (2.5) is satisfied).

More precisely, the following result characterizes the subclass of CQF filters leading to a multiresolution analysis and orthonormal basis of wavelets.

Theorem 2.1. *Let $m_0(\omega)$ be a Conjugate Quadrature Filter. Then, the infinite product* (2.11) *leads to a multiresolution analysis if and only if there exist a compact set $K \subset \mathbb{R}$ such that,*

i) *K contains a neighbourhood of the origin,*
ii) *$|K| = 2\pi$ and for all ω in $[-\pi, \pi]$, there exist $n \in \mathbb{Z}$ such that $\omega + 2n\pi \in K$,*
iii) *for all $n > 0$, $m_0(2^{-n}\omega)$ does not vanish on K.*

The set K is said to be "congruent to $[-\pi, \pi]$ modulo 2π" (figure 3). The proof of this result can be found in [Co1]. It exploits the continuity of m_0, the compactness of K and $m_0(0) = 1$ to show that (iii) is equivalent to $\hat{\varphi}(\omega) \geq c > 0$ on K. This is then sufficient to derive the L^2 convergence of the φ_n by Lebesgue's Theorem. We shall use a multidimensional generalization of Theorem 2.1 in the fourth section.

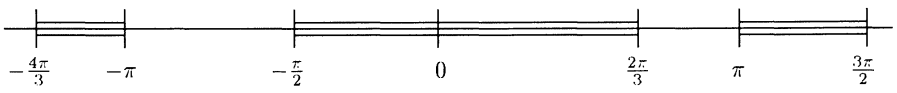


Figure 3
Example of compact set congruent to $[-\pi, \pi]$ modulo 2π.

II.1.b. The biorthogonal case.

The conjugate quadrature filters are a very particular case of subband coding scheme with perfect reconstruction, because identical filters (up to a complex conjugation) are used for both the decomposition and the reconstruction stages. If we do not impose this restriction, then the scheme uses four different filters: $\overline{\tilde{m}_0(\omega)}$ and $\overline{\tilde{m}_1(\omega)}$ for the decomposition, $m_0(\omega)$ and $m_1(\omega)$ for the reconstruction. Perfect reconstruction for any discrete signal is then ensured if,

$$\begin{cases} \overline{m_0(\omega)}\,\tilde{m}_0(\omega) + \overline{m_1(\omega)}\,\tilde{m}_1(\omega) = 1 \\ \tilde{m}_0(\omega+\pi)\,\overline{m_0(\omega)} + \tilde{m}_1(\omega+\pi)\,\overline{m_1(\omega)} = 0\,. \end{cases} \tag{2.12}$$

$\tilde{m}_0(\omega)$ and $\tilde{m}_1(\omega)$ may thus be regarded as the solutions of a linear system. However, to avoid the infinite impulse response solutions, we shall force the determinant of this system to be $\alpha e^{ik\omega}$, $\alpha \neq 0$, $k \in \mathbb{Z}$. For sake of convenience we take $\alpha = -1$ and $k = 1$ (a change of these values would only mean a shift and a scalar multiplication on the impulse response of our filters). This leads to

$$\overline{m_0(\omega)}\,\tilde{m}_0(\omega) + \overline{m_0(\omega+\pi)}\,\tilde{m}_0(\omega+\pi) = 1\,, \tag{2.13}$$

and

$$m_1(\omega) = e^{-i\omega}\,\overline{\tilde{m}_0(\omega+\pi)}\,, \quad \tilde{m}_1(\omega) = e^{-i\omega}\,\overline{m_0(\omega+\pi)}\,. \tag{2.14}$$

The formulas (2.13) and (2.14) are thus the most general setting for finite impulse response subband coders with exact reconstruction (in the two channel case). The functions $m_0(\omega)$ and $\tilde{m}_0(\omega)$ are called "dual filters". It is clear that the special case $m_0(\omega) = \tilde{m}_0(\omega)$ corresponds to the conjugate quadrature filters of II.1.a. However, dual filters are easier to design than CQF's. For example, if m_0 is fixed, $\tilde{m}_0$ can be found as the solution of a Bezout problem which is equivalent to a linear system. The coefficients of these filters can be very simple numerically (in particular they can have finite binary expansion which is very useful for practical implementation), furthermore they can be chosen symmetrical ("linear phase filter"), a property which is impossible to satisfy in the CQF case.

We can mimic, in this more general framework, the construction of orthonormal wavelets from CQF. Assuming that $m_0(0) = \tilde{m}_0(0) = 1$ and $m_0(\pi) = \tilde{m}_0(\pi) = 0$, we define

$$\hat{\varphi}(\omega) = \prod_{k=1}^{+\infty} m_0(2^{-k}\omega)\,, \tag{2.15}$$

$$\hat{\psi}(2\omega) = m_1(\omega)\hat{\varphi}(\omega) , \tag{2.16}$$

$$\hat{\tilde{\varphi}}(\omega) = \prod_{k=1}^{+\infty} \tilde{m}_0(2^{-k}\omega) , \tag{2.17}$$

$$\hat{\tilde{\psi}}(2\omega) = \tilde{m}_1(\omega)\hat{\tilde{\varphi}}(\omega) . \tag{2.18}$$

In [CDF], the following theorem was proved,

Theorem 2.2.

- *If* $\hat{\varphi}_n(\omega) = \prod_{k=1}^{n} m_0(2^{-k}\omega)\chi_{[-2^n\pi, 2^n\pi]}(\omega)$ *and* $\hat{\tilde{\varphi}}_n(\omega) = \prod_{k=1}^{n} \tilde{m}_0(2^{-k}\omega)\ \chi_{[-2^n\pi, 2^n\pi]}(\omega)$ *converge in* $L^2(\mathbb{R})$ *respectively to* $\hat{\varphi}(\omega)$ *and* $\hat{\tilde{\varphi}}(\omega)$, *then the following duality relations are satisfied*

$$\langle \varphi(x-k), \tilde{\varphi}(x-k') \rangle = \delta_{k,k'} \tag{2.19}$$

$$\langle \psi_k^j, \tilde{\psi}_{k'}^{j'} \rangle = \delta_{j,j'}\ \delta_{k,k'} \tag{2.20}$$

and for all f *in* $L^2(\mathbb{R})$ *one has the unique decomposition*

$$f = \lim_{J \to +\infty} \sum_{j=-J}^{J} \sum_{k \in \mathbb{Z}} \langle f, \psi_k^j \rangle\ \tilde{\psi}_k^j \tag{2.21}$$

(*in the* L^2 *sense*).

- *If* φ *and* $\tilde{\varphi}$ *satisfy* $|\hat{\varphi}(\omega)| + |\hat{\tilde{\varphi}}(\omega)| \leq C\,(1 + |\omega|)^{-1/2-\varepsilon}$ *for some* $\varepsilon > 0$, *then the families* $\{\psi_k^j\}_{j,k\in\mathbb{Z}}$ *and* $\{\tilde{\psi}_k^j\}_{j,k\in\mathbb{Z}}$ *are frames of* $L^2(\mathbb{R})$.
- *When these two properties hold, then* $\{\psi_k^j,\ \tilde{\psi}_k^j\}_{j,k\in\mathbb{Z}}$ *are biorthogonal (or dual) Riesz bases of* $L^2(\mathbb{R})$.

Many examples of these systems can be found in [CDF] and a sharper analysis of the frame conditions is developed in [CD]. We now recall a practical way of constructing φ and ψ numerically from a given subband coding scheme.

II.2. The cascade algorithm.

In the last section we saw that the scaling function $\varphi(x)$ could be approximated, at least in $L^2(\mathbb{R})$, by a sequence of band limited functions

$\{\varphi_n\}_{n>0}$ defined by

$$\hat{\varphi}_n(\omega) = \prod_{j=1}^{n} m_0(2^{-j}\omega)\ \chi_{[-2^n\pi, 2^n\pi]}(\omega)\ . \tag{2.22}$$

These functions are characterized by their sampled values at the points $2^{-n}k$ ($k \in \mathbb{Z}$), *i.e.*

$$s_k^n = \varphi_n(2^{-n}k)\ . \tag{2.23}$$

This sequence can also be considered as the impulse response of the transfer function

$$S_n(\omega) = 2^n \prod_{j=1}^{n-1} m_0(2^j\omega)\ . \tag{2.24}$$

$S_n(\omega)$ can be obtained recursively by the formula

$$S_{n+1}(\omega) = 2\, m_0(\omega)\ S_n(2\omega)\ . \tag{2.25}$$

In the time domain, (2.25) becomes an interpolation scheme; the sequence s_k^n is dilated by insertion of zeros ($S_n(\omega) \rightarrow S_n(2\omega)$) before being filtered (multiplication by $2\, m_0(\omega)$). We have thus,

$$s_p^{n+1} = 2 \sum_{k\in\mathbb{Z}} c_{p-2k}\ s_k^n\ . \tag{2.26}$$

This iterative process, which computes the $\{s_k^n\}_{k\in\mathbb{Z}}$ sequences from an initial Dirac sequence $\delta_{0,k}$ is called the "cascade algorithm". We illustrate it on figure 4 (our sequences are represented by piecewise constant functions).

Note that it identifies exactly with the reconstruction stage in the FWT algorithm described in II.1.a. The scaling function is thus approached by the reconstructed signal from a single approximation coefficient at a coarse scale. Similarly, the wavelet will be obtained by starting the reconstruction from a detail coefficient at a coarse scale (and thus applying $m_1(\omega)$ at the first step of the cascade).

This explains why subband coding schemes associated with regular wavelets are particularly interesting: the smoothness of the wavelet

determines the appearance of the coarse scale components of the reconstructed signal. A smooth appearance is important for many applications such as compression where a big part of the finer scale information is thrown away.

In the biorthogonal case, the analysis and the synthesis wavelets (ψ and $\tilde{\psi}$) need not have the same regularity. As just discussed, smoothness

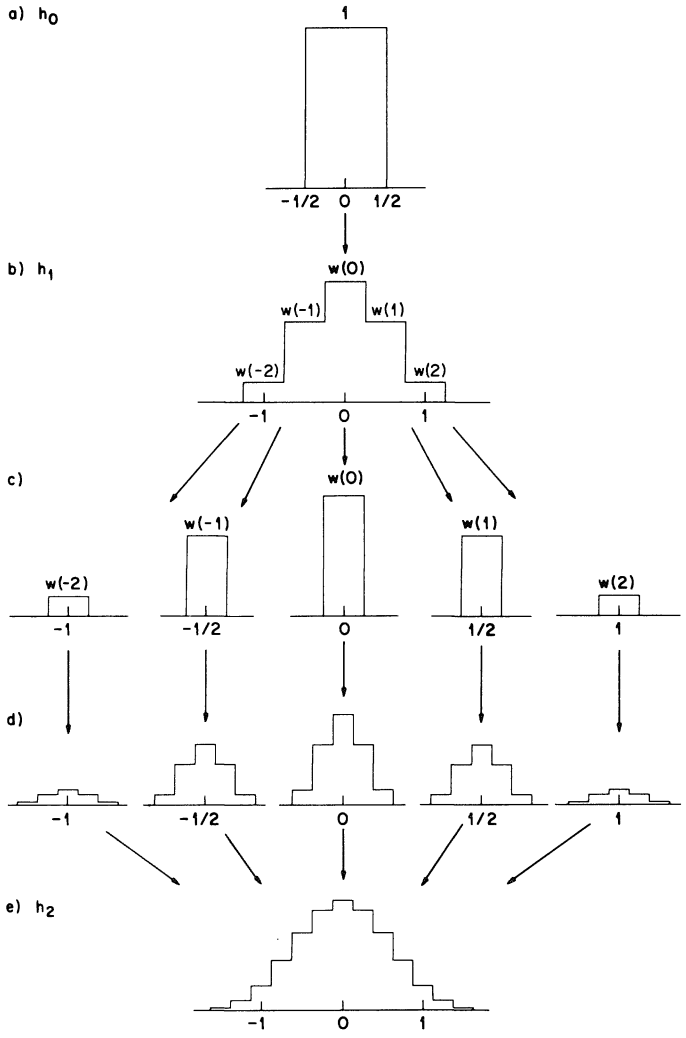


Figure 4
The cascade algorithm (from [Dau1]).

is important for the reconstructing function; the analyzing function needs only to be sufficiently regular to ensure that the wavelet bases are unconditional, so that the FWT algorithm is stable. Note that an important property on the analyzing wavelet is cancellation, *i.e.* vanishing moments, ensuring small high scale coefficients for smooth regions in the function or signal to be analyzed.

Let us finally mention that this type of "refinement method" is well known in approximation theory as "stationary subdivision" (*e.g.* [CDM], [DyL]). Most of these papers are motivated by interpolation problems, where smooth curves or surfaces need to be constructed, connecting (or close to) given sparse data points. Consequently, they are mainly concerned with what we call the reconstruction stage and they do not study the existence of an associated subband coding scheme. This also means that they do not care about an easy way of encoding or representing the extra "detail information" ($\longleftrightarrow W_j$) that can be added in going from one refinement level to the next one ($V_j \to V_{j-1}$). On the other hand, the subband coding literature seldom mentions the importance of the smoothness appearing in the cascade of the reconstruction from the low scales. Orthonormal and biorthogonal wavelet bases lead to an elegant combination of these two approaches.

We now present several different methods to estimate the regularity of the wavelets associated to a given subband coding scheme. We shall concentrate on the regularity of the scaling function which determines the regularity of the wavelet itself because $\psi(x)$ is a finite linear combination of translates of $\varphi(2x)$. Whatever the method used, if a global regularity of order r is achieved, then the cascade algorithm also converges uniformly up to this order (see [Dau1], [DL], [Co2]).

II.3. Regularity: the spectral approach.

II.3.a. A Fourier estimation of the Hölder exponent.

Let us denote by C^α the Hölder space defined as follows. For $\alpha = n + \beta$, $\beta \in [0, 1[$, $f \in C^\alpha$ if and only if it is n times continuously differentiable and for all $x \neq y$,

$$\frac{|f^n(x) - f^n(y)|}{|x - y|^\beta} \leq C(f) .$$

Define also

$$\mathcal{F}_p^\alpha = \{f : \ (1 + |\omega|)^\alpha \hat{f}(\omega) \in L^p\} \qquad (\alpha \geq 0, \ p \geq 1) . \tag{2.27}$$

It is well known (and easy to check) that $\mathcal{F}_\infty^{\alpha+1+\varepsilon} \subset \mathcal{F}_1^\alpha \subset C^\alpha$, for $\varepsilon > 0$. For compactly supported functions f, we also have

$$f \in C^\alpha \text{ implies } f \in \mathcal{F}_\infty^\alpha \tag{2.28}$$

so that the decay of the Fourier transform can be used to evaluate the global regularity. To estimate this decay in the case of the scaling function, it is possible to use the factorization of $m_0(\omega)$; due to its cancellation at $\omega = \pi$, we have indeed

$$m_0(\omega) = \left(\frac{1+e^{i\omega}}{2}\right)^N p(\omega) . \tag{2.29}$$

The infinite product (2.11) is thus divided in two parts. The first part, which comes from the factor $\left((1+e^{i\omega})/2\right)^N$ gives decay, since

$$\left|\prod_{k=1}^{+\infty}\left(\frac{1+e^{i2^{-k}\omega}}{2}\right)\right| = \left|\prod_{k=2}^{+\infty}\cos(2^{-k}\omega)\right| = \left|\frac{2}{\omega}\sin(\frac{\omega}{2})\right| . \tag{2.30}$$

The second part, which involves the factor $p(\omega)$, can be controlled by a polynomial expression. Indeed, since $p(0) = 1$ and p is a regular function, the infinite product generated by the second factor satisfies

$$\left|\prod_{k=1}^{+\infty} p(2^{-k}\omega)\right| \leq C \prod_{1\leq k<\log(1+|\omega|)/\log 2} |p(2^{-k}\omega)| . \tag{2.31}$$

Defining, for $j > 0$,

$$B_j = \sup_{\omega\in\mathbb{R}} \left|\prod_{k=0}^{j-1} p(2^k\omega)\right| \tag{2.32}$$

and

$$b_j = \frac{\log B_j}{j \log 2} , \tag{2.33}$$

we obtain

$$\left|\prod_{k=1}^{+\infty} p(2^{-k}\omega)\right| \leq C(B_j)^{\log(1+|\omega|)/\log 2} \leq C\,(1+|\omega|)^{b_j} \tag{2.34}$$

and

$$|\tilde{\varphi}(\omega)| \leq C\,(1+|\omega|)^{b_j - N}\,. \tag{2.35}$$

Consequently, φ is in $\mathcal{F}_1^\alpha$ and C^α if $\alpha < N - b_j - 1$ for some $j > 0$. We see here that N must be large to allow high regularity since b_j is always positive. In fact, one can prove that if the wavelet is r times continuously differentiable then it has at least $r+1$ vanishing moments (see [Me1], [Dau1]), *i.e.*

$$\left(\frac{d}{d\omega}\right)^n (\hat{\psi})(0) \;=\; \left(\frac{d}{d\omega}\right)^n (m_0)(\pi) \;=\; 0\,,$$

for $n \;=\; 0,\ldots,r+1$ and thus $N \geq r+1$. These cancellations are also known as the Fix-Strang conditions [FS]; they are equivalent to the property that the polynomials of order $N-1$ can be expressed as linear combinations of the $\{\varphi(x-k)\}_{k\in\mathbb{Z}}$. However, these conditions are necessary but not sufficient to ensure the regularity of the scaling function since the effect of N may be killed by a large value of b_j. Fortunately, this can be avoided by a careful choice of the filter $m_0(\omega)$ (and, in the biorthogonal case, additionally $\tilde{m}_0(\omega)$).

In the CQF-orthonormal case, a particular family of FIR filters indexed by N has been constructed in [Dau1]. This construction uses the polynomial

$$P_N(y) \;=\; \sum_{j=0}^{N-1} \binom{N-1+j}{j}\, y^j \tag{2.36}$$

(with the shorthand notation $y \;=\; \sin^2(\omega/2)$), which is the lowest degree solution of the Bezout problem

$$P_N(y)(1-y)^N \;+\; y^N P_N(1-y) \;=\; 1\,. \tag{2.37}$$

The corresponding filters are defined by

$$m_0^N(\omega) \;=\; \left(\frac{1+e^{i\omega}}{2}\right)^N p_N(\omega) \tag{2.38}$$

with

$$|p_N(\omega)|^2 \;=\; P_N(y) \;=\; P_N\left(\frac{1-\cos\omega}{2}\right)\,. \tag{2.39}$$

The Fejer-Riesz lemma guarantees that there exists a FIR filter $p_N(\omega)$ which satisfies (2.39). It is clear that the CQF condition (2.5) is equivalent to (2.36) and the conditions in Theorem 2.1 are trivially satisfied with $K = [-\pi, \pi]$. For large values of N, the regularity $\alpha(N)$ of the associated scaling function is approximately $0.2\,N$ and the exact asymptotic ratio between $\alpha(N)$ and N can be determined. Intuitively speaking, this means that the contribution of $p_N(\omega)$ removes "eighty percent of the regularity" brought by the factor $\left((1+e^{i\omega})/2\right)^N$. For this estimation, we need to optimize the inequality (2.35), *i.e.* find the best possible exponent for the decay of $\hat{\varphi}(\omega)$.

II.3.b. Optimal and asymptotical Fourier estimation: The role of fixed points.

We start by defining "the critical exponent of $m_0(\omega)$":

$$b = \inf_{j>0} b_j = \inf_{j>0} \max_{\omega\in\mathbb{R}} \frac{1}{j\log 2} \log\left|\prod_{k=0}^{j-1} p(2^k\omega)\right| . \tag{2.40}$$

Then, it was proved in [Co2] that under the hypothesis $|p(\pi)| > |p(0)| = 1$ (satisfied in the present case (2.39)), $\hat{\varphi}(\omega)$ cannot have a better decay at infinity than $|\omega|^{b-N}$. If the infimum b is attained for some finite j, $b = b_j$, then this estimate is optimal.

How can we estimate the critical exponent? A first method consists in evaluating b_j for large values of j. Indeed, b is also the limit of the sequence b_j because the boundedness of p implies $b_J \le b_j + O(j/J)$. This may however require heavy computations.

In several cases, it is possible to use a more powerful method based on the transformation $\tau : \omega \mapsto 2\omega \bmod 2\pi$ and the fixed points of its powers τ^n, $n > 0$. Indeed, let ω_0 be a fixed point of τ^n for $n > 0$ and define its orbit $\omega_j = \tau^j\omega_0$, for $j = 0, \ldots, n-1$. Since $p(\omega)$ has period 2π, we have

$$p(2^{nk}\omega_j) = p(\omega_j), \quad \text{for all } k > 0 \tag{2.41}$$

and consequently

$$b_{nk} \ge \frac{1}{n\log 2} \log\left|\prod_{j=0}^{n-1} p(\omega_j)\right| . \tag{2.42}$$

Letting k go to $+\infty$, this leads to

$$b \geq \frac{1}{n \log 2} \log \left| \prod_{j=0}^{n-1} p(\omega_j) \right| . \tag{2.43}$$

Fixed points of τ lead therefore to lower bounds for b and upper bounds for the regularity index. In fact they can do much better and provide optimal estimates for certain types of filters. Let us consider the smallest orbit of τ different from $\{0\}$, namely the pair $\{-2\pi/3,\ 2\pi/3\}$. Note that, because our filters have real coefficients, $|m_0(\omega)|$ and $|p(\omega)|$ are even functions so that $|p(2\pi/3)| = |p(-2\pi/3)|$. The following result associates the value $|p(2\pi/3)|$ and the critical exponent b.

Theorem 2.3. *Suppose that $p(\omega)$ satisfies*

$$|p(\omega)| \leq \left| p\left(\frac{2\pi}{3}\right) \right| \quad \text{if } |\omega| \leq \frac{2\pi}{3} , \tag{2.44}$$

$$|p(\omega)p(2\omega)| \leq \left| p\left(\frac{2\pi}{3}\right) \right|^2 \quad \text{if } \frac{2\pi}{3} \leq |\omega| \leq \pi . \tag{2.44'}$$

Then

$$b = \frac{1}{\log 2} \log \left| p\left(\frac{2\pi}{3}\right) \right| . \tag{2.45}$$

PROOF. We already know from (2.43) that $b \geq \log |p(2\pi/3)| / \log 2$. We now use the bounds on p to find an upper bound for b_j, $j > 0$. We can regroup the factors in (2.32) by packets of one or two elements in order to apply either (2.44) or (2.44') on each block. Since only the last factor can miss one of these two inequalities, we obtain

$$\left| \prod_{k=0}^{j-1} p(2^k \omega) \right| \leq \left| p\left(\frac{2\pi}{3}\right) \right|^{j-1} \sup |p| , \tag{2.46}$$

and thus,

$$b_j \leq \frac{1}{\log 2} \left[\frac{j-1}{j} \log \left| p\left(\frac{2\pi}{3}\right) \right| + \frac{\sup [\log |p|]}{j} \right] , \tag{2.47}$$

which leads to

$$b \leq \frac{1}{\log 2} \log \left| p\left(\frac{2\pi}{3}\right) \right| \tag{2.48}$$

and to (2.45).

The equality (2.45) means that the worst decay of $\hat{\varphi}(\omega)$ occurs for the sequence $\omega_k = 2^n \pi/3$, $n > 0$. This is interesting, because (2.44) and (2.44′) turn out to be satisfied in many cases and in particular for the whole family of CQF defined by (2.38), (2.39). This is easy to check directly for small values of N, since the inequalities can be rewritten as

$$P_N(y) \leq P_N\left(\frac{3}{4}\right) \quad \text{if } y \leq \frac{3}{4}\,, \tag{2.49}$$

$$P_N(y)\ P_N(4y(1-y)) \leq \left(P_N\left(\frac{3}{4}\right)\right)^2 \quad \text{if } \frac{3}{4} \leq y \leq 1\,. \tag{2.49'}$$

The discussion for general N is more difficult and we refer to [CC] for a complete proof of (2.49), (2.49′). However, a similar result can be obtained in a simple way. To characterize the asymptotical behavior of the critical exponent when N goes to $+\infty$, one does not need the full force of (2.44), (2.44′), however. It can also be derived from a weaker, asymptotically valid inequality, as proved by H. Volkner in [V].

Theorem 2.4. *Let $b(N)$ be the critical exponent associated to $m_0^N(\omega)$ and $\alpha(N)$ the Hölder exponent of the corresponding scaling function. Then*

$$\lim_{N \to +\infty} \frac{b(N)}{N} = \frac{\log 3}{2 \log 2} \tag{2.50}$$

and

$$\lim_{n \to +\infty} \frac{\alpha(N)}{N} = \lim_{N \to +\infty} \frac{N - b(N)}{N} = 1 - \frac{\log 3}{2 \log 2} \simeq 0.2075\,. \tag{2.50'}$$

PROOF. This result can be viewed as a consequence of Theorem 2.3, but it can also be proved directly by using some properties of $P_N(y)$. Let us write (2.36) in the following form:

$$P_N(y) = \sum_{j=0}^{N-1} \binom{N-1+j}{j} \left(\frac{1}{2}\right)^j (2y)^j\,. \tag{2.51}$$

From (2.36) we see that $P_N(1/2) = 2^{N-1}$; since P_N is an increasing function between 0 and 1, we have

$$P_N \leq (\max\{4y, 2\})^{N-1} = |g(y)|^{N-1} . \tag{2.52}$$

It is now trivial to check that (2.49) and (2.49′) are satisfied if we replace $P_N(y)$ by $g(y)$. The same argument used in the proof of Theorem 2.3 leads then to

$$b(N) \leq \frac{N-1}{2\log 2} \log\left|g\left(\frac{3}{4}\right)\right| = \frac{N-1}{2\log 2} \log 3 \tag{2.53}$$

but from (2.43) we get

$$\begin{aligned} b(N) &\geq \frac{1}{2\log 2} \log\left|P_N\left(\frac{3}{4}\right)\right| \\ &\geq \frac{1}{2\log 2} \log\left|\binom{2N-2}{N-1}\left(\frac{3}{4}\right)^{N-1}\right| \geq \frac{N-2}{2\log 2} \log 3 . \end{aligned} \tag{2.54}$$

This proves the limit (2.50), and consequently (2.50′) since the decay index of the Fourier transform is equivalent to the Hölder exponent when both tend to $+\infty$.

The use of fixed points for optimal estimations of the spectral decay is thus very efficient when one is looking for arbitrarily high regularity since a sharp asymptotical result is obtained. For small filters, this method does not give a good result because the error on the exact regularity may have the same order as the value of the Hölder exponent itself. For such filters, other methods, which take advantage of the small number of taps in the filter, can be used to derive more precise estimations. We now describe these methods; they are typically based on matrix computations.

II.4. Regularity: Matrix based sharper estimates.

II.4.a. The Littlewood-Paley approach.

We first recall some aspects of the Littlewood-Paley theory. Let $\gamma(x)$ be a real-valued, symmetrical function of the Schwartz class $\mathcal{S}(\mathbb{R})$, which satisfies

$$\begin{cases} \hat{\gamma}(\omega) = 0 & \text{if } |\omega| \leq 1/2 \text{ or } |\omega| \geq 5/2 , \\ \hat{\gamma}(\omega) > 0 & \text{if } 1/2 < |\omega| < 5/2 , \end{cases} \tag{2.55}$$

so that the frequency axis is covered by the dyadic dilations of γ. Indeed, we have

$$0 < C_1 \leq \sum_{j=-\infty}^{+\infty} \hat{\gamma}(2^j\omega) \leq C_2 \quad \text{if } \omega \neq 0 \,. \tag{2.56}$$

Define for any f in $\mathcal{S}'(\mathbb{R})$ the dyadic blocks $\Delta_j(f)$ by

$$\Delta_j(f) = 2^j\gamma(2^j\cdot) * f \iff \hat{\Delta}_j(f) = \hat{\gamma}(2^{-j}\cdot)\hat{f} \,. \tag{2.57}$$

The Littlewood-Paley theory tells us that several functional spaces can be characterized by examining only the L^p norm of these blocks. This is the case in particular for the Sobolev spaces $W^{p,s}$ and the Hölder spaces C^α, $\alpha > 0$. To do this, it is necessary to change slightly the definition of C^α when α is an integer; we shall say that a bounded function f is in C^n if and only if f^{n-1} belongs to the Zygmund class Λ, *i.e.* there exists a constant C such that, for all x and y, we have

$$\left|f^{n-1}(x+y) + f^{n-1}(x-y) - 2f^{n-1}(x)\right| \leq C\,|y| \,. \tag{2.58}$$

With this convention, the Hölder space C^α is characterized by the following conditions,

$$\|\Delta_j(f)\|_{L^\infty} \leq C\,2^{-\alpha j} \quad \text{when } j \geq 0 \,, \tag{2.59}$$

$$f \text{ is a bounded continuous function.} \tag{2.59'}$$

Note that the choice (2.55) for γ is arbitrary and that more general functions could be chosen to divide the Fourier domain into dyadic blocks. To derive these types of estimates on the scaling function φ, we introduce a tool which will be very useful in the bidimensional case.

Definition 2.1. *Let $L^2[0, 2\pi]$ be the space of 2π-periodic, square integrable functions on $[0, 2\pi]$, and $C[0, 2\pi]$ the space of 2π-periodic continuous functions. Then, for any $m(\omega)$ in $C[0, 2\pi]$, we define the transition operator T_m associated to $m(\omega)$ by*

$$\begin{cases} T_m : L^2[0, 2\pi] \longrightarrow L^2[0, 2\pi] \\ f \mapsto T_m f(\omega) = m\left(\dfrac{\omega}{2}\right) f\left(\dfrac{\omega}{2}\right) \\ \qquad\qquad + m\left(\dfrac{\omega}{2}+\pi\right) f\left(\dfrac{\omega}{2}+\pi\right) \,. \end{cases} \tag{2.60}$$

Note that when $m(\omega)$ is a trigonometric polynomial, the study of T_m can be made in a finite dimensional space. More precisely, if we define

$$(2.61)\quad E(N_1, N_2) = \left\{ \sum_{n=N_1}^{N_2} h_n e^{in\omega} : \quad (h_{N_1}, \ldots, h_{N_2}) \in \mathbb{C}^{N_2-N_1+1} \right\}$$

then we have clearly

$$(2.62)\qquad (f, m) \in [E(N_1, N_2)]^2 \text{ implies } T_m f \in E(N_1, N_2) .$$

This is due to the contraction $\omega \mapsto \omega/2$ which appears in the definition (2.60) of T_m. If c_n is the n-th Fourier coefficient of $m(\omega)$, then the matrix of T_m in the basis of the complex exponentials is given by

$$(2.63)\qquad T_{\ell,n} = (2\, c_{2\ell-n}) .$$

The size of this matrix P in $E(N_1, N_2)$ is $L \times L$ with $L = N_2 - N_1 + 1$. This operator has been studied by J. P. Conze and A. Raugi and several ideas presented below are due to their work [CR], [Con]. We shall use it to derive Littlewood-Paley type of estimations for the Hölder continuity of the scaling function. For this, we need the following result.

Lemma 2.5. *For all $n > 0$,*

$$(2.64)\qquad \int_{-\pi}^{\pi} (T_m)^n f(\omega) d\omega = \int_{-2^n\pi}^{2^n\pi} f(2^{-n}\omega) \prod_{k=1}^{n} m(2^{-k}\omega) d\omega .$$

PROOF. We prove it by induction. It is clear for $n = 1$ since

$$\begin{aligned}\int_{-\pi}^{\pi} T_m f(\omega) d\omega &= \int_{-\pi}^{\pi} \left[m\left(\frac{\omega}{2}\right) f\left(\frac{\omega}{2}\right) + m\left(\frac{\omega}{2}+\pi\right) f\left(\frac{\omega}{2}+\pi\right) \right] d\omega \\ &= 2\int_{-\pi/2}^{\pi/2} [m(\omega)f(\omega) + m(\omega+\pi)f(\omega+\pi)]\, d\omega \\ &= 2\int_{-\pi}^{\pi} m(\omega)f(\omega)d\omega = \int_{-2\pi}^{2\pi} m\left(\frac{\omega}{2}\right) f\left(\frac{\omega}{2}\right) d\omega .\end{aligned}$$

Assuming (2.64) for n, we obtain at the next step,

$$\int_{-\pi}^{\pi} (T_m)^{n+1} f(\omega)\, d\omega = \int_{-\pi}^{\pi} (T_m)^n T_m f(\omega)\, d\omega$$

$$= \int_{-2^n\pi}^{2^n\pi} \left[\prod_{k=1}^{n} m(2^{-k}\omega)\right] \left[m(2^{-n-1}\omega) f(2^{-n-1}\omega) + m(2^{-n-1}\omega+\pi) f(2^{-n-1}\omega+\pi)\right] d\omega$$

$$= 2^{n+1} \int_{-\pi/2}^{\pi/2} \left[\prod_{k=1}^{n} m(2^{k}\omega)\right] \left[m(\omega) f(\omega) + m(\omega+\pi) f(\omega+\pi)\right] d\omega$$

$$= \int_{-2^{n+1}\pi}^{2^{n+1}\pi} \left[\prod_{k=1}^{n+1} m(2^{-k}\omega)\right] f(2^{-n-1}\omega)\, d\omega \,.$$

This concludes the proof.

We now suppose that $m(\omega)$ is a positive trigonometric polynomial in $E_M = E(-M, M)$ and that $m(0) = 1$ and $m(\pi) = 0$. Then m can be factorized as

$$m(\omega) = \cos^{2N}\left(\frac{\omega}{2}\right) p(\omega) \tag{2.65}$$

where $p(\omega)$ is a trigonometric polynomial that does not vanish for $\omega = \pi$. Note that necessarily $N \leq M$. From this cancellation property, we can derive,

Lemma 2.6. *$\{1, 1/2, \ldots, 2^{-2N+1}\}$ are eigenvalues of T_m. The row vectors $p_j = (n^j)_{n=-M,\ldots,M}$, for $0 \leq j \leq 2N-1$ generate a subspace which is left invariant by T_m and contains one eigenvector for each of these $2N$ eigenvalues.*

Consequently, the orthogonal subspace defined by

$$F_N = \left\{ \sum_{n=-M}^{M} h_n e^{-in\omega} : \sum_{n=-M}^{M} n^j h_n = 0,\ j = 0, \ldots, 2N-1 \right\} \tag{2.66}$$

is right invariant by T_m .

PROOF. The factorization in (2.65) is equivalent to the cancellation rules

$$\sum_{n=-M}^{M} (-1)^n\, n^j c_n = 0 \quad \text{for } j = 0, \ldots, 2N-1 \,. \tag{2.67}$$

In particular, for $j = 0$, we have

$$\sum_n c_{2n} = \sum_n c_{2n+1} = \frac{1}{2} \quad \text{(because } m(0) = 1\text{)} . \tag{2.68}$$

This means that the sum of each column in the matrix of T (2.63) is equal to 1 and that $p_0 = (1, \ldots, 1)$ is a left eigenvector for the eigenvalue 1. For $0 < j \leq 2N-1$ we define $q_j = p_j P = (q_j^{-M}, \ldots, q_j^M)$; we have,

$$q_j^\ell = \sum_n n^j c_{2n-\ell} . \tag{2.69}$$

Thus, if ℓ is even

$$q_j^\ell = \sum_n \left(n + \frac{\ell}{2}\right)^j c_{2n} \tag{2.70}$$

and if ℓ is odd

$$q_j^\ell = \sum_n \left(n + \frac{1}{2} + \frac{\ell}{2}\right)^j c_{2n+1} . \tag{2.70'}$$

Using the binomial formula and the cancellation rules (2.67), we see that q_j is a linear combination of p_k for $k = 0, \ldots, j$. The coefficient of p_j is given by the last term of the binomial and is thus equal to 2^{-j}. Consequently $\{p_j\}_{j=0,\ldots,2N-1}$ is a triangular basis for the left action of T_m and the eigenvalues are $\{2^{-j}\}_{j=0,\ldots,2N-1}$.

We now come back to the scaling function φ, given by the infinite product

$$\hat{\varphi}(\omega) = \prod_{k=0}^{+\infty} m(2^{-k}\omega) . \tag{2.71}$$

Theorem 2.7. *Let F_N be the invariant subspace of T_m defined by (2.66). If λ is the eigenvalue of T_m restricted to F_N with largest modulus, and if $|\lambda| < 1$, then, we have, with $\alpha = -\log|\lambda|/\log 2$ (> 0),*

- *φ is in $C^{\alpha-\varepsilon}$ for all $\varepsilon > 0$,*

- *φ is in C^α if the restriction of T_m to the invariant subspace F_λ of eigenvalue λ is purely diagonal (i.e. $= \lambda I$).*

These two estimates are optimal if $\hat{\varphi}(\omega)$ does not vanish on $[-\pi, \pi]$.

PROOF. Consider the trigonometric polynomial

$$C_N(\omega) = (1 - \cos\omega)^N \ . \tag{2.72}$$

It clearly belongs to F_N.

Consequently, for all $n > 0$,

$$\begin{aligned} &\int_{-\pi}^{\pi} (T_m)^n C_N(\omega) d\omega \\ &\qquad \leq (2\pi)^{1/2} \left(\int_{-\pi}^{\pi} |(T_m)^n C_N(\omega)|^2 d\omega \right)^{1/2} \\ &\qquad \leq C(|\lambda| + \varepsilon)^n \quad \text{or } C\,|\lambda|^n \text{ if } T_m|_{F_\lambda} = \lambda I \,. \end{aligned} \tag{2.73}$$

We now use Lemma 2.5 combined with the inequality

$$C_N(\omega) \geq 1 \quad \text{when } \frac{\pi}{2} \leq |\omega| \leq \pi \ . \tag{2.74}$$

This leads us to

$$\begin{aligned} \int_{2^{n-1}\pi \leq |\omega| \leq 2^n \pi} \hat{\varphi}(\omega)\, d\omega &\leq C \int_{2^{n-1}\pi \leq |\omega| \leq 2^n \pi} \prod_{k=1}^{n} m(2^{-k}\omega)\ d\omega \\ &\leq C \int_{-2^n\pi}^{2^n\pi} C_N(2^{-n}\omega) \prod_{k=1}^{n} m(2^{-k}\omega)\ d\omega \\ &= C \int_{-\pi}^{\pi} (T_m)^n\ C_N(\omega)\, d\omega \ . \end{aligned}$$

Consequently the Littlewood-Paley blocks satisfy the inequality

$$\|\hat{\Delta}_j(\varphi)\|_{L^1} \ \leq\ C\, 2^{-(\alpha-\varepsilon)j},\ \varepsilon > 0,\ \alpha = -\log(|\lambda|)/\log 2 \tag{2.75}$$

$$\|\hat{\Delta}_j(\varphi)\|_{L^1} \ \leq\ C\, 2^{-\alpha j}, \quad \text{if } T_m|_{F_\lambda} \text{ is purely diagonal.} \tag{2.75'}$$

Since $\|\Delta_j(\varphi)\|_{L^\infty} \leq \|\hat{\Delta}_j(\varphi)\|_{L^1}$ we obtain the announced regularity.

To prove that these estimates are optimal, we need to reverse all the inequalities which have been used. First, note that since $m(\omega)$ and $\hat{\varphi}(\omega)$ are positive, we have $\|\Delta_j(\varphi)\|_{L^\infty} = \|\hat{\Delta}_j(\varphi)\|_{L^1}$.

Let f_λ be an eigenfunction in F_λ. If $\int f_\lambda(\omega)\, d\omega > 0$, then

$$\int_{-2^n\pi}^{2^n\pi} f_\lambda(2^{-n}\omega) \prod_{k=1}^{n} m(2^{-k}\omega) d\omega = \int_{-\pi}^{\pi} (T_m)^n f_\lambda(\omega) d\omega = \lambda^n \int_{-\pi}^{\pi} f_\lambda(\omega) d\omega \geq C\lambda^n. \tag{2.76}$$

If $\int_{-\pi}^{\pi} f_\lambda(\omega)\, d\omega < 0$, then we replace f_λ by $-f_\lambda$. If $\int_{-\pi}^{\pi} f_\lambda(\omega)\, d\omega = 0$, then the argument has to be modified slightly; see below (after (2.78)). Since we have supposed that $\hat{\varphi}(\omega)$ does not vanish on $[-\pi, \pi]$, we have

$$\hat{\varphi}(\omega) \geq C \prod_{k=1}^{n} m(2^{-k}\omega) \quad \text{for all } n > 0 \text{ and } |\omega| \leq 2^n\pi\ . \tag{2.77}$$

Note that this hypothesis corresponds to the condition of Theorem 2.1 with $K = [-\pi, \pi]$. In a more general setting, we could replace the integrals on $[-2^n\pi, 2^n\pi]$ by integrals on 2^nK and the same results would hold. Combining (2.76) and (2.77) gives

$$\int_{-2^n\pi}^{2^n\pi} |\hat{\varphi}(\omega)|\ |f_\lambda(2^{-n}\omega)| d\omega \geq C|\lambda|^n\ . \tag{2.78}$$

(If $\int_{-\pi}^{\pi} f_\lambda(\omega)\, d\omega = 0$, then a slightly more sophisticated argument will do the trick. Lemma 2.5 still holds if the measure $d\omega$ is replaced by any other measure of the type $g(\omega)\, d\omega$ where g is a 2π-periodic, strictly positive, continuous function. We can always choose g such that

$$\int_{-\pi}^{\pi} f_\lambda(\omega)\ g(\omega)\, d\omega > 0\ ;$$

(2.76) then holds if $d\omega$ is replaced everywhere by $g(\omega)\, d\omega$. Since g is strictly positive, this modified version of (2.76) combined with (2.77), still implies (2.78)).

Since f_λ has a zero of order $2N$ at the origin, the function $\gamma(x)$, defined by $\hat{\gamma}(\omega) = |f_\lambda(\omega)|\, \chi_{[-\pi,\pi]}(\omega)$ is convenient for the Littlewood-Paley analysis of Hölder regularity less than $2N$. This is the case for φ since $2N+1$ vanishing moments would be necessary for a higher Hölder exponent than $2N$ (see [FS], [DL] or [DyL]). Consequently (2.78) tells us that φ cannot be more regular than C^α. To prove the optimality of $C^{\alpha-\varepsilon}$ when $T_m|_{F_\lambda}$ is not purely diagonal, it suffices to replace f_λ by a function g_λ such that $T_m g_\lambda = \lambda g_\lambda + \mu f_\lambda$ with $\mu \neq 0$. This leads to

$$\int_{-2^n\pi}^{2^n\pi} |\hat{\varphi}(\omega)|\, |g_\lambda(2^{-n}\omega)|\, d\omega \geq C\, n\lambda^n \tag{2.78'}$$

which proves the optimality of $C^{\alpha-\varepsilon}$.

The theorem is thus completely proved.

REMARKS.

- The estimates (2.75) and (2.75′) can be found by an equivalent technique, using the transition operator T_p corresponding to the factor $p(\omega)$ in (2.65). We simply consider the eigenvalue λ_p with largest $|\lambda|_p$ and iterate T_p on $f \equiv 1$. This leads to

$$\begin{aligned}\int_{2^{j-1}\pi\leq\omega\leq 2^j\pi} \hat{\varphi}(\omega)d\omega &\leq C \int_{2^{j-1}\pi\leq|\omega|\leq 2^j\pi} |\omega|^{-2N} \left[\prod_{k=1}^{j} p(2^{-k}\omega)\right] d\omega \\ &\leq C\, 2^{-2Nj} \int_{-\pi}^{\pi} (T_p)^j 1\; d\omega \\ &\leq C\,(|\lambda_p| + \varepsilon)^j\; 2^{-2Nj} \\ &\quad \left(\text{or } C|\lambda_p|^j\; 2^{-2Nj} \quad \text{if } T_p/F_{\lambda_p} = \lambda_p I\right)\end{aligned}$$

and thus $\varphi \in C^{\alpha-\varepsilon}$ with $\alpha = 2N - \log|\lambda_p|/\log 2$. This estimate is in fact the same as (2.75). Indeed, if μ is an eigenvalue of T_m in F_N, then its associated eigenfunction can be written as

$$f_\mu = \left(\sin^2\left(\frac{\omega}{2}\right)\right)^N g_\mu(\omega)\,. \tag{2.79}$$

Replacing $m(\omega)$ by its factorized form in

$$\mu f_\mu(\omega) = f_\mu\left(\frac{\omega}{2}\right) m\left(\frac{\omega}{2}\right) + f_\mu\left(\frac{\omega}{2}+\pi\right) m\left(\frac{\omega}{2}+\pi\right) \tag{2.80}$$

we obtain, after dividing by $\left[\sin^2(\omega/2)\,\cos^2(\omega/2)\right]^N$,

$$\mu\, 2^{2N} g_\mu(\omega) = g_\mu\left(\frac{\omega}{2}\right) p\left(\frac{\omega}{2}\right) + g_\mu\left(\frac{\omega}{2}+\pi\right) p\left(\frac{\omega}{2}+\pi\right) . \tag{2.81}$$

We see here that the eigenvalues of T_p are exactly given by $\mu_p = 2^{2N}\mu$. This proves the equivalence between the two techniques.

- In general $m(\omega)$ is not a positive function. One can then define $M(\omega) = |m(\omega)|^2$ and use the operator T_M associated to $M(\omega)$. The result is an estimate of the L^2 norms of $\Delta_j(\varphi)$. Using the Cauchy-Schwarz inequality, we derive the following corollary,

Corollary 2.8. *Suppose that $M(\omega) = |m(\omega)|^2$ has a zero of order $2N$ at $\omega = \pi$. Define λ, the largest eigenvalue of T_M on F_N and $\alpha = -\log\lambda/(2\log 2)$. Then, $\varphi \in H^{\alpha-\varepsilon} \subset C^{\alpha-1/2-\varepsilon}$ where H^s is the Sobolev space of index s. The value α is attained if $T_M|_{F_\lambda} = \lambda I$.*

Note that the Hölder exponent has no chance of being optimal because we have used the Cauchy-Schwarz inequality and $\hat{\varphi}(\omega)$ is not a positive function. The Sobolev exponent however is optimal. The regularity of compactly supported wavelets was estimated with this method in [Dau1].

The transition operator plays also a crucial role in the biorthogonal wavelet theory: we show in Appendix A how it can be used to prove that the families $\{\psi_k^j\}_{j,k\in\mathbb{Z}}$ and $\{\tilde{\psi}_k^j\}_{j,k\in\mathbb{Z}}$ are unconditional bases, with weaker assumptions than the boundedness of $(1+|\omega|)^{1/2+\varepsilon}(|\hat{\varphi}(\omega)| + |\hat{\tilde{\varphi}}(\omega)|)$ imposed in Theorem 2.2.

The optimal estimate for the global and local Hölder regularity of any wavelet can be estimated by another method developed by I. Daubechies and J. Lagarias in [DL]. We now recall its main points.

II.4.b. The time domain approach.

Let $m(\omega) = \sum_{n=0}^{N} c_n e^{in\omega}$ be a trigonometric polynomial such that $m(0) = 1$ and $m(\pi) = 0$. We do not require that $m(\omega)$ be positive. Let $\varphi(x)$ be the scaling function defined by the infinite product (2.71). It is at least a compactly supported distribution in $[0, N]$.

In the time domain approach, we represent $\varphi(x)$ by its "vector" form $w(x) : [0,1] \to \mathbb{R}^N$

$$[w(x)]_n = \varphi(x+n-1), \qquad n = 1, \dots, N. \tag{2.82}$$

From the two scale difference equation (1.5) we get

$$w(x) = \begin{cases} T_0\, w(2x) & \text{if } x \le 1/2, \\ T_1\, w(2x-1) & \text{if } x \ge 1/2, \end{cases} \tag{2.83}$$

where T_0 and T_1 are $N \times N$ matrices defined by

$$(T_0)_{i,j} = c_{2i-j-1} \qquad 1 \le i,j \le N, \tag{2.84}$$

$$(T_1)_{i,j} = c_{2i-j} \qquad 1 \le i,j \le N. \tag{2.84'}$$

Using the notations

$$d_n(x) = n^{\text{th}} \text{ binary digit of } x \in [0,1]$$
$$\tau(x) = \begin{cases} 2x & \text{if } x \le 1/2 \\ 2x-1 & \text{if } x \ge 1/2 \end{cases} \qquad \text{(binary shift)},$$

we can rewrite (2.83) as a "fixed point" equation

$$w(x) = T_{d_1(x)}\, w(\tau(x)). \tag{2.85}$$

This leads to an evaluation of $w(x)$ and its derivative by an iterative process. The regularity of the result depends of course on the spectral properties of T_0 and T_1. Note that when $m(\omega)$ has a zero of order L (as for the transition operator studied in the previous section), then the space F_L orthogonal to the vector $p_j = (n^j)_{n=1,\dots,N}$ for $j = 0, \dots, L-1$ is invariant by T_0 and T_1. This method gives sharp estimates on the local regularity in x by considering the products $T_{d_1(x)} \cdots T_{d_n(x)}$ for all $n \ge 0$. The main result on global regularity proved in [DL; Theorem 3.1] is the following

Theorem 2.9. *Suppose that there exist $\rho < 1$ such that, for all binary sequence $(d_j)_{j\in\mathbb{Z}}$ and all $m > 0$, we have*

$$\| T_{d_1} T_{d_2} \cdots T_{d_m} |_{F_L} \| \le C\rho^m. \tag{2.86}$$

Define $\alpha = -\log \rho / \log 2$. *Then,*

- *if* α *is not an integer,* φ *belongs to* C^α,
- *if* α *is an integer,* $\varphi^{\alpha-1}$ *is almost Lipschitz: for almost all* x, t,

$$|\varphi^{\alpha-1}(x+t) - \varphi^{\alpha-1}(x)| \leq C|t| \; |\log |t|| \; .$$

REMARK.

- The "generalized spectral norm"

$$\rho(T_0, T_1) = \limsup_{m \to \infty} \max_{\substack{d_j = 0 \text{ or } 1 \\ j = 1, \ldots, m}} \| T_{d_1} T_{d_2} \cdots T_{d_m} |_{F_L} \|^{1/m} \tag{2.87}$$

  gives a sharp estimate of the global regularity. Note that it is in general superior to the spectral radius of T_0 and T_1. When N is not too large it is possible to compute the exact value of $\rho(T_1, T_2)$. For example, in the case of orthonormal wavelets, the optimal Hölder exponent was found in [DL] for $N = 4$, 6 and 8. The same evaluation becomes more difficult for larger filters.
- The generalization of this approach in higher dimensions is not trivial. In particular, it involves nonstandard binary expansions depending on the dilation matrix which is used. We describe these techniques in Appendix B.

As a conclusion of this review of regularity estimators, we could say that these three approach are complementary: the time domain method gives sharp results but it is only practicable for small filters, the Littlewood-Paley estimates can be derived for longer filters but they will be optimal only if $m(\omega)$ is a positive function and finally, the Fourier approach is less precise but appropriate to asymptotical results on very large filters. Let us also mention that another method recently developed by O. Rioul [Ri] and based on $\ell^1(\mathbb{Z})$ norms estimates of the iterated filters leads to interesting results; in particular, it is still manageable for larger filters than the time domain method of [DL].

We are now ready to deal with the bidimensional wavelets. We start by examining the different subband coding schemes that can be used to build these non-separable multiscale bases.

III. Two channel bidimensional subband coding schemes.

As mentioned previously, we shall concentrate on the dilation matrices of determinant equal to 2 or -2. In such conditions, the subband coding scheme that we consider split the signal in two channels (instead of four in the separable case) and only one wavelet is then necessary to characterize the detail coefficients at each scale. We first present a short summary of the equations satisfied by these filter. They are immediate generalizations of the results presented in II.1.

III.1. General conditions for exact reconstruction.

As in the one dimensional case, the scheme that we are considering here is based on four fundamental operations:

- The action of two analyzing filters, one low pass

$$\tilde{M}_0(\omega) = \tilde{M}_0(\omega_1, \omega_2)$$

  and one high pass $\tilde{M}_1(\omega) = \tilde{M}_1(\omega_1, \omega_2)$,
- Decimation on each channel by keeping only the samples on the sublattice $\Gamma = D\mathbb{Z}^2$,
- Insertion of zero values at the intermediate points of $\mathbb{Z}^2/\Gamma$,
- Interpolation by two synthesis filters, one low pass

$$M_0(\omega) = M_0(\omega_1, \omega_2)$$

  and one high pass $M_1(\omega) = M_1(\omega_1, \omega_2)$, followed by reconstruction of the original signal by summation.

We see here that the conditions for perfect reconstruction will not depend on the dilation matrix D but only on the sublattice $\Gamma = D\mathbb{Z}^2$ that is generated (different matrices may lead to the same Γ). More precisely, there exist only two types of grid corresponding to a decimation of a factor 2 in $\mathbb{Z}^2$:

- The quincunx sublattice, shown on figure 5, is generated by the integer combinations of $(1,1)$ and $(1,-1)$.
- The column sublattice, shown on figure 6, is generated by the integer combinations of $(0,1)$ and $(2,0)$. It is of course equivalent to the row sublattice, by exchange of the coordinates.

The same arguments that were used in II.1.b show that perfect reconstruction is achieved by FIR filters, if and only if they satisfy (up to a shift) the following equations, which are similar to (2.13) and (2.14).

- In the quincunx case,

$$\overline{M_0(\omega)}\,\tilde{M}_0(\omega) + \overline{M_0(\omega+(\pi,\pi))}\,\tilde{M}_0(\omega+(\pi,\pi)) = 1 \tag{3.1}$$

and

$$\begin{aligned} M_1(\omega) &= e^{-i(\omega_1+\omega_2)}\,\overline{\tilde{M}_0(\omega+(\pi,\pi))}, \\ \tilde{M}_1(\omega) &= e^{-i(\omega_1+\omega_2)}\,\overline{M_0(\omega+(\pi,\pi))}. \end{aligned} \tag{3.2}$$

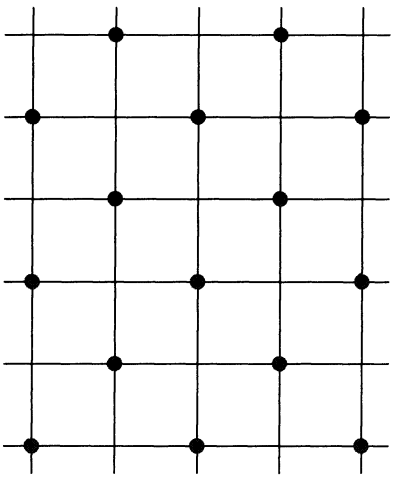


Figure 5
Quincunx decimation.

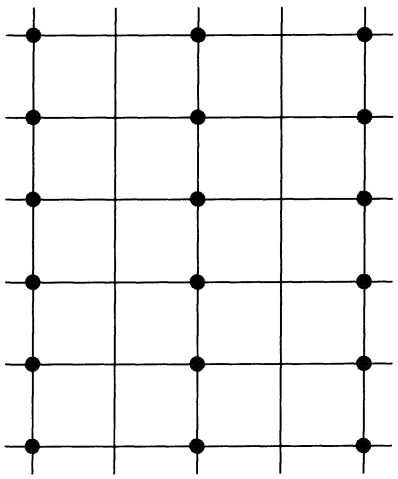


Figure 6
Column decimation.

- In the column case,

$$\overline{M_0(\omega)}\,\tilde{M}_0(\omega) + \overline{M_0(\omega+(\pi,0))}\tilde{M}_0(\omega+(\pi,0)) = 1 \tag{3.3}$$

and

$$\begin{aligned} M_1(\omega) &= e^{-i\omega_1}\,\overline{\tilde{M}_0(\omega+(\pi,0))}\,, \\ \tilde{M}_1(\omega) &= e^{-i\omega_1}\,\overline{M_0(\omega+(\pi,0))}\,. \end{aligned} \tag{3.4}$$

If the analysis and synthesis filters are equal, we find two generalizations of the CQF condition (2.5). The formulas (3.1) and (3.2) become

$$\begin{aligned} |M_0(\omega)|^2 + |M_0(\omega+(\pi,\pi))|^2 &= 1\,, \\ M_1(\omega) &= e^{-i(\omega_1+\omega_2)}\,\overline{M_0(\omega+(\pi,\pi))}\,; \end{aligned} \tag{3.5}$$

whereas (3.3) and (3.4) become

$$\begin{aligned} |M_0(\omega)|^2 + |M_0(\omega+(\pi,0))|^2 &= 1\,, \\ M_1(\omega) &= e^{-i\omega_1}\,\overline{M_0(\omega+(\pi,0))}\,. \end{aligned} \tag{3.6}$$

As in the one dimensional situation, we want to build from these schemes the associated scaling function which can be viewed as the limit of the cascade-reconstruction algorithm.

III.2. Non-separable scaling function and wavelets.

If c_{mn} are the Fourier coefficients of $M_0(\omega)$, *i.e.*

$$M_0(\omega) = M_0(\omega_1,\omega_2) = \sum_{m,n} c_{mn}\, e^{-i(m\omega_1+n\omega_2)}\,, \tag{3.7}$$

then the associated scaling function $\phi(x) = \phi(x_1, x_2)$ satisfies a two scale difference equation,

$$\phi(x) = 2\sum_{m,n} c_{mn}\,\phi(Dx-(m,n)) \tag{3.8}$$

and its Fourier transform can be expressed as an infinite product

$$\hat{\phi}(\omega) = \prod_{k=1}^{+\infty} M_0(D^{-k}\omega) \tag{3.9}$$

which is convergent if and only if $M_0(0) = 1$.

This scaling function has compact support if and only if $M_0(\omega)$ is an FIR filter. We see from (3.9) that ϕ will be highly dependent on the choice of D. For the same sublattice and the same filter, the results can be completely different for different D. The column sublattice for example is generated by both matrices $D_1 = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$ and $D_2 = \begin{pmatrix} 0 & 2 \\ 1 & 0 \end{pmatrix}$, but the first one cannot lead to an L^2 scaling function. Indeed, we would have

$$\hat{\phi}_1(0, 2n\pi) = \prod_{k=1}^{+\infty} M_0(D_1^{-k}(0, 2n\pi)) = 1 ,$$

for all $n \geq 0$. But since ϕ_1 is compactly supported and belongs to $L^2(\mathbb{R})$, it is also in $L^1(\mathbb{R})$ and its Fourier transform should tend to zero at infinity. We can also remark that only the eigenvalues of D_2 have their modulus strictly superior to 1.

The choice of the dilation matrix is thus very important. In fact, although the equations (3.1)-(3.2) are different from (3.3)-(3.4), the choice of the sublattice is less important: Indeed, for any dilation matrix D_1 such that $D_1\mathbb{Z}^2$ is the column sublattice, we can define

$$D_2 = P\, D_1 P^{-1} \quad \text{with } P = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} . \tag{3.10}$$

Clearly, the image of $\mathbb{Z}^2$ by D_2 is now the quincunx sublattice. Then, for any filter $M_0^1(\omega)$ satisfying the column-CQF condition (3.6), the corresponding scaling function ϕ_1 can be written in the following way,

$$\hat{\phi}_1(\omega) = \prod_{k=1}^{+\infty} M_0^1(D_k^{-k}\omega) = \prod_{k=1}^{+\infty} M_0^1(P^{-1}D_2^{-k}P\omega) = \hat{\phi}_2(P\omega)$$

where $\hat{\phi}_2$ is also a scaling function defined by

$$\begin{cases} \hat{\phi}_2(\omega) = \prod_{k=1}^{+\infty} M_0^2(D_2^{-k}\omega) , \\ M_0^2(\omega) = M_0^1(P^{-1}\omega) . \end{cases} \tag{3.11}$$

Since $P^{-1} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$, we have

$$|M_0^2(\omega)|^2 + |M_0^2(\omega + (\pi, \pi))|^2 = |M_0^1(\omega_1, \omega_1 + \omega_2)|^2 + |M_0^1(\omega_1 + \pi, \omega_1 + \omega_2 + 2\pi)|^2 = 1 .$$

And thus M_0^2 satisfies the quincunx-CQF condition (3.5). A similar result holds of course if we start from two dual filters M_0^1 and $\tilde{M}_0^1$ which satisfy (3.3). This shows that the scaling functions associated to D_1 and D_2 are linked by the simple relation $\phi_2(x) = \phi_1(Px)$. Consequently we can restrain our study to the quincunx case. More generally, if D_1 and D_2 satisfy

$$D_2 = PD_1P^{-1} \tag{3.12}$$

where P is a matrix having integer entries and determinant equal to 1, then we also have the same type of equivalence between the scaling functions. For this reason, we shall only consider the two simplest dilation matrices of determinant 2, which cannot be related as in (3.12) since they do not have the same eigenvalues:

$$R = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \left(\text{Rotation of } \frac{\pi}{4} \text{ and dilation of } \sqrt{2}\right) \tag{3.13}$$

and

$$S = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \left(\text{Symmetry with respect to } (\sqrt{2}+1, 1) \text{ and dilation of } \sqrt{2}\right) . \tag{3.13'}$$

In both of these cases the image of $\mathbb{Z}^2$ is the quincunx sublattice. The wavelet ψ is then defined by

$$\hat{\psi}(D\omega) = M_1(\omega)\hat{\phi}(\omega) \qquad \text{with} \quad D = R \text{ or } S , \tag{3.14}$$

where $M_1(\omega)$ is defined by (3.5) in the orthogonal case, and by (3.2) in the biorthogonal case where we also have a dual wavelet defined by

$$\hat{\tilde{\psi}}(D\omega) = \tilde{M}_1(\omega)\, \hat{\tilde{\phi}}(\omega) \qquad \text{with} \quad D = R \text{ or } S . \tag{3.15}$$

The goal is now to design filters leading to regular scaling functions and wavelets. We end this section by presenting two important families of

filters. The regularity of the associated ϕ, ψ, $\hat{\phi}$ and $\tilde{\psi}$ will be estimated in sections IV and V by different techniques which all are natural generalizations of the one dimensional tools that we introduced previously.

III.3. Filter design.

III.3.a. The orthonormal case.

Recall (see [Dau1]) that in $1D$, the CQF filter can be designed in the following way, in order to obtain wavelets with an arbitrarily high regularity:

1) For a given number N of vanishing moments, define m_0 by

$$|m_0(\omega)|^2 = \left[\cos^2\left(\frac{\omega}{2}\right)\right]^N P_N\left[\sin^2\left(\frac{\omega}{2}\right)\right] \tag{3.16}$$

where $P_N(y)$ is a polynomial, solution of the Bezout problem

$$y^N P_N(1-y) + (1-y)^N P_N(y) = 1 . \tag{3.17}$$

The minimal degree choice is given by

$$P_N(y) = \sum_{j=0}^{N-1} \binom{N-1+j}{j} y^j .$$

2) Find the function $m_0(\omega)$ by using the Riesz lemma which guarantees that there exist a trigonometric polynomial solving (3.16).

Unfortunately, this last result does not generalize to higher dimensions. We thus have to find other means to build trigonometric polynomials which satisfy (3.5). One possible method is the "polyphase component" construction used by Vaidyanathan [Va] and M. Vetterli [Ve], [VK]. It is based on the remark that $M_0(\omega)$ satisfies (3.5) if and only if the polyphase matrix

$$H_0(\omega) = \frac{1}{\sqrt{2}} \begin{pmatrix} M_0(\omega) + M_0(\omega + (\pi, \pi)) & M_1(\omega) + M_1(\omega + (\pi, \pi)) \\ M_0(\omega) - M_0(\omega + (\pi, \pi)) & M_1(\omega) - M_1(\omega + (\pi, \pi)) \end{pmatrix} \tag{3.18}$$

is unitary for all ω. Since the product of two polyphase matrices is also a polyphase matrix for a third pair of filter, infinite families can be constructed by multiplying elementary building blocks of the type (3.18) as soon as we know some simple filters which satisfy (3.5). The disadvantage of this method is that it does not furnish the vanishing moments in a natural way. Recall (see [Me1]) that the N times differentiability of the function ψ implies

$$|\hat{\psi}(\omega)| \leq C\left(|\omega_1|^{N+1} + |\omega_2|^{N+1}\right), \qquad (|\omega| \mapsto 0) \tag{3.19}$$

and thus $M_0(\omega)$ has necessarily a zero or order $N+1$ at the frequency $\omega = (\pi, \pi)$. This can also be viewed as the Fix-Strang condition (see [SF]) for the regularity of the scaling function ϕ.

The simplest way to build such filters with N arbitrarily high is to remark that if $m_0(\omega)$ is a $1D$ solution of the CQF equation (2.5), then the $2D$ filter defined by

$$M_0(\omega) = M_0(\omega_1, \omega_2) = m_0(\omega_1) \tag{3.20}$$

satisfies the equation (3.5). It is apparently a good candidate for building regular wavelets since it has the same order of cancellation in (π, π) as $m_0(\omega)$ in π. This allows us to build an infinite family of filters with an arbitrarily high number of vanishing moments by posing

$$M_0^N(\omega) = m_0^N(\omega_1) \tag{3.21}$$

where $\{m_0^N(\omega)\}_{N>0}$ is the family of filters designed in [Dau1], defined by (2.35), (2.37) and (2.38). Note that the filter (3.21) has a unidimensional structure but since the dilation D contains either a rotation or a symmetry, the final analysis (using iterates of the filter) is performed in all the directions of the plane. In Section IV, we shall take a closer look at the associated wavelets and their regularity. If $D = R$, then one can also derive another family of "almost" one-dimensional filters M_0 from unidimensional m_0 (they get again fanned out to other directions by applying R^{-1}). Explicitly,

$$\begin{aligned} M_0(\omega_1, \omega_2) = {} & \frac{1}{2}\left[m_0\left(\frac{\omega_1 - \omega_2}{2}\right) + m_0\left(\frac{\omega_1 - \omega_2}{2} + \pi\right)\right] \\ & + \frac{1}{2}\left[m_0\left(\frac{\omega_1 - \omega_2}{2}\right) - m_0\left(\frac{\omega_1 - \omega_2}{2} + \pi\right)\right] e^{i(\omega_1 + \omega_2)/2}. \end{aligned}$$

This construction corresponds to a filter with taps on two diagonals, $h_{n_1,n_2} = 0$ if $n_1 \neq -n_2$ and $n_1 \neq -n_2 + 1$. It is easy to check that this M_0 satisfies (3.5) if m_0 satisfies $|m_0(\omega)|^2 + |m_0(\omega + \pi)|^2 = 1$. If $m_0(0) = 1$, $m_0(\pi) = 0$, then $M_0(\pi, \pi) = 0$ follows, so that M_1, as defined in (3.5), satisfies $M_1(0,0) = 0$, as it should. One easily checks, however, that $\partial_{\omega_1} M_0(\pi, \pi)$ and $\partial_{\omega_2} M_0(\pi, \pi)$ cannot both be zero for these examples, so that the corresponding bases cannot possibly be C^1. Only the small examples are therefore of any interest; it seems possible (numerical experiment) to construct a continuous ϕ corresponding to a 4-tap filter in this way.

III.3.b. The biorthogonal case.

The filter design is clearly easier in the biorthogonal situation. One can start from a given filter $M_0(\omega)$ and find the dual $\tilde{M}_0(\omega)$ by solving linear equations.

In particular we can look for filters which have more isotropy than those of the family (3.21). Here, again, the one dimensional theory can help us to build families of filters in a simple way. Several examples of real and symmetrical dual filters have been designed by the authors and J. C. Feauveau in [CDF].

In these one dimensional construction the symmetry allows us to use the variable $y = \sin^2(\omega/2)$ and to write the transfer functions as

$$m_0(\omega) = p(y) \quad \text{and} \quad \tilde{m}_0(\omega) = \hat{p}(y) \tag{3.22}$$

where p and $\hat{p}$ are two polynomial satisfying

$$p(y)\hat{p}(y) + p(1-y)\hat{p}(1-y) = 1\ . \tag{3.23}$$

In two dimensions, consider the variables $y_1 = \sin^2(\omega_1/2)$ and $y_2 = \sin^2(\omega_2/2)$. If the filters are symmetrical with respect to the vertical and the horizontal axes, the duality condition in (3.3) can be rewritten as

$$P(y_1, y_2)\hat{P}(y_1, y_2) + P(1-y_1, 1-y_2)\hat{P}(1-y_1, 1-y_2) = 1\ , \tag{3.24}$$

where

$$P(y_1, y_2) = M_0(\omega_1, \omega_2)\,, \qquad \hat{P}(y_1, y_2) = \tilde{M}_0(\omega_1, \omega_2)\ .$$

We see that a possible choice for P and $\hat{P}$ is given by

$$P(y_1, y_2) = p(\alpha y_1 + (1-\alpha)y_2) \tag{3.25}$$

$$\hat{P}(y_1, y_2) = \hat{p}(\alpha y_1 + (1-\alpha)y_2) \tag{3.25'}$$

where α is in $[0,1]$. For an optimal isotropy it is natural to choose $\alpha = 1/2$; in this case the diagonals are also symmetry axes. This choice is known in signal processing as the McClellan transform of the $1D$ filters p and $\tilde{p}$. Using the variable $z = (y_1 + y_2)/2$ we can thus write

$$M_0(\omega) = p(z) \quad \text{and} \quad \tilde{M}_0(\omega) = \hat{p}(z) \tag{3.26}$$

where p and $\hat{p}$ are polynomials satisfying (3.24). These polynomials must also satisfy

$$p(0) = \hat{p}(0) = 1 \quad \text{and} \quad p(1) = \hat{p}(1) = 0 \tag{3.27}$$

which are necessary for the construction of wavelet bases. Note that we have

$$\begin{aligned} z &= \frac{1}{2}\left(\sin^2\left(\frac{\omega_1}{2}\right) + \sin^2\left(\frac{\omega_2}{2}\right)\right) \\ &= \frac{1}{8}\left(4 - e^{i\omega_1} - e^{i\omega_2} - e^{-i\omega_1} - e^{-i\omega_2}\right) \end{aligned} \tag{3.28}$$

and thus z can be regarded as the transfer function of the filter which computes the discrete Laplacian with the formula

$$\begin{aligned} (\Delta_d x)_{m,n} = \frac{1}{8}\,(& 4\,x_{m,n} - x_{m-1,n} - x_{m+1,n} \\ & - x_{m,n-1} - x_{m,n+1}\ )\,. \end{aligned} \tag{3.29}$$

Since a Laplacian scheme has frequently been proposed in image processing to detect the edges with a maximum isotropy (see [AB], [M]), it seems tempting to use z or one of its powers as a high pass analyzing filter (and thus $1 - z$ as the corresponding low pass synthesis filter). This can be achieved in a very simple way, by a method already used to build biorthogonal bases in $L^2(\mathbb{R})$. Recall that

$$P_N(z) = \sum_{j=0}^{N-1} \binom{N-1+j}{j} z^j$$

is the lowest degree solution of the Bezout problem

$$z^N P_N(1-z) + (1-z)^N P_N(z) = 1 \ . \tag{3.30}$$

If we fix the reconstruction low pass as $M_0^N(\omega) = (1-z)^N$ (so that the analyzing high pass is, up to a shift, the N-th power of the Laplacian), then a possible choice for the dual filter is given by

$$\tilde{M}_0^{N,L}(\omega) \ = \ (1-z)^L \ P_{N+L}(z) \tag{3.31}$$

where L is a positive integer indicating the cancellation order of $\tilde{M}_0$ at $\omega = (\pi, \pi)$. L has to be chosen large enough so that both functions $\varphi(x)$ and $\tilde{\varphi}(x)$ satisfy the necessary conditions to generate a pair $\{\psi_k^j, \tilde{\psi}_k^j\}_{j\in\mathbb{Z}, k\in\mathbb{Z}^2}$ of unconditional Riesz bases (see Theorem 2.1 and Appendix A). We shall examine the properties of these functions and give an estimate of the minimal value of L in Section V.

We have now at hand two families of filters, orthonormal and biorthogonal, with an arbitrarily high number of vanishing moments. We still have to know if these filters allow us to build wavelet bases with an arbitrarily high regularity as in the one dimensional case ([Dau1], [Co2]). As we shall see in the next two sections, the results of our investigations are very surprising and show that the multidimensional situation contains a lot of new difficulties from this point of view.

IV. Orthonormal bases of non-separable wavelets.

Let us consider the family of CQF filters defined by

$$M_0^N(\omega_1, \omega_2) \ = \ m_0^N(\omega_1) \tag{4.1}$$

with

$$|m_0^N(\omega)|^2 = \left[\cos^2\left(\frac{\omega}{2}\right)\right]^N \sum_{j=0}^{N-1} \binom{N-1+j}{j} \left[\sin^2\left(\frac{\omega}{2}\right)\right]^j \tag{4.2}$$

and the associated scaling functions for the dilations S and R,

$$\hat{\phi}_{N,S}(\omega) = \prod_{k=1}^{\infty} M_0^N(S^{-k}\omega), \tag{4.3}$$

$$\hat{\phi}_{N,R}(\omega) = \prod_{k=1}^{\infty} M_0^N(R^{-k}\omega)\,. \tag{4.4}$$

IV.1. Orthonormality of the translates.

A first requirement is that the $\mathbb{Z}^2$-translates of $\phi_{N,S}$ or $\phi_{N,R}$ are orthonormal. This is a necessary and sufficient condition to generate multiresolution analyses and orthonormal bases of wavelets.

Theorem 4.1. *For all $N > 0$, the functions $\phi_{N,D}$ have orthonormal translates and generate wavelet bases of the type*

$$2^{-j/2}\psi(D^{-j}x - k)\,, \qquad j \in \mathbb{Z},\ k \in \mathbb{Z}^2\,,$$

where $D = S$ or R.

PROOF. By a trivial generalization of Theorem 2.1, this orthonormality is ensured if and only if $|\hat{\phi}(\omega)| \geq C > 0$ on a compact set K congruent to $[-\pi, \pi]^2$ modulo $2\pi\mathbb{Z}^2$ which contains a neighbourhood of the origin.

It is clear that $M_0^N(\omega)$ vanishes only on the vertical lines $\omega_1 = (2k+1)\pi$, $k \in \mathbb{Z}$. Consequently we see that the simple choice $K = [-\pi, \pi]^2$ is not convenient since for both dilations, we have

$$D^{-1}(\pi, \pi) = (\pi, 0) \tag{4.5}$$

and thus

$$\hat{\phi}(\pi, \pi) = 0\,. \tag{4.6}$$

Recall that in the one dimensional case, the trivial choice $K = [-\pi, \pi]$ was convenient for the family $m_0^N(\omega)$. Here we have to use a compact set K slightly different from $[-\pi, \pi]^n$ so that $D^{-j}K \cap \{\omega_1 = (2k+1)\pi\}$ is empty for all $j > 0$ and for all k in $\mathbb{Z}$. This can be done very easily by removing small neighbourhoods of (π, π) and $(-\pi, -\pi)$ and translating them by $(-2\pi, 0)$ and $(2\pi, 0)$ as shown in figure 7.

One checks easily that all the sets $D^{-j}K$ for $j > 0$ are contained in the strip $|\omega_1| \leq \pi - \varepsilon$, $\varepsilon > 0$ where $M_0^N(\omega)$ does not vanish.

We now have to check the regularity of the scaling functions which have been obtained. We shall see that the results are completely different depending on whether one chooses S or R as the dilation matrix.

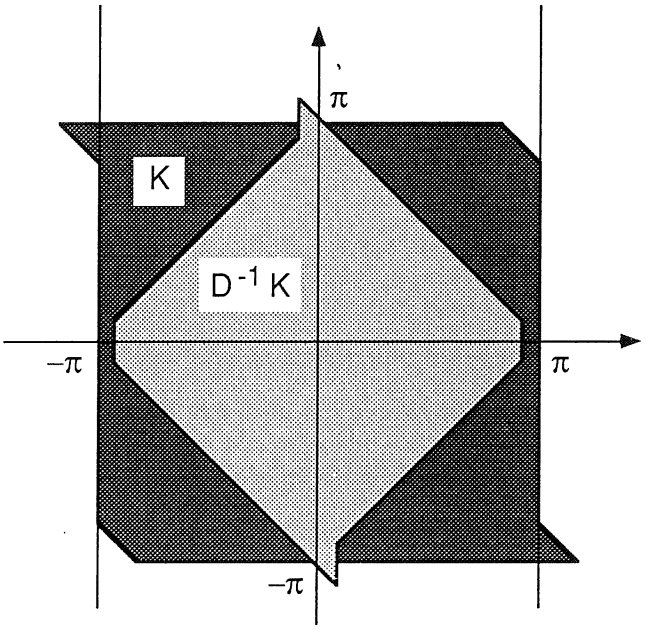


Figure 7
The convenient compact set K congruent to $[-\pi, \pi]^2$: Neighbourhoods of (π, π) and $(-\pi, -\pi)$ have been shifted so that $\hat{\varphi}$ does not vanish on K.

IV.2. The symmetry dilation case.

In this case the dilation matrix is $S = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ and its inverse is $S^{-1} = \frac{1}{2}S$. Since $M_0^N(\omega) = m_0^N(\omega_1)$, we have to consider the sequence $\{[S^{-j}\omega]_1\}_{j>0}$ for a given $\omega = (\omega_1, \omega_2)$. Clearly, it has the following form:

$$\frac{1}{2}(\omega_1 + \omega_2),\ \frac{1}{2}\omega_1,\ \frac{1}{4}(\omega_1 + \omega_2),\ \frac{1}{4}\omega_1, \ldots, 2^{-j}(\omega_1 + \omega_2),\ 2^{-j}\omega_1, \ldots$$

Since $S^{-2} = \frac{1}{2}I$, the odd and the even parts are simple dyadic sequences

and this leads to

$$\hat{\phi}_{N,S}(\omega) = \hat{\varphi}_N(\omega_1 + \omega_2)\ \hat{\varphi}_N(\omega_1) \tag{4.7}$$

or

$$\phi_{N,S}(x) = \varphi_N(x_2)\ \varphi_N(x_1 - x_2) \tag{4.8}$$

where φ_N is the one dimensional scaling function. The associated wavelet is defined by

$$\hat{\psi}_{N,S}(\omega) = M_1(\omega)\hat{\phi}_{N,S}(\omega) = \hat{\psi}_N(\omega_1 + \omega_2)\ \hat{\varphi}_N(\omega_1) \tag{4.9}$$

or

$$\psi_{N,S}(x) = \psi_N(x_2)\ \varphi_N(x_1 - x_2)\ . \tag{4.10}$$

We see here that the scaling function and wavelet are in this case separable in the sense that they can be expressed directly in terms of the one dimensional functions φ_N and ψ_N. This separability can be explained by the fact that S is similar to the matrix $\begin{pmatrix} 0 & 1 \\ 2 & 0 \end{pmatrix}$, which is simply a dilation by a factor 2 (in one) direction, followed by an exchange of the axes. The regularity can of course be made arbitrarily high since it is directly given by the Hölder exponent of φ_N.

REMARK. Theorem 4.1 is not necessary here to prove the orthonormality of the translates since it is a trivial consequence of the separability formulas (4.7) and (4.8).

We now consider the case of the matrix R which is by far less trivial.

IV.3. The rotation dilation case.

We now have $R = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$ and $R^{-1} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$. The sequence $\left\{[R^{-j}\omega]_1\right\}_{j>0}$ is then,

$$\frac{1}{2}(\omega_1 + \omega_2),\ \frac{1}{2}\omega_2,\ \frac{1}{4}(\omega_2 - \omega_1),\ -\frac{1}{4}\omega_1,\ -\frac{1}{8}(\omega_1 + \omega_2),$$
$$-\frac{1}{8}\omega_2,\ \frac{1}{16}(\omega_1 - \omega_2),\ \frac{1}{16}(\omega_1),\ \frac{1}{32}(\omega_1 + \omega_2),\ \frac{1}{32}\omega_2\ ,\ldots$$

Here the first power of R^{-1} proportional to the identity is $R^{-4} = -\frac{1}{4}I$. Consequently, it is not possible to use the one dimensional scaling functions and wavelets to express the ϕ_N and ψ_N in a separable way. We first consider the case $N = 1$ which corresponds to the Haar filter. The result of the cascade algorithm with this filter shows how different the situation is when R is used instead of S.

IV.3.a. The twin dragon.

For $M_0^1(\omega) = (1 + e^{-i\omega_1})/2$, the function $\phi_{1,R}$ satisfies

$$\phi_{1,R}(x) = \phi_{1,R}(Rx) + \phi_{1,R}(Rx - (1,0)) \tag{4.11}$$

and

$$\hat{\phi}_{1,R} = \prod_{k=1}^{\infty} M_0^1(R^{-k}\omega)\,. \tag{4.12}$$

By iteration of the cascade algorithm, one finds that ϕ is the characteristic function of a well known fractal set called the "twin dragon" (see [K]) shown in figure 8. This set can be defined directly in the complex plane as

$$\Delta = \left\{ \sum_{n=1}^{\infty} \varepsilon_n \left(\frac{1-i}{2}\right)^n : \quad \{\varepsilon_n\}_{n\in\mathbb{N}} \in \{0,1\}^{\mathbb{N}} \right\} \tag{4.13}$$

and it is clear that $\phi_{1,R} = \chi_\Delta$ solves (3.41) since we have

$$\begin{aligned} \Delta &= \left(\frac{1-i}{2}\right)\Delta \cup \left(\frac{1-i}{2}\right)(\Delta + 1) \\ &\sim R^{-1}\Delta \cup R^{-1}(\Delta + (0,1))\,. \end{aligned} \tag{4.14}$$

The self-similarity of Δ is thus expressed by the two scale difference equation (4.11), but furthermore, since the family $\{\phi_{1,R}(x-k)\}_{k\in\mathbb{Z}^2}$ is orthonormal (by Theorem 4.1) and since $|\Delta| = \hat{\phi}_{1,R}(0) = 1$, these integer translates constitute a fractal tiling of the whole plane $\mathbb{R}^2$ (similarly to the squares obtained in the tensor product situation with the same filter). This beautiful property has been observed independently by W. Madych and K. Gröchenig [MG] and W. Lawton and H. Resnikoff

[LR]. More generally, such tilings can be derived by considering a two scale difference equation of the type

$$\phi(x) = \sum_{i=1}^{d} \phi(Dx + e_i) \tag{4.15}$$

where D is a dilation matrix and $\{e_i\}_{i=1,\ldots,d}$ are d representatives of $\mathbb{Z}^n/D\mathbb{Z}^n$ ($d = |\det D|$). This scaling function and the corresponding wavelet do not seem however of great interest for image processing: not only are they discontinuous but the set of discontinuity is a very chaotic fractal curve. Nevertheless the twin dragon is important in estimating the regularity (local and global) of the wavelets with dilation matrix R.

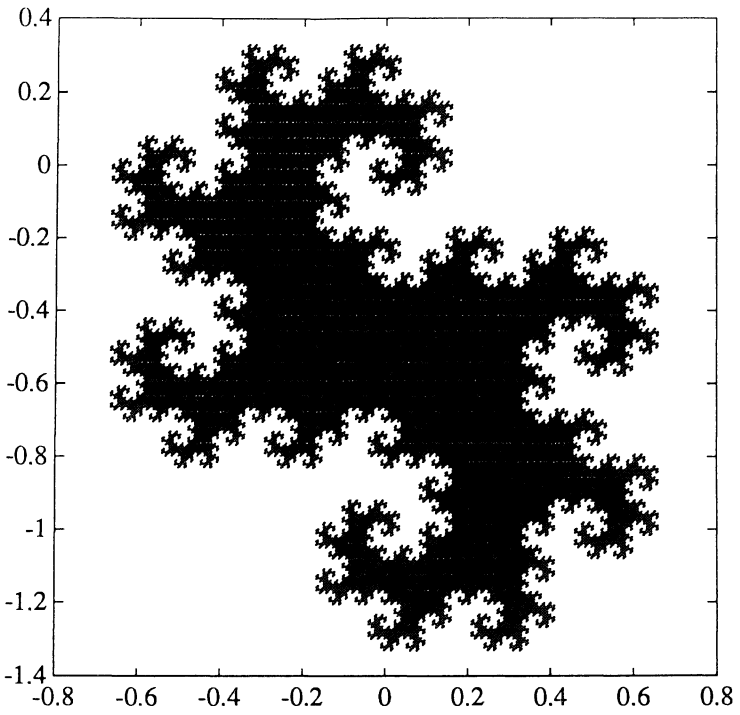


Figure 8
The "twin dragon" set Δ.

Indeed, if we want to generalize the method of [DL] (see Section II.4.6), it is necessary to consider the expansion of any point in $\mathbb{C}$ in terms of the power of $(1+i)/2$ ($\sim R^{-1}$), which also means that the point is considered as the limit of a "dragonic sequence" $\{\Delta_j\}_{j\in\mathbb{Z}}$ with $\Delta_j \subset \Delta_{j-1}$ and $|\Delta_j| = 2^{-j}$. These "dragonic expansion" techniques are described in Appendix B.

Let us now examine the functions obtained with higher order filters which have more vanishing moments.

IV.3.b. Higher order filters.

We are interested in the family of scaling function $\phi_{N,R}$, $N > 1$.

Recall that in the one dimensional case, the asymptotic result ensuring arbitrarily high regularity (Theorem 2.4, Section II.3.6) is based on the value of $|m_0(\pm 2\pi/3)|$ since $\{-2\pi/3, 2\pi/3\}$ is a cyclic orbit of $\omega \mapsto 2\omega$ modulo 2π. In the present case similar considerations for a fixed orbit of $\omega \mapsto R\omega$ modulo $2\pi\mathbb{Z}^2$, lead to an opposite result: arbitrarily high regularity cannot be obtained by increasing the number of vanishing moments. More precisely, we have

Theorem 4.2. *For all $N > 0$, the function $\phi_{N,R}$ is not in $C^1(\mathbb{R}^2)$.*

PROOF. This is of course true for $N = 1$ since we obtain the twin dragon. For $N > 1$, we shall prove a stronger result: the decay at infinity of $\hat{\phi}_{N,R}(\omega)$ cannot be majorated by $C\,|\omega|^{-1}$ (which is a necessary condition for $\phi_{N,R}$ to be in C^1 because it is a compactly supported function). For this we consider the orbit of $\omega \mapsto R\omega$ modulo $2\pi\mathbb{Z}^2$ given by the four points $(2\pi/5, 4\pi/5)$, $(2\pi/5, -4\pi/5)$, $(-2\pi/5, -4\pi/5)$ and $(-2\pi/5, 4\pi/5)$. Let us denote $v_0 = (2\pi/5, 4\pi/5)$ and $v_j = R^j v_0$. One checks easily that

$$|\hat{\phi}_{N,R}(v_0)| = C_N \neq 0 \quad \text{for all } N > 0\,. \tag{4.16}$$

We then have, for all $N > 0$,

$$|\hat{\phi}_{N,R}(v_j)| = C_N \left| m_0^N\left(\frac{2\pi}{5}\right)\right|^j\,. \tag{4.17}$$

From the definition of m_0^N we have

$$\left| m_0^N\left(\frac{2\pi}{5}\right)\right|^2 = \left[\cos^2\left(\frac{\pi}{5}\right)\right]^N P_N\left(\sin^2\left(\frac{\pi}{5}\right)\right) \tag{4.18}$$

and we know from (2.51) that

$$P_N(y) \leq (4y)^{N-1}, \quad \text{if } \frac{1}{2} \leq y \leq 1 \,. \tag{4.19}$$

Because $\cos^2(\pi/5) > \cos^2(\pi/4) = 1/2$, we can write

$$\begin{aligned}\left|m_0^N\left(\frac{2\pi}{5}\right)\right|^2 &= 1 - \left[\sin^2\left(\frac{\pi}{5}\right)\right]^N P_N\left[\cos^2\left(\frac{\pi}{5}\right)\right] \\ &\geq 1 - \left[\sin^2\left(\frac{\pi}{5}\right)\right]^N \left[4\cos^2\left(\frac{\pi}{5}\right)\right]^{N-1} \\ &= 1 - \sin^2\left(\frac{\pi}{5}\right)\left[\sin^2\left(\frac{2\pi}{5}\right)\right]^{N-1}\end{aligned}$$

and thus, since $|v_j| \geq 2^{j/2}$,

$$\begin{aligned}|\hat{\phi}_{N,R}(v_j)| &\geq C_N\left[1 - \sin^2\left(\frac{\pi}{5}\right)\left[\sin^2\left(\frac{2\pi}{5}\right)\right]^{N-1}\right]^{j/2} \\ &\geq C_N|v_j|^{-\alpha_N}\end{aligned}$$

with $\alpha_N = \left|\log\left(1 - \sin^2(\pi/5)\left(\sin^2(2\pi/5)\right)^{N-1}\right)\right| / \log 2$. Clearly α_N is decreasing with N. Since $\alpha_1 \simeq 0.6115 < 1$, this ends the proof.

In fact, these wavelets do not even seem continuous although we have no mathematical proof for this. A simple look at the result of the cascade algorithm for the 4 tap filter (which corresponds to a .55 Hölder continuous one dimensional wavelet) shows how chaotic the functions $\phi_{R,N}$ can be (figure 9). The design of FIR filters leading to regular wavelet bases with R as the dilation matrix seems to be a difficult problem. Using a polyphase component approach M. Vetterli and J. Kovacevic ([KV], p. 32) have constructed a filter for which the result of the cascade looks continuous but no infinite family with arbitrarily high regularity has been designed so far.

The main difficulty which makes this design unpracticable is the absence of the Riesz lemma in more than one dimension and thus the impossibility to start by designing the square modulus of $M_0(\omega)$ in an appropriate way. Apart from this problem, the CQF filters (in particular the family (3.21) that we have introduced) cannot be symmetrical. We must keep in mind that one of the interests of the quincunx grid

decimation is to have a more isotropic analysis; this is only achieved if the filter coefficients are themselves symmetrical around the horizontal, vertical and diagonal directions.

These two reasons encourage us to construct biorthogonal bases of wavelets from dual filters for which the Riesz lemma is not necessary and linear phase can be achieved.

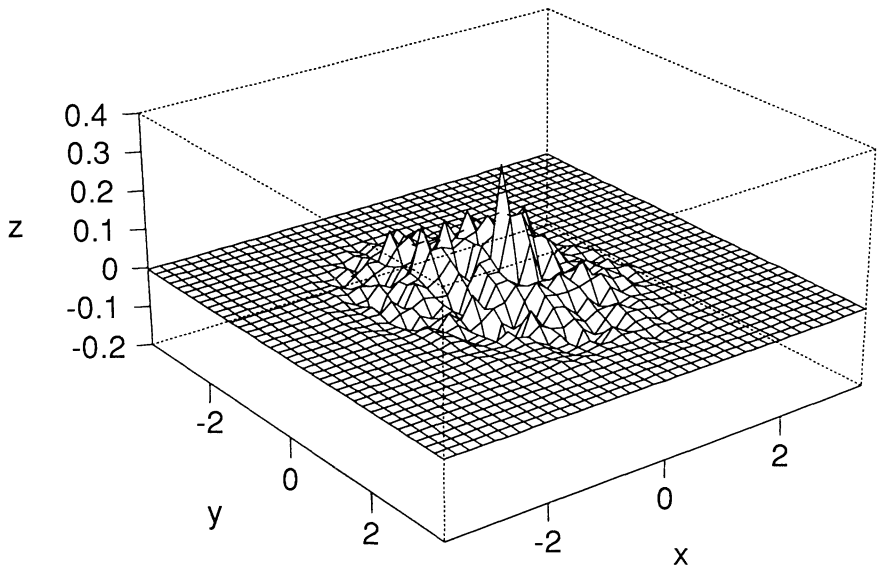


Figure 9
Approximation of the scaling function ϕ_{2R} .

V. Biorthogonal bases of nonseparable wavelets.

Let us recall the family of dual filters introduced in III.3.b. It is based on the variable

$$z = \frac{1}{2}\left(\sin^2\left(\frac{\omega_1}{2}\right) + \sin^2\left(\frac{\omega_2}{2}\right)\right).$$

We have chosen

$$M_0^N(\omega) = (1-z)^N, \tag{5.1}$$

and

$$\tilde{M}_0^{N,L}(\omega) = (1-z)^L \, P_{N+L}(z), \tag{5.2}$$

where L is still to be fixed.

A first remark is that the action of the dilation matrices R and S on the variable z are equivalent. This is due to the fact that z is invariant if we exchange ω_1 and ω_2 or if we change the sign of one of these variables. We shall thus consider a dilation matrix D which can be equal to R or S. To express its action on z we still need the two variables

$$y_1 = \sin^2\left(\frac{\omega_1}{2}\right) \quad \text{and} \quad y_2 = \sin^2\left(\frac{\omega_2}{2}\right). \tag{5.3}$$

We then have

$$\begin{aligned} z = \frac{1}{2}(y_1 + y_2) &\xmapsto{D} z = \frac{1}{2}(y_1 + y_2 - 2\,y_1 y_2) \\ &\xmapsto{D} z = \frac{1}{2}(4\,y_1(1-y_1) + 4\,y_2(1-y_2)) = \frac{1}{2}(y_1' + y_2') \\ &\xmapsto{D} z = \frac{1}{2}(y_1' + y_2' - 2\,y_1' y_2') \ldots \end{aligned}$$

We shall start by studying the scaling function ϕ_1 associated to the filter $M_0^1(\omega) = 1 - z$, because it is the elementary building block for the family ϕ_N $(= (*)^N \phi_1)$.

V.1. The quincunx Laplacian scheme.

The coefficients of $M_0^1(\omega)$ are centered around the origin and have the following form:

$$\frac{1}{8}\begin{pmatrix} & 1 & \\ 1 & 4 & 1 \\ & 1 & \end{pmatrix}. \tag{5.4}$$

Note that this is the simplest symmetrical filter (with respect to the horizontal, vertical and diagonal directions) which satisfies the cancellation condition $M_0^1(\pi, \pi) = 0$. To estimate the decay of $\hat{\phi}_1(\omega)$ we could hope for a bidimensional formula equivalent to

$$\prod_{k=1}^{+\infty} \cos(2^{-k}\omega) = \frac{\sin\omega}{\omega}, \tag{5.5}$$

used in the one dimensional case. Note that (5.5) is based on the iteration of $\sin\omega = 2\sin(\omega/2)\cos(\omega/2)$. Unfortunately, similar relations do not exist in the bidimensional case for the dilation matrix D. In particular the infinite product

$$\hat{\phi}_1(\omega) = \prod_{k=1}^{+\infty} M_0(D^{-j}\omega) \tag{5.6}$$

has no simple expression and one checks easily that, unlike (5.5), it does not have uniform decay at infinity. Indeed, let us consider the sets $\{(2\pi/5, 4\pi/5)\}$ and $\{(2\pi/3, 2\pi/3), (2\pi/3, 0)\}$. These are two cyclic orbits of $\omega \mapsto D\omega$ modulo $2\pi\mathbb{Z}^2$ and modulo the exchange of coordinates and sign changes which do not affect the variable z. Consequently, if we define $v_j = D^j(2\pi/5, 4\pi/5)$ and $\mu_j = D^j(2\pi/3, 2\pi/3)$, we have, when j goes to $+\infty$,

$$\hat{\phi}_1(v_j) \sim C\left[\frac{\cos^2(\pi/5) + \cos^2(2\pi/5)}{2}\right]^j \sim C\,|v_j|^{-\alpha_v} \tag{5.7}$$

and

$$\hat{\phi}_1(\mu_j) \sim C\left[\left(\frac{\cos^2(\pi/3)+1}{2}\right)\cos^2\left(\frac{\pi}{3}\right)\right]^{j/2} \sim C\,|\mu_j|^{-\alpha_\mu} \tag{5.8}$$

with

$$\alpha_v = -\frac{2}{\log 2} \log\left[\frac{\cos^2(\pi/5) + \cos^2(2\pi/5)}{2}\right] \simeq 2.83 \tag{5.9}$$

and

$$\alpha_\mu = -\frac{1}{\log 2} \log\left[\left(\frac{\cos^2(\pi/3)+1}{2}\right)\cos^2\left(\frac{\pi}{3}\right)\right] \simeq 2.68 \neq \alpha_v \,. \tag{5.10}$$

Still we would like to find a global exponent for the decay of $\hat{\phi}_1(\omega)$ at infinity. For this we shall introduce an "artificial" function which will play the same role as $\cos\omega$ in (5.5). We define

$$C(\omega) = \frac{\sin^2\left(\frac{\omega_1+\omega_2}{2}\right) + \sin^2\left(\frac{\omega_1-\omega_2}{2}\right)}{2\left[\sin^2\left(\frac{\omega_1}{2}\right) + \sin^2\left(\frac{\omega_2}{2}\right)\right]} \,, \quad C(0) = 1 \,. \tag{5.11}$$

Contrarily to $M_0^1(\omega)$, $C(\omega)$ is not a trigonometric polynomial, but it is a bounded regular function which vanishes at the point (π, π) with the same order of cancellation as $M_0^1(\omega)$. Moreover, it satisfies by construction

$$\prod_{k=1}^{+\infty} C(D^{-j}\omega) = \frac{2\left[\sin^2(\omega_1/2) + \sin^2(\omega_2/2)\right]}{\omega_1^2 + \omega_2^2} \leq C(1+|\omega|)^{-2} \,. \tag{5.12}$$

The decay of this infinite product is now uniform and, for this reason, $C(\omega)$ will play an important role in the construction of our dual bases. For the moment, by comparing $C(\omega)$ and $M_0^1(\omega)$, we obtain the following result:

Proposition 5.1. *The decay of $\hat{\phi}_1(\omega)$ at infinity is controlled by*

$$|\hat{\phi}_1(\omega)| \leq C(1+|\omega|)^{-2} \,. \tag{5.13}$$

Furthermore, this exponent is globally optimal, i.e. there exists a sequence $\{\omega_j\}_{j>0}$ such that $\lim_{j\to+\infty}|\omega_j| = +\infty$ and $|\hat{\phi}_1(\omega_j)| \sim C|\omega_j|^{-2}$.

PROOF. Using the variables $y_1 = \sin^2(\omega_1/2)$ and $y_2 = \sin^2(\omega_2/2)$ we can rewrite $C(\omega)$ as

$$C(\omega) = \frac{y_1 + y_2 - 2y_1y_2}{y_1+y_2} = \frac{(1-y_1)y_2 + (1-y_2)y_1}{y_1+y_2} \,. \tag{5.14}$$

We thus have

$$\begin{aligned} C(\omega) - M_0^1(\omega) &= \frac{(1-y_1)y_2 + (1-y_2)y_1}{y_1+y_2} - \frac{(1-y_1)+(1-y_2)}{2} \\ &= \frac{(1-y_1)(y_2-y_1)+(1-y_2)(y_1-y_2)}{2(y_1+y_2)} \\ &= \frac{(y_1-y_2)^2}{2(y_1+y_2)} \geq 0 \ . \end{aligned}$$

Thus $M_0^1(\omega) \leq C(\omega)$ and by (5.12) $|\hat{\phi}_1(\omega)| \leq C\,(1+|\omega|)^{-2}$. To prove that this exponent is optimal we consider a small vector $\rho \neq 0$ in $\mathbb{R}^2$ and define

$$\omega_j \ = \ D^j(\pi, \pi) + \rho \ , \tag{5.15}$$

so that

$$\hat{\phi}_1(\omega_j) \ = \ \prod_{k=1}^{+\infty} M_0^1 \left(D^{j-k}(\pi, \pi) + D^{-k}\rho\right) \ . \tag{5.16}$$

Let us divide this product in three parts

$$\begin{aligned} \hat{\phi}_1(\omega_j) = &\left[\hat{\phi}_1\left((\pi, \pi) + D^{-j}\rho\right)\right] \\ &\cdot \left[\prod_{k=1}^{j-1} M_0^1 \left(D^{j-k}(\pi, \pi) + D^{-k}\rho\right)\right] \\ &\cdot \left[M_0^1\left((\pi, \pi) + D^{-j}\rho\right)\right] \\ = \ &A(j)\ B(j)\ C(j) \ . \end{aligned} \tag{5.17}$$

One checks easily that $\hat{\phi}_1(\pi, \pi) \neq 0$ and thus, for j large enough or sufficiently small ρ, we have $0 < C_1 \leq A(j) \leq 1$. It is also clear that for $1 \leq k \leq j-1$, $M_0^1\left(D^{j-k}(\pi, \pi)\right) = 1$ and that for $\ell \geq 1$, $M_0^1(D^\ell(\pi, \pi) + \sigma) \geq 1 - C\,\|\sigma\|$ for σ small enough, with $C > 0$. Consequently, if ρ has been chosen small enough, $1 \geq B(j) \geq \prod_{\ell=1}^{\infty} [1 - C\,2^{-\ell}\,\|\rho\|] \geq C_2 > 0$. Finally since (π, π) is a second order zero of $M_0(\omega)$, the third factor satisfies

$$\begin{aligned} 2^{-j} C_3\,\|\rho\|^2 \ = \ C_3\,\|D^{-j}\rho\|^2 \ &\leq \ C(j) \\ &\leq \ C_4\,\|D^{-j}\rho\|^2 \ = \ 2^{-j} C_4\,\|\rho\|^2 \ . \end{aligned} \tag{5.18}$$

This shows that $\hat{\phi}_1(\omega_j)$ behaves like $2^{-j} \sim |\omega_j|^{-2}$ when j goes to $+\infty$ and the proposition is proved.

Note that from the decay of $\hat{\phi}_1(\omega)$ we cannot even conclude that it belongs to $L^1(\mathbb{R}^2)$ or that $\phi_1(x)$ is a continuous function. Yet both are true; we are going to prove this by the Littlewood-Paley method explained in II.4.a. The filter $M_0^1(\omega)$ and the scaling function $\hat{\phi}_1(\omega)$ are particularly well adapted for this approach since they are positive so that the regularity estimation is optimal (because $\|\Delta_j(\phi_1)\|_{L^\infty} \sim \|\widehat{\Delta_j}(\phi_1)\|_{L^1}$; see Section II.4.a).

Proposition 5.2. *The optimal global Hölder exponent for $\phi_1(x)$ is*

$$\alpha = \frac{2}{\log 2} \log\left(\frac{1+\sqrt{5}}{4}\right) \simeq .61$$

PROOF. We consider the transition operator defined by

$$TF(D\omega) = M_0^1(\omega)F(\omega) + M_0^1(\omega + (\pi, \pi))F(\omega + (\pi, \pi)). \tag{5.19}$$

As in the one dimensional case T can be studied in a finite dimensional space but this subspace cannot be defined as simply as $E(N_1, N_2)$ in (2.61). One way of finding an invariant subspace is to apply T to the constant 1 and then iterate it on the characters $e^{i(k_1\omega_1 + k_2\omega_2)}$ which are obtained until a stable set is attained. With M_0^1 corresponding to (5.4), this subspace is trivial, since $T_1 = 1$. Lemma 2.5 then guarantees the integrability of $\hat{\phi}_1$, hence the continuity of ϕ_1. To estimate the Hölder exponent of ϕ_1 we need a larger subspace, which we obtain by iterating T on 1 and on $\cos\omega_1 + \cos\omega_2$. The size of the matrix representing the action of T on this subspace can be seriously reduced by exploiting the symmetries, *i.e.* the invariance under $\omega_1 \longleftrightarrow -\omega_1$, $\omega_2 \longleftrightarrow -\omega_2$ and $\omega_1 \longleftrightarrow \omega_2$.

Using the subspace E generated by the basis

$$e_1 = 1, \; e_2 = \cos\omega_1 + \cos\omega_2, \; e_3 = \cos(\omega_1 + \omega_2) + \cos(\omega_2 - \omega_1) \tag{5.20}$$

we obtain the following matrix

$$T = \begin{pmatrix} 1 & 1/2 & 0 \\ 0 & 1/2 & 1 \\ 0 & 1/4 & 0 \end{pmatrix} \tag{5.21}$$

which has the eigenvalues $\{1, (1+\sqrt{5})/4, (1-\sqrt{5})/4\}$. The two last eigenvalues correspond to the subspace $E_0 \subset E$ defined by

$$E_0 = \{F(\omega) \in E : F(0) = 0\} . \tag{5.22}$$

Similarly to the one dimensional case, we iterate T on the positive function $e_1 - \frac{1}{2}e_2$ which is clearly in E_0 and this leads us to

$$\|\Delta_{j/2}(\phi_1)\|_{L^\infty} \sim \left\|\hat{\Delta}_{j/2}(\phi_1)\right\|_{L^1} \sim C\left(\frac{1+\sqrt{5}}{4}\right)^j , \tag{5.23}$$

where $\Delta_{j/2}(\phi_1)$ is the Littlewood-Paley block corresponding to the region $D^j([-\pi, \pi]^2)/D^{j-1}([-\pi, \pi]^2)$, situated at a distance $2^{j/2}$ of the origin. Consequently, if we define

$$\alpha = -\frac{2}{\log 2} \log\left(\frac{1+\sqrt{5}}{4}\right) \simeq 0.61 , \tag{5.24}$$

then it follows from (5.23) that

$$(1+|\omega|)^\alpha \hat{\phi}_1(\omega) \in L^1(\mathbb{R}^2) \quad \text{and} \quad \phi_1(x) \in C^\alpha(\mathbb{R}^2) . \tag{5.25}$$

Consequently ϕ_1 is Hölder continuous with regularity 0.61.

This property appears in the graph of ϕ_1 on figure 10 (obtained by the cascade algorithm) which presents a smooth aspect with several pointwise cusps. Note that this regularity is not sufficient to derive a better decay of $\hat{\phi}_1(\omega)$ than $|\omega|^{-0.61}$; Propositions 5.1 and 5.2 are thus complementary.

REMARKS.

- Note that, since we have

$$M_0^1(\omega) + M_0^1(\omega + (\pi, \pi)) = 1 , \tag{5.26}$$

we can derive the L^1 convergence of the truncated products $\hat{\phi}_{1n} = \prod_{j=1} M_0(D^{-j}\omega)\chi_{D^n([-\pi,\pi]^2)}(\omega)$ with the same method as in the

orthonormal case for the L^2 convergence (Theorem 2.1). This leads us to a Poisson summation formula

$$\sum_{k\in\mathbb{Z}^2} \hat{\phi}_1(\omega + 2k\pi) = 1 \tag{5.27}$$

which is equivalent to

$$\phi_1(n_1, n_2) = 1 \text{ if } n_1 = n_2 = 0,\ 0 \text{ if } (n_1, n_2) \in \mathbb{Z}^2/\{0\}\ . \tag{5.28}$$

This interpolating property of ϕ_1 has been noticed in approximation theory by Deslaurier and Dubuc [DD]. It explains the four cusps surrounding the center at the points $(0,1)$, $(1,0)$, $(0,-1)$ and $(-1,0)$ which are visible on figure 10. However, a sharper analysis shows that the isolated points where $\phi_1(x) = 0$ are an infinite family.

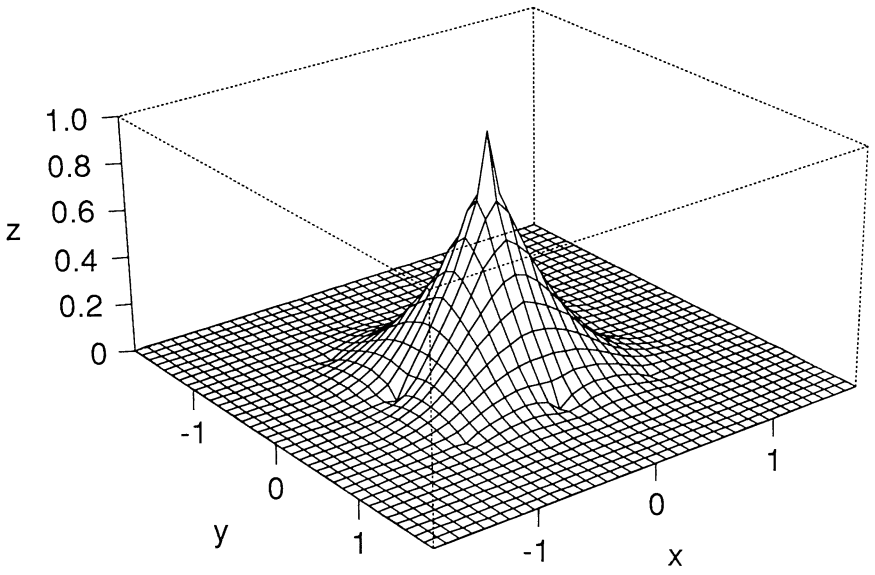


Figure 10
The scaling function $\phi_1(x)$.

- As mentioned in Section III.3.b, the variable $z = (y_1 + y_2)/2$ can be replaced by, more generally, $z_\lambda = \lambda y_1 + (1-\lambda)y_2$ with $\lambda \in [0,1]$; $M_0^1(\omega) = 1 - z_\lambda$ is still positive. Let us now distinguish the dilation matrices R and S. Then, a similar analysis in the case of $D = R$ leads to a 5×5 matrix in the basis

$$(e_1, e_2, e_3, e_4, e_5) = (1, \cos\omega_1, \cos\omega_2, \cos(\omega_1 + \omega_2), \cos(\omega_1 - \omega_2))$$

$$T_\lambda = \frac{1}{2}\begin{pmatrix} 2 & \lambda & 1-\lambda & 0 & 0 \\ 0 & 1-\lambda & \lambda & 0 & 2 \\ 0 & 1-\lambda & \lambda & 2 & 0 \\ 0 & \lambda & 0 & 0 & 0 \\ 0 & 0 & 1-\lambda & 0 & 0 \end{pmatrix} \tag{5.29}$$

and numerical computations show that the "isotropic value" $\lambda = 1/2$ gives the highest index of regularity. The lowest index of regularity is attained for $\lambda = 0$ or 1. Note that $\lambda = 1$ corresponds to the convolution product $g(x) = \chi_\Delta * \chi_\Delta$ where Δ is the twin dragon introduced in IV.3.a. The Hölder exponent is then $\alpha \simeq 0.47$.

- To estimate the decay of $\hat{g}(\omega)$ $(= (\hat{\chi}_\Delta(\omega))^2)$, one can again use the function $C(\omega)$ of Proposition 5.1, in a slightly different way. Remark that, if we define $G(\omega) = 1 - z_1 = 1 - y_1$, then

$$\begin{aligned} C(\omega) - G(\omega) &= \frac{(1-y_1)y_2 + (1-y_2)y_1}{y_1 + y_2} - (1 - y_1) \\ &= \frac{y_1(y_1 - y_2)}{y_1 + y_2} \geq 0 \end{aligned}$$

if $y_1 \geq y_2$, and

$$\begin{aligned} 2\,C(\omega) - G(\omega) &= \frac{2\,[(1-y_1)y_2 + (1-y_2)y_1]}{y_1 + y_2} - (1 - y_1) \\ &= \frac{(1-y_1)(y_2 - y_1) + 2\,y_1(1-y_2)}{y_1 + y_2} \geq 0 \end{aligned}$$

if $y_2 \geq y_1$. On the other hand

$$|\hat{g}(\omega)| = \prod_{k=1}^{\infty} G(R^{-k}\omega)\ ; \tag{5.30}$$

to majorate $|\hat{g}(\omega)|$ for $2^{j/2} \leq |\omega| \leq 2^{(j+1)/2}$ we only need to majorate the j first factors in (5.30). Since R rotates by $\pi/4$, half of the factors can be majorated by $C(\omega)$ and the others by $2C(\omega)$. This leads to

$$|\hat{g}(\omega)| \leq C\, 2^{\log(1+|\omega|)/\log 2} \prod_{1 \leq k \leq 2\log(1+|\omega|)/\log 2} C(R^{-k}\omega) \tag{5.31}$$

and thus

$$\hat{g}(\omega) \leq C\,(1+|\omega|)^{-1}\ . \tag{5.32}$$

It is easy to check (in a similar way as for $\hat{\phi}_1(\omega)$) that this estimate is optimal. An immediate consequence is that the Fourier transform of the twin dragon characteristic function χ_Δ satisfies

$$\hat{\chi}_\Delta(\omega) \leq C\,(1+|\omega|)^{-1/2} \tag{5.33}$$

which was not obvious since we did not have a formula similar to (5.5) for $\hat{\chi}_\Delta$.

We now return to the construction of our biorthogonal bases and attack the problem of obtaining isotropic wavelet bases with arbitrarily high regularity.

V.2. Biorthogonal wavelet bases with arbitrarily high regularity.

We now consider the whole family of filters

$$\left\{M_0^N(\omega),\ \tilde{M}_0^{N,L}(\omega)\right\}_{N,L>0}$$

defined by (5.1) and (5.2).

A first remark is that the regularity of the functions ϕ_N increases linearly with N. More precisely, since

$$\phi_N(x) \;=\; (*)^N \phi_1(x)\ , \tag{5.34}$$

we can use the characterization of the optimal decay exponent for $\hat{\phi}_1(\omega)$ established in Proposition 5.1 to estimate the regularity index $\alpha(N)$ of $\phi_N(x)$. This leads to

$$2N-2 \leq \alpha(N) \leq 2N \tag{5.35}$$

and thus to

$$\lim_{N \to +\infty} \frac{\alpha(N)}{N} = 2 . \tag{5.36}$$

The estimate (5.35) is of course more interesting for large values of N than for small values where the error is comparable with the regularity.

For $N = 1$, we have seen that $\alpha \simeq 0.61$.

For $N = 2$, the Littlewood-Paley approach is still reasonable; using the symmetries reduces the size of the matrix to 9×9. Analyzing the eigenvalues, one finds that ϕ_2 is in C^α with $\alpha \simeq 2.93$. The function $\phi_2 = \phi_1 * \phi_1$ looks very smooth indeed on figure 13.

For $N \geq 3$, the matrix becomes too large to tackle by hand. In all cases the regularity of the wavelet $\psi_{N,L}$ will of course be the same as that of ϕ_N. The problem is now to find the appropriate dual function for the analysis. More precisely we want to design the filter

$$\tilde{M}_0^{N,L}(\omega) = (1 - z)^L \, P_{N+L}(z) \tag{5.37}$$

by choosing the number L in such way that the hypotheses of Theorem 2.2 (in its bidimensional generalization) are satisfied, *i.e.* that we have at least

$$\left| \hat{\tilde{\phi}}_{N,L}(\omega) \right| \leq C \, (1 + |\omega|)^{-1-\varepsilon} , \qquad \varepsilon > 0 . \tag{5.38}$$

To show that such a choice is possible for any value of N (*i.e.* for an arbitrarily regular synthesis function), we need an asymptotical result of the same nature as Theorem 2.4. We want to be sure that the regularizing action of the factor $(1 - z)^L$ can compensate the inverse effect of P_{N+L} if L is large enough.

Using a similar approach, we consider the simplest fixed point of $\omega \longmapsto D\omega$ modulo $2\pi\mathbb{Z}^2$, and modulo sign changes and the exchange of ω_1 and ω_2. This fixed point is $\omega_0 = (2\pi/5, 4\pi/5)$ which corresponds to $z_0 = z(\omega_0) = 5/8$.

We now decompose $\tilde{M}_0^{N,L}$ into three factors, by introducing the function $C(\omega)$ defined by (5.11):

$$\tilde{M}_0^{N,L}(\omega) = [C(\omega)]^L \, [Q(\omega)]^L \, P_{N+L}(z) = [C(\omega)]^L \, B_{N,L}(\omega) \tag{5.39}$$

with

$$Q(\omega) = \frac{M_0^1(\omega)}{C(\omega)} = \frac{(y_1 + y_2)(2 - y_1 - y_2)}{2\,(y_1 + y_2 - 2y_1 y_2)} .$$

We already know from Section II.3.b that

$$P_N(z) \leq (4z)^{N-1} \qquad \text{if } z \geq \frac{1}{2} . \tag{5.40}$$

From the Bezout relation (3.30), we also have

$$P_N(z) \leq \left(\frac{1}{1-z}\right)^N . \tag{5.41}$$

Consequently, we can roughly majorate $P_N(z)$ by

$$P_N(z) \leq \left[\min\{\frac{1}{1-z},\ \max\{4z,2\}\}\right]^N \qquad \text{if } z \in [0,1] . \tag{5.42}$$

Defining $H(\omega) = \min\{1/(1-z),\ \max\{4z,2\}\}$ and $G(\omega) = H(\omega)Q(\omega)$, (5.39) leads us to

$$\tilde{M}_0^{N,L}(\omega) \leq [C(\omega)]^L\ [G(\omega)]^L\ [H(\omega)]^N . \tag{5.43}$$

We are now facing a similar situation as in Theorem 2.4 where we had shown that the function $g(y) = \max\{2, 4y\} = h(\omega)$ satisfied

(5.44)
$$\begin{cases} h(\omega) = g(y) \leq g\,(3/4) & \text{if } y \leq 3/4\,, \\ h(\omega)h(2\,\omega) = g(y)\,g(4y(1-y)) \leq [g\,(3/4)]^2 & \text{if } 3/4 \leq y \leq 1\,. \end{cases}$$

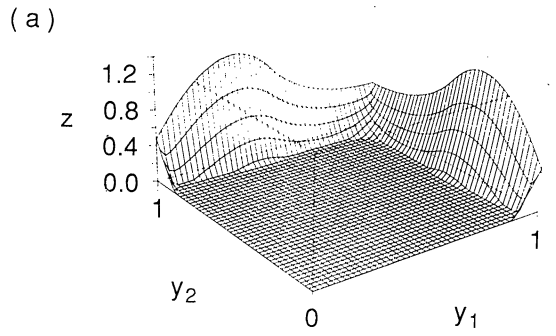
In the present case, although we do not dispose of any simple mathematical proof, numerical evidence shows that we have

$$\begin{cases} G(\omega)G(D\omega) \leq [G(\omega_0)]^2 & \text{or if not,} \\ G(\omega)G(D\omega)G(D^2\omega) \leq [G(\omega_0)]^3 \end{cases} \tag{5.45}$$

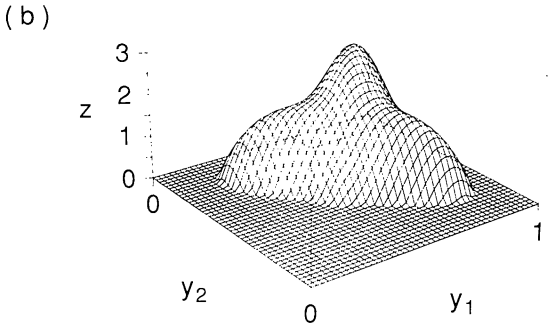
and similarly

$$\begin{cases} H(\omega)H(D\omega) \leq [H(\omega_0)]^2 & \text{or if not,} \\ H(\omega)H(D\omega)H(D^2\omega) \leq [H(\omega_0)]^3 \end{cases} \tag{5.46}$$

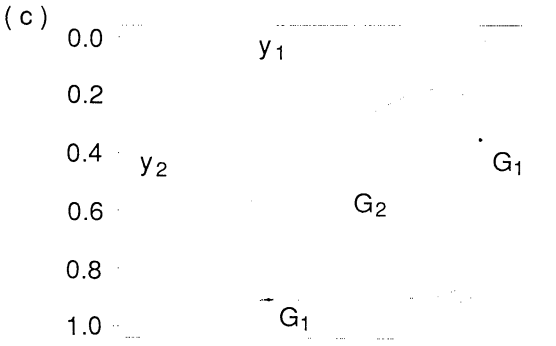
These two statements are illustrated respectively in figures 11 and 12.



a) Graph of $G_1(y_1, y_2) = \max\{G(\omega)G(D\omega), [G(\omega_0)]^2\} - [G(\omega_0)]^2$



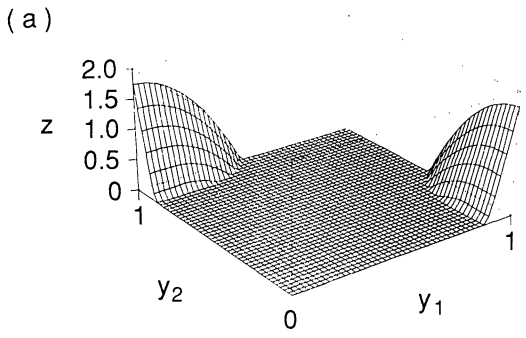
b) Graph of $G_2(y_1, y_2) = \max\{G(\omega)G(D\omega)G(D^2\omega), [G(\omega_0)]^3\} - [G(\omega_0)]^3$



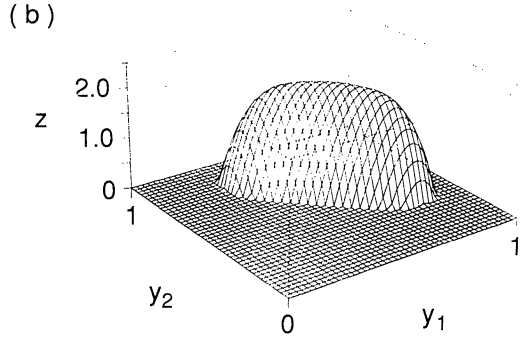
c) Compared supports of $G_1(y_1, y_2)$ and $G_2(y_1, y_2)$

Figure 11

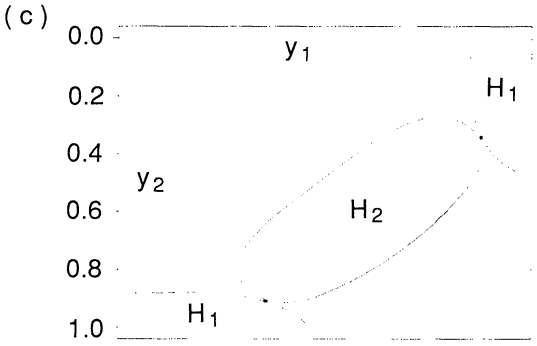
Graphic proof of (5.45)



a) Graph of $H_1(y_1, y_2) = \max\{H(\omega)H(D\omega), [H(\omega_0)]^2\} - [H(\omega_0)]^2$



b) Graph of $H_2(y_1, y_2) = \max\{H(\omega)H(D\omega)H(D^2\omega), [H(\omega_0)]^3\} - [H(\omega_0)]^3$



c) Compared supports of $H_1(y_1, y_2)$ and $H_2(y_1, y_2)$

Figure 12

Graphic proof of (5.46)

On a) and b) of each of these figures we have plotted the functions

$$\max\{F(\omega)F(D\omega), [F(\omega_0)]^2\} \; - \; [F(\omega_0)]^2$$

and

$$\max\{F(\omega)F(D\omega)F(D^2\omega), [F(\omega_0)]^3\} \; - \; [F(\omega_0)]^3$$

for $F = G$ and H (the coordinates are $(y_1, y_2) \in [0,1]^2$). On c) the supports of a) and b) are shown to be disjoint regions in $[0,1]^2$.

We now estimate $\hat{\tilde{\phi}}_{N,L}(\omega)$. From(5.39) and (5.43) we get

$$\begin{aligned}
\hat{\tilde{\phi}}_{N,L}(\omega) &= \prod_{k=1}^{+\infty} \left[C(D^{-k}\omega)\right]^L B_{N,L}(D^{-k}\omega) \\
&\leq C\,(1+|\omega|)^{-2L} \prod_{1\leq k\leq 2\log(1+|\omega|)/\log 2} B_{N,L}(D^{-k}\omega) \\
&\leq C\,(1+|\omega|)^{-2L} \Big[\prod_{1\leq k\leq 2\log(1+|\omega|)/\log 2} G(D^{-k}\omega)\Big]^L \\
&\quad \cdot \Big[\prod_{1\leq k\leq 2\log(1+|\omega|)/\log 2} H(D^{-k}\omega)\Big]^N .
\end{aligned}$$

Using (5.45) and (5.46) to divide these products in groups of two of three factors which satisfy one of the inequalities, this leads to

$$\hat{\tilde{\phi}}_{N,L}(\omega) \leq C\,(1+|\omega|)^{-2L+2(L\log G(\omega_0) \,+\, N\log H(\omega_0))/\log 2} \tag{5.47}$$

or

$$\hat{\tilde{\phi}}_{N,L}(\omega) \leq C\,(1+|\omega|)^{2L(\alpha-1)+2N\beta} \tag{5.47'}$$

with

$$\alpha = \frac{\log(G(\omega_0))}{\log 2} \simeq 0.907 \quad \text{and} \quad \beta = \frac{\log(H(\omega_0))}{\log 2} \simeq 1.322 \, .$$

Fortunately $\alpha < 1$. This means $\hat{\phi}_{L,N}(x)$ can be made arbitrarily regular by choosing L large enough. In particular, (5.38) will be satisfied if we have

$$2L(\alpha-1) + 2N\beta \; < \; -1 \, . \tag{5.48}$$

The smallest L such that M_0^N and $\tilde{M}_0^{N,L}$ generate unconditional multiscale bases is therefore given asymptotically by

$$L(N) \simeq \frac{\beta N}{1-\alpha} = \frac{\log 5 - \log 2}{\log 16 - \log 15} N \simeq 14.2\, N \,. \tag{5.49}$$

This asymptotical estimate is moreover optimal. Indeed define $\omega_j = D^j \omega_0 = D^j (2\pi/5, 4\pi/5)$. Because of the fixed point property of ω_0, we clearly have

$$\hat{\tilde{\phi}}^{N,L}(\omega_j) \sim C \left[\tilde{M}_0^{N,L}(\omega_0)\right]^j \sim C\, |\omega_j|^{-\gamma} \tag{5.50}$$

with

$$\gamma = \frac{-2 \log \tilde{M}_0^{N,L}(\omega_0)}{\log 2} \,. \tag{5.51}$$

From the definition (5.2) of $\tilde{M}_0^{N,L}$, we get

$$\begin{aligned} \tilde{M}_0^{N,L}(\omega_0) &= \left(\frac{3}{8}\right)^L P_{N+L}\left(\frac{5}{8}\right) \\ &\geq \left(\frac{3}{8}\right)^L \binom{2(N+L-1)}{N+L-1} \left(\frac{5}{8}\right)^{N+L-1} \\ &\geq C \left(\frac{3}{8}\right)^L \left(\frac{5}{2}\right)^{N+L} = C \left(\frac{15}{16}\right)^L \left(\frac{5}{2}\right)^N \end{aligned}$$

and thus

$$\begin{aligned} \gamma &\leq C + 2L \frac{\log 16/15}{\log 2} - 2N \frac{\log 5/2}{\log 2} \\ &= C + 2L(1-\alpha) - 2N\beta \,. \end{aligned}$$

It follows that the estimate (5.49), if true, is certainly optimal. While we expect (5.45), (5.46), hence (5.49), to be true, we have unfortunately no rigorous proof. However, we *can* prove inequalities which are slightly less strong than (5.45), (5.46), leading to a non-optimal but rigorous estimate for $L(N)$. More precisely, we can prove that $\Omega = [-\pi, \pi]^2$ can be split up as $\Omega = \Omega_1 \cup \Omega_2 \cup \Omega_3$, with

$$\begin{cases} G(\omega) \leq \xi & \omega \in \Omega_1 \,, \\ G(\omega)\, G(D\omega) \leq \xi^2 & \omega \in \Omega_2 \,, \\ G(\omega)\, G(D\omega)\, G(D^2\omega) \leq \xi^3 & \omega \in \Omega_3 \,, \end{cases} \tag{5.52}$$

with $\xi/2 \simeq .9588 < 1$, resulting in (5.47') with

$$\alpha = \frac{\log \xi}{\log 2} \simeq .93982 .$$

If we use the crude estimate $H(\omega) \leq 4$ for all $\omega \in [-\pi, \pi]^2$, corresponding to $\beta = 2$, then this leads to

$$L(N) \simeq \frac{\beta}{1-\alpha} N \simeq 32.959 \; N \; ;$$

this factor is about twice as large as in (5.49). The detailed proof of this estimate is in Appendix C.

All these results can be summarized in the following theorem.

Theorem 5.3. *The family of dual filters* $\{M_0^N(\omega), \; \tilde{M}_0^{N,L}(\omega)\}_{N,L>0}$ *generates biorthogonal bases of compactly supported wavelets with arbitrarily high regularity. For large values of* N, *the Hölder exponent of* $\phi_N(x)$ *is equivalent to* $2N$ *and the minimal choice for* L *is asymptotically proportional to* N,

$$L(N) \simeq \mathcal{K} N \; , \tag{5.53}$$

with $14.215 \leq \mathcal{K} \leq 32.959$.

Here the upper bound on $\mathcal{K}$ is not tight, and we expect $\mathcal{K} = 14.215$ to hold, as indicated above.

REMARK. By taking L larger than $L(N)$, $\hat{\phi}^{N,L}$ can also be made arbitrarily regular. However, in many applications such as coding, approximation, data storage and compression, we do not really care about the regularity of the analyzing functions $\tilde{\psi}$ and $\hat{\phi}$; only the synthesis function ψ and ϕ have to be smooth since this property is important for the cascade-reconstruction algorithm. This justifies the choice of the minimal value $L(N)$ such that the families $\{2^{j/2} \; \psi_{N,L}(D^j x - k)\}_{j \in \mathbb{Z}, \; k \in \mathbb{Z}^2}$ and $\{2^{j/2} \; \tilde{\psi}_{N,L}(D^j x - k)\}_{j \in \mathbb{Z}, \; k \in \mathbb{Z}^2}$ are unconditional dual bases of $L^2(\mathbb{R})$. Recall that the existence of frame bounds is essential for the stability of the subband coding scheme.

We end this section by taking a closer look at the size of these dual filters.

V.3. Size and optimal implementation of the dual filters.

The asymptotical ratio $L(N)/N \simeq 14.2$ is big in the sense that the filter $\tilde{M}_0^{NL(N)}$ may have a very large number of taps. More precisely, a polynomial $P(z)$ of degree p corresponds to a filter with $p^2 + (p+1)^2$ nonzero coefficients. For example, if $N = 3$,

$$\tilde{M}_0^{NL(N)}(\omega) = (1-z)^{L(N)}\ P_{N+L(N)}(z)$$

is according to (5.49) a polynomial of degree $p = N + 2L(N) \simeq 87$ in z. Consequently it is the transfer function of a filter with approximately 1350 taps!

It seems thus that the dual filter is, even for small values of N, much too large for a realistic implementation. This is not quite true for several reasons.

First, one can factorize the polynomial $P_{N+L(N)}(z)$ and express $\tilde{M}_0^{N,L}$ as a product of p monomials in z. By applying successively these monomial filters instead of using directly their product, the number of multiplications per sample in the filtering process is reduced from order p^2 to p. Note that this complexity reduction associated with the factorization is due to the multidimensional situation and does not occur in the $1D$ case.

Second, the filter corresponding to the variable z, *i.e.* the laplacian discrete scheme, has coefficients $c_{0,0} = 1/2$ and $c_{1,0} = c_{-1,0} = c_{0,1} = c_{0,-1} = -1/8$. It can thus be implemented by using binary shifts instead of multiplications. This is very important since a binary shift is usually performed 10 times faster than an addition and 100 times faster than a multiplication in most processors. This shows that only the additions count here. If t is the time for one multiplication, each monomial filter will generate one sample in approximately $3t/5$ and the same operation will take $3pt/5$ for the whole filter. For $N = 3$ and $p = 87$, this corresponds to the complexity of a 52 tap filter which is much more reasonable than the first estimation.

Finally, for small values of N, it is clear that the asymptotical estimate (5.49) of $L(N)$ is far from sharp, just as, in $1D$, the asymptotical estimate on regularity of Section II.3.b was ill-suited to small filters. A better estimate for $L(N)$ can be found by checking that the optimal decay exponent for $\hat{\tilde{\phi}}(\omega)$ is exactly determined by the value of $\tilde{M}_0^{N,L}$ at $\omega_0 = (2\pi/5,\ 4\pi/5)$. More precisely recall that we have

$$\tilde{M}_0^{N,L}(\omega) \ = \ [C(\omega)]^L\ B_{N,L}(\omega)\ .$$

For the small values of N and L considered below, one can check by the same graphical arguments that the inequalities (5.45) or (5.46) are also satisfied by $B_{N,L}(\omega)$, *i.e.*

$$\text{(5.54)} \qquad \begin{cases} B_{N,L}(\omega)\, B_{N,L}(D\omega) \leq [B_{N,L}(\omega_0)]^2 & \text{or if not,} \\ B_{N,L}(\omega)\, B_{N,L}(D\omega)\, B_{N,L}(D^2\omega) \leq [B_{N,L}(\omega_0)]^3 \ . \end{cases}$$

In order for (5.38) to be satisfied, we therefore only need

$$\text{(5.55)} \qquad \tilde{M}_0^{N,L}(\omega_0) \leq \frac{\sqrt{2}}{2}$$

and this will be sufficient for these small values of N and L for which (5.52) holds. Using the definition (5.2) of $\tilde{M}_0^{N,L}$ we obtain

- for $N = 1$, $L(1) = 3$,
- for $N = 2$, $L(2) = 12$,
- for $N = 3$, $L(3) = 22$.

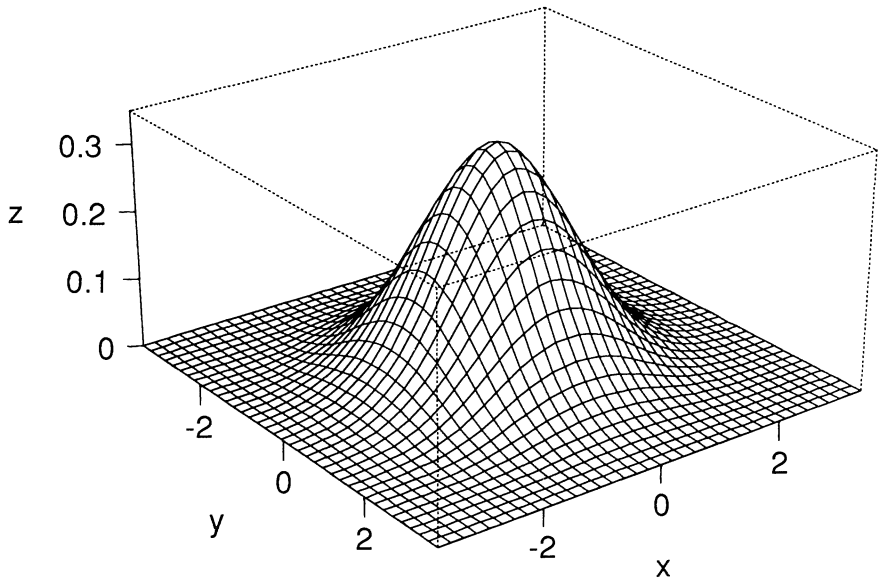
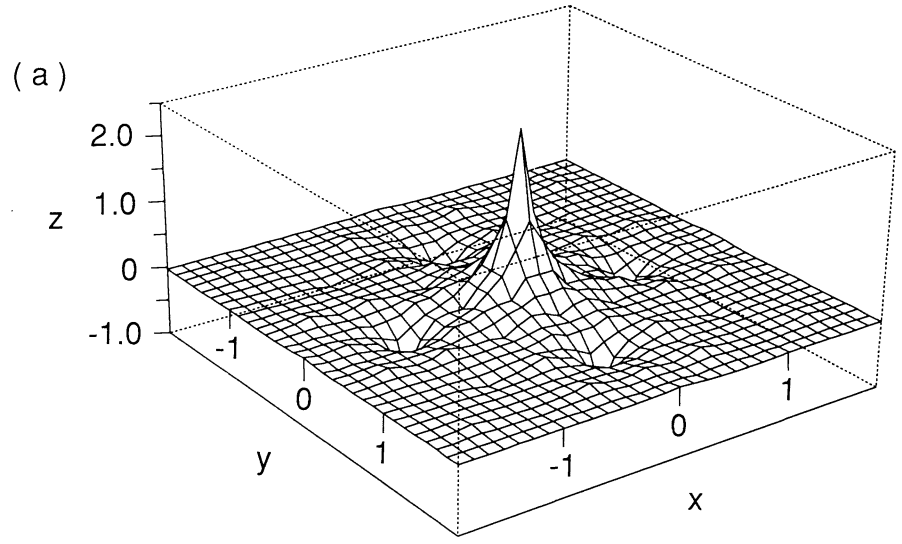
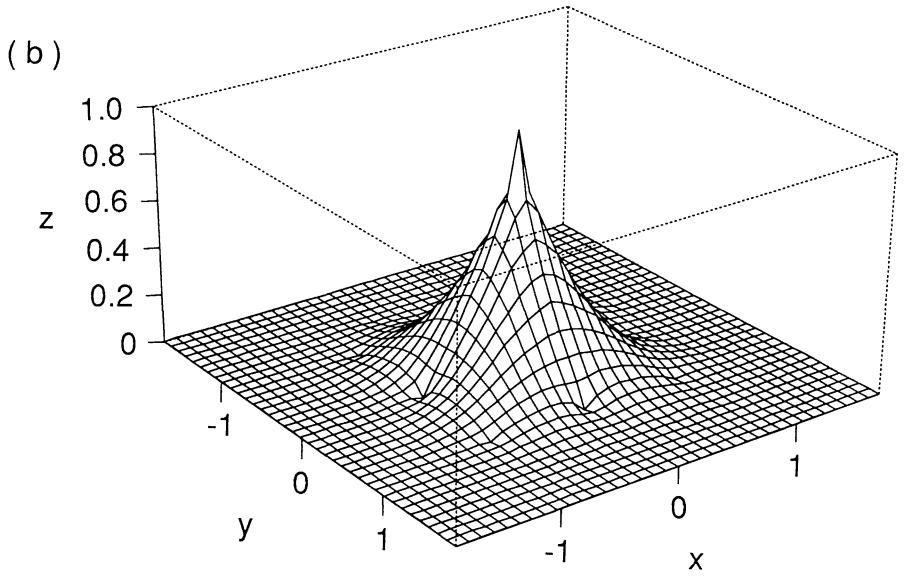


Figure 13
The scaling function ϕ_2 $(= \phi_1 * \phi_1)$.



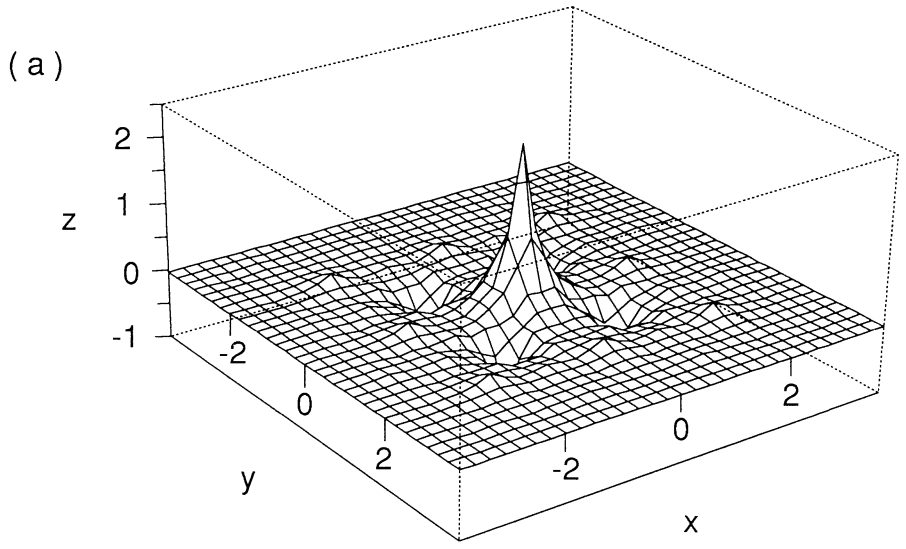
a) $\hat{\phi}_{12}$



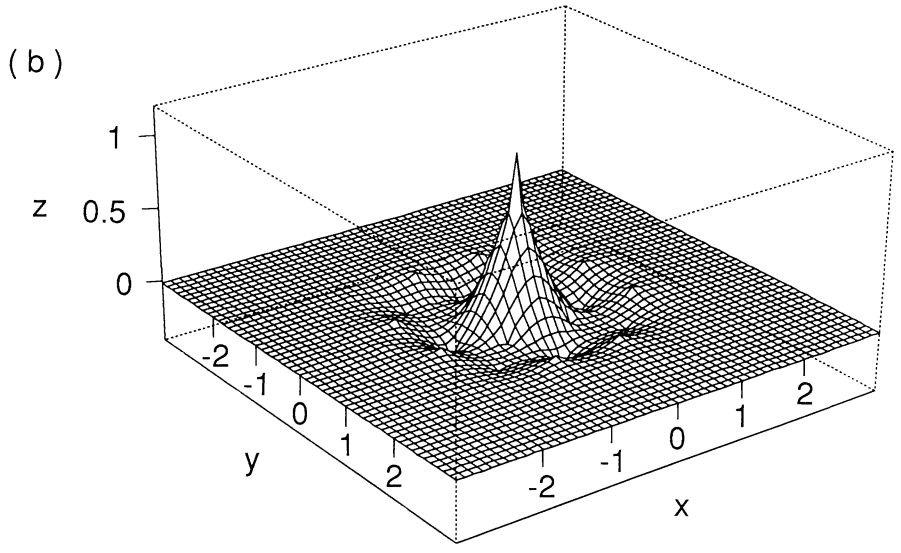
b) ϕ_1

Figure 14

Analysis and synthesis scaling function for $N = 1$ and $L = 2$.



a) $\tilde{\psi}_{12}$



b) ψ_{12}

Figure 15

Analysis and synthesis wavelets for $N = 1$ and $L = 2$.

Clearly these estimates are much better than (5.49). Finally, $L(N)$ can be even more reduced, for small values of N, if an even sharper criterion that the frequency decay (5.38) is used to ensure the existence of frame bounds. We show indeed in Appendix A that the spectral analysis of the transition operators T_0 and $\tilde{T}_0$ corresponding to the functions $|M_0|^2$ and $|\tilde{M}_0|^2$ can be used to derive both the frame property and the L^2 convergence needed to have a pair of dual Riesz bases. In [CD] we prove that this criterion is sharp so that the value of $L(N)$ is here optimal. Unfortunately the matrices of T_0 and $\tilde{T}_0$ can be very big, even for small N and L.

For $N = 1$, we now obtain $L(N) = 2$ so that the two filters M_0^1 and $\tilde{M}_0^{12}$ are of small size. We show on figure 14 and 15 the scaling functions and wavelets obtained from such a choice.

Appendix A: A sharp criterion for frame bounds.

We want to give here a better result than Theorem 2.2 to characterize the dual filter pair $(m_0, \tilde{m}_0)$ which lead to biorthogonal Riesz bases of wavelets. The method that we show here uses the transition operators associated to the positive functions $|m_0|^2$ and $|\tilde{m}_0|^2$ (see Section II.4.a).

First, recall that the φ, $\tilde{\varphi}$, ψ and $\tilde{\psi}$ are defined by

$$\text{(A.1)} \qquad \begin{cases} \hat{\varphi}(\omega) = \prod_{k=1}^{+\infty} m_0(2^{-k}\omega)\,, & \hat{\psi}(2\omega) = m_1(\omega)\,\hat{\varphi}(\omega)\,, \\ \hat{\tilde{\varphi}}(\omega) = \prod_{k=1}^{+\infty} \tilde{m}_0(2^{-k}\omega)\,, & \hat{\tilde{\psi}}(2\omega) = \tilde{m}_1(\omega)\,\hat{\tilde{\varphi}}(\omega)\,. \end{cases}$$

As mentioned in Theorem 2.2 the duality relations (2.19), (2.20) and the decomposition formula (2.21) are ensured as soon as the partial products

$$\hat{\varphi}_n(\omega) = \prod_{k=1}^{n} m_0(2^{-k}\omega)\chi_{[-2^n\pi,\,2^n\pi]}(\omega)\,,$$

$$\hat{\tilde{\varphi}}_n(\omega) = \prod_{k=1}^{n} \tilde{m}_0(2^{-k}\omega)\chi_{[-2^n\pi,\,2^n\pi]}(\omega)$$

converge in $L^2(\mathbb{R})$ respectively to $\hat{\varphi}(\omega)$ and $\hat{\tilde{\varphi}}(\omega)$.

The main difficulty is then to obtain the frame bounds A, B, $\tilde{A}$ and $\tilde{B}$ all strictly positive such that for all f in $L^2(\mathbb{R})$,

$$\text{(A.2)} \qquad \begin{cases} A\,\|f\|^2 \leq \sum_{j,k\in\mathbb{Z}} |\langle f, \psi_k^j\rangle|^2 \leq B\,\|f\|^2\,, \\ \tilde{A}\,\|f\|^2 \leq \sum_{j,k\in\mathbb{Z}} |\langle f, \tilde{\psi}_k^j\rangle|^2 \leq \tilde{B}\|f\|^2\,. \end{cases}$$

It is sufficient to obtain the two upper bounds of (A.2) because the lower bounds are then obtained by using (2.21) and the Schwarz inequality which give

$$\text{(A.3)} \qquad \|f\|^2 \leq \left(\sum_{j,k} |\langle f, \psi_k^j\rangle|^2\right)^{1/2} \left(\sum_{j,k} |\langle f, \tilde{\psi}_k^j\rangle|^2\right)^{1/2}$$

In [CDF] we used the following assumption

$$\text{(A.4)} \qquad |\hat{\psi}(\omega)|^2 + |\hat{\tilde{\psi}}(\omega)|^2 \leq C\,(1+|\omega|)^{-1/2-\varepsilon}$$

which can also be formulated with φ and $\tilde{\varphi}$ instead of ψ and $\tilde{\psi}$. Here, we shall prove the L^2-convergence of $\{\varphi_n, \tilde{\varphi}_n\}_{n>0}$ and the frame inequalities (A.2) using weaker assumptions. More precisely let T_0 and $\tilde{T}_0$ be the two transition operators associated to the functions $|m_0|^2$ and $|\tilde{m}_0|^2$, as defined in Section II.4.a. They both operate in two spaces of trigonometric polynomials E_N and $E_{\tilde{N}}$. We have proved in Lemma 2.2 that the subspaces $F_N = \{f \in E_N : f(0) = 0\}$ and $F_{\tilde{N}} = \{f \in E_{\tilde{N}} : f(0) = 0\}$ are invariant under the action of T_0 and $\tilde{T}_0$. The following result gives a sharp characterization of the dual filter pairs associated to biorthogonal wavelet bases.

Theorem A.1. *Let λ (respectively $\tilde{\lambda}$) be the largest eigenvalue of T_0 (respectively, $\tilde{T}_0$) in the subspace F_N (respectively, $F_{\tilde{N}}$). Then if $|\lambda|$ and $|\tilde{\lambda}|$ are both strictly inferior to 1, the functions ψ and $\tilde{\psi}$ defined by* (A.1) *generate biorthogonal Riesz bases of wavelets $\{\psi_k^j, \tilde{\psi}_k^j\}_{j,k\in\mathbb{Z}}$.*

PROOF. We shall prove here that this condition on the eigenvalues of T_0 and $\tilde{T}_0$ is sufficient to obtain biorthogonal wavelet bases. In fact, it is also a necessary condition. This result is detailed in [CD].

We first show that φ and ψ are in $L^2(\mathbb{R})$. As in Theorem 2.7, we apply T_0^n to the function $C_1(\omega) = 1 - \cos\omega$ which is in F_N and by using Lemma 2.5, we obtain

$$\begin{aligned}\int_{2^{n-1}\pi \le |\omega| \le 2^n\pi} |\hat{\varphi}(\omega)|^2 d\omega &\le C \int_{2^{n-1}\pi \le |\omega| \le 2^n\pi} |\hat{\varphi}_n(\omega)|^2 d\omega \\ &\le C \int_{-2^n\pi}^{2^n\pi} C_1(2^{-n}\omega) |\hat{\varphi}_n(\omega)|^2 d\omega \\ &\le C \left(\frac{\gamma+1}{2}\right)^n\end{aligned}$$

because $(\gamma+1)/2 > \gamma$. Since we also have $(\gamma+1)/2 < 1$, it follows that the dyadic blocks in the Littlewood-Paley decomposition of φ satisfy the inequality

$$\|\Delta_j(\varphi)\|_{L^2} \le C\, 2^{-\varepsilon j} \quad \text{for some } \varepsilon > 0 \,. \tag{A.5}$$

This proves that φ and ψ are even better than L^2: They belong to a Besov space $B_2^{\varepsilon,\infty}$ ($\subset L^2(\mathbb{R})$) for some $\varepsilon > 0$. We shall use this property to prove the frame inequalities. Similarly $\tilde{\varphi}$ and $\tilde{\psi}$ belong to $B_2^{\tilde{\varepsilon},\infty}$ for some $\tilde{\varepsilon} > 0$. To prove the L^2 convergence of the sequence φ_n to φ, we remark that since $m_0(0) = 1$, for α in $]0, \pi]$ small enough we have

$$|\omega| \le \alpha \quad \text{implies} \quad |\hat{\varphi}(\omega)| \ge C > 0 \,. \tag{A.6}$$

We now introduce the sequence φ_n^α defined by

$$\hat{\varphi}_n^\alpha(\omega) \;=\; \prod_{k=1}^{n} m_0(2^{-k}\omega)\chi_{[-2^n\alpha,\, 2^n\alpha]}(\omega) \,. \tag{A.7}$$

It is clear that $\hat{\varphi}_n^\alpha(\omega)$ converges pointwise to $\hat{\varphi}(\omega)$, but (A.6) also implies $|\hat{\varphi}_n^\alpha(\omega)| \le |\hat{\varphi}(\omega)|/C$ for all $n > 0$. By the Lebesgue dominated convergence theorem we get

$$\lim_{n\to\infty} \|\varphi_n^\alpha - \varphi\|_{L^2} \;=\; 0 \,. \tag{A.8}$$

We now use the hypothesis on the eigenvalues to evaluate the L^2 norm of the difference $\varphi_n - \varphi_n^\alpha$

$$\int |\hat{\varphi}_n(\omega) - \hat{\varphi}_n^\alpha(\omega)|^2 d\omega = \int_{|\omega|>\alpha} |\hat{\varphi}_n(\omega)|^2 d\omega$$

$$\leq \frac{1}{C_1(\alpha)} \int_{-2^n\pi}^{2^n\pi} C_1(2^{-n}\omega)|\hat{\varphi}_n(\omega)|^2 d\omega$$
$$\leq C\, 2^{-\varepsilon n} \underset{n\to\infty}{\longrightarrow} 0\,.$$

Consequently φ_n converges to φ in $L^2(\mathbb{R})$ and the same holds for $\tilde{\varphi}_n$ and $\tilde{\varphi}$.

It remains to establish the upper frame inequalities in (A.2). We shall obtain them by using the following lemma.

Lemma A.2. *Let ψ be a function in $L^2(\mathbb{R})$ such that for some $\sigma > 0$,*

$$\sum_{k\in\mathbb{Z}} |\hat{\psi}(\omega + 2k\pi)|^{2-\sigma} \leq C_1 \tag{A.9}$$

$$\sum_{j\in\mathbb{Z}} |\hat{\psi}(2^{-j}\omega)|^{\sigma} \leq C_2 \tag{A.10}$$

uniformly in ω. Define, for j, k in $\mathbb{Z}$, $\psi_k^j(x) = 2^{-j/2}\psi(2^{-j}x - k)$. Then, for all f in $L^2(\mathbb{R})$,

$$\sum_{j,k\in\mathbb{Z}} |\langle f, \psi_k^j\rangle|^2 \leq C_1 C_2 \|f\|^2\,. \tag{A.11}$$

Let us first assume that this result is true to conclude the proof of the theorem. We thus have to check that there exist a $\sigma > 0$ such that (A.9) and (A.10) are satisfied for ψ and $\tilde{\psi}$.

To check (A.10), we define $I_j = [-2^{j+1}\pi, -2^j\pi] \cup [2^j\pi, 2^{j+1}\pi]$. For $j \leq 1$, we can use the cancellation of $\hat{\psi}(\omega)$ at the origin to obtain

$$\max_{\omega\in I_j} |\hat{\psi}(\omega)| \leq C\, 2^j \qquad \text{for } j \leq 1\,. \tag{A.12}$$

For $j \geq 2$, we know that $\hat{\psi}(\pm 2^j\pi) = 0$ since $\hat{\varphi}(2k\pi) = 0$ for $k \in \mathbb{Z}\backslash\{0\}$. We thus have

$$\max_{\omega\in I_j} |\hat{\psi}(\omega)|^2 \leq \int_{I_j} \frac{d}{d\omega}\left(|\hat{\psi}|^2\right) d\omega$$
$$\leq 2\int_{I_j} \left|\hat{\psi}(\omega)\frac{d\hat{\psi}}{d\omega}(\omega)\right| d\omega$$

$$\leq 2\Big[\int_{I_j} |\hat{\psi}(\omega)|^2 \, d\omega\Big]^{1/2} \Big[\int_{\mathbb{R}} \Big|\frac{d\hat{\psi}}{d\omega}(\omega)\Big|^2 \, d\omega\Big]^{1/2} .$$

The first factor can be majorated by $2^{-\varepsilon j}$ because we have proved that ψ belongs to $B_2^{\varepsilon,\infty}$. The second factor is finite since it is proportional to $\int |x\psi(x)|^2 \, dx$ and ψ is a compactly supported L^2 function. Consequently

$$\max_{\omega \in I_j} |\hat{\psi}(\omega)| \leq C \, 2^{-\varepsilon j/2} \qquad \text{for } j \geq 2 . \tag{A.13}$$

Combining (A.12) and (A.13) we see that (A.10) holds for all $\sigma > 0$, since we have

$$\begin{aligned} \max_{\omega \in \mathbb{R}} \sum_{j \in \mathbb{Z}} |\hat{\psi}(2^j \omega)|^\sigma &\leq \sum_{j \in \mathbb{Z}} \Big[\max_{\omega \in I_j} |\hat{\psi}(\omega)|\Big]^\sigma \\ &\leq C\Big[\sum_{j \leq 1} 2^{\sigma j} + \sum_{j \geq 2} 2^{-\varepsilon \sigma j/2}\Big] \\ &\leq C_2(\sigma) . \end{aligned}$$

We now check that (A.9) is satisfied for some $\sigma > 0$. Because the wavelet satisfies $\hat{\psi}(4k\pi) = 0$ for all $k \in \mathbb{Z}$, we can derive

$$\begin{aligned} \sum_{k \in \mathbb{Z}} |\hat{\psi}(\omega + 2k\pi)|^{2-\sigma} &\leq \sum_{k \in \mathbb{Z}} \int_{2k\pi}^{2k\pi + 2\pi} \Big|\frac{d}{d\omega}(|\hat{\psi}|^{2-\sigma})\Big| \, d\omega \\ &\leq \int_{\mathbb{R}} \Big|\frac{d}{d\omega}[|\hat{\psi}|^2]^{1-\sigma/2}\Big| \, d\omega \\ &\leq |2-\sigma| \int_{\mathbb{R}} |\hat{\psi}(\omega)|^{1-\sigma} \Big|\frac{d\hat{\psi}}{d\omega}\Big| \, d\omega \\ &\leq |2-\sigma|\Big[\int_{\mathbb{R}} |\hat{\psi}(\omega)|^{2-2\sigma}\Big]^{1/2} \Big[\int_{\mathbb{R}} \Big|\frac{d\hat{\psi}}{d\omega}\Big|^2 \, d\omega\Big]^{1/2} . \end{aligned}$$

We already saw that the second factor was finite (in the proof of (A.10)). The first factor is also finite for σ small enough. Indeed, using $\psi \in B_2^{\varepsilon,\infty}$ and the Hölder inequality, we obtain

$$\int_{I_j} |\hat{\psi}(\omega)|^{2-2\sigma} d\omega \leq \left[\int_{I_j} |\hat{\psi}(\omega)|^2 \, d\omega\right]^{1-\sigma} (2^{j+1}\pi)^\sigma$$

$$\leq C\ 2^{j(\sigma-2\varepsilon(1-\sigma))}\ .$$

We thus have to choose $\sigma > 0$ such that $\sigma - 2\varepsilon(1-\sigma) < 0$, *i.e.* $\sigma < 2\varepsilon/(1+2\varepsilon)$. Since the same results also hold for the dual wavelet $\tilde{\psi}$, the theorem is proved modulo Lemma A.2 that we tackle now . Using the Plancherel and the Poisson formulas, we derive for any f in $L^2(\mathbb{R})$

$$\begin{aligned}
\sum_{k\in\mathbb{Z}} |\langle f, \psi_k^j\rangle|^2 &= \frac{1}{4\pi^2} \sum_{k\in\mathbb{Z}} 2^j \left| \int_{\mathbb{R}} \hat{f}(\omega)\ \overline{\hat{\psi}(2^j\omega)}\ e^{-i2^j k\omega}\ d\omega \right|^2 \\
&= \frac{1}{4\pi^2} \sum_{k\in\mathbb{Z}} 2^{-j} \left| \int_{\mathbb{R}} \hat{f}(2^{-j}\omega)\ \overline{\hat{\psi}(\omega)}\ e^{-ik\omega}\ d\omega \right|^2 \\
&= \frac{1}{2\pi} \int_0^{2\pi} 2^{-j} \left| \sum_{\ell\in\mathbb{Z}} \hat{f}(2^{-j}(\omega+2\ell\pi))\ \overline{\hat{\psi}(\omega+2\ell\pi)} \right|^2 d\omega \\
&\leq \frac{2^{-j}}{2\pi} \int_0^{2\pi} \left(\sum_{\ell\in\mathbb{Z}} |\hat{f}(2^{-j}(\omega+2\ell\pi)|\ |\hat{\psi}(\omega+2\ell\pi)|^{\sigma/2} \right. \\
&\qquad \left. \cdot\ |\hat{\psi}(\omega+2\ell\pi)|^{1-\sigma/2} \right)^2 d\omega \\
&\leq \frac{2^{-j}}{2\pi} \int_0^{2\pi} \left(\sum_{\ell\in\mathbb{Z}} |\hat{f}(2^{-j}(\omega+2\ell\pi))|^2\ |\hat{\psi}(\omega+2\ell\pi)|^{\sigma} \right) \\
&\qquad \cdot \left(\sum_{\ell\in\mathbb{Z}} |\hat{\psi}(\omega+2\ell\pi)|^{2-\sigma} \right) d\omega \\
&\leq C_1\ \frac{2^{-j}}{2\pi} \int_{\mathbb{R}} |\hat{f}(2^{-j}\omega)|^2\ |\hat{\psi}(\omega)|^{\sigma}\ d\omega \\
&= \frac{1}{2\pi}\ C_1 \int_{\mathbb{R}} |\hat{f}(\omega)|^2\ |\hat{\psi}(2^j\omega)|^2\ d\omega\ .
\end{aligned}$$

Summing on all the scales $j \in \mathbb{Z}$ and using (A.10), we get

$$\text{(A.14)} \qquad \sum_{j,k\in\mathbb{Z}} |\langle f, \psi_k^j\rangle|^2 \leq \frac{C_1 C_2}{2\pi} \int_{\mathbb{R}} |\hat{f}(\omega)|^2 d\omega\ =\ C_1 C_2\ \|f\|^2$$

and this concludes the proof.

Appendix B. Dragonic expansions.

In this Appendix we want to show how the one-dimensional techniques in [DL] can be extended to multidimensional situations. As an example we discuss the two-dimensional case, with the dilation matrix $R = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$.

A first multidimensional extension of [DL] can be found in [Mo]. Even though he looks at general matrices, Mongeau effectively reduces his analysis to pure dilations by considering the smallest n such that $\overline{D} = D^n$ is a multiple of the identity, and rewriting (by iteration) the two-scale equation so that it involves only $\overline{D}$. This procedure can drastically increase the number of different terms in the equation. We choose here to work directly with $D = R$ itself.

When the two-scale equation is one-dimensional, and the dilation factor is 2, the regularity at x of the function ϕ solving the equation is regulated by the binary expansion of x (for dilation factor p, the same role is played by the p-ary expansion). Moreover, $\mathbb{R}$ and in particular $\operatorname{supp}\phi$ is tiled with integer translates of the interval $[0,1]$, which can be viewed as the set of numbers equal to the decimal part only of their dyadic expansion; if N such tiles are needed to cover the support of ϕ, then the two-scale functional equation can be rewritten as an equation for an N-dimensional vector-valued function involving two matrices T_0 and T_1. The spectral properties of T_0, T_1 then determine the regularity of ϕ, both local and global [DL].

In the two-dimensional case with dilation matrix R, the role of elementary tile is now played by the twin dragon set Δ. It is defined by

$$\text{(B.1)} \qquad \left\{ x \in \mathbb{R}^2 : \quad x = \sum_{j=1}^{\infty} R^{-j} p_j \quad \text{where} \right.$$
$$\left. p_j \in L = \mathbb{Z}^2 / R\mathbb{Z}^2 = \{(0,0),(1,0)\} \right\}.$$

Under the standard identification of $\mathbb{R}^2$ with $\mathbb{C}$, with $(x,y) \simeq x+iy$, Δ can also be written as

$$\text{(B.2)} \quad \Delta = \left\{ z \in \mathbb{C} : \; z = \sum_{j=1}^{\infty} d_j \left(\frac{1+i}{2} \right)^j \text{ where } d_j = 0 \text{ or } 1 \right\} .$$

This set Δ is compact, has fractal boundary, is selfsimilar, and its $\mathbb{Z}^2$-translates tile the plane. The indicator function of Δ is the solution to the two-scale equation

$$\phi(x) \;=\; \phi(Rx) \;+\; \phi(Rx-(1,0))$$

(see [GM]). Δ is called the twin dragon set [K]. We shall give the name *dragonic expansions* to expansions of x or z as in (B.1), (B.2). Note that (as in the binary case) some points may have two different dragonic expansions, *e.g.* $.01000\ldots = ((1+i)/2)^2 \;=\; i/2 \;=\; .1011111\ldots$ (This example also illustrates that addition follows rules very different from the binary case, since $.0100\ldots \;+\; .0100\ldots \;=\; .1111\ldots$)

Suppose we are interested in various regularity properties of L^1-solutions ϕ of

$$\phi(x) \;=\; \sum_{k\in\Lambda} c_k\ \phi(Rx-k)\ , \tag{B.3}$$

where Λ is a finite subset of $\mathbb{Z}^2$. Such solutions are uniquely defined up to normalization and have necessarily compact support. One can determine the minimal set $\Gamma \subset \mathbb{Z}^2$ so that $R^{-1}(\Gamma+\Lambda-L)\subset\Gamma$; then $\operatorname{supp}\phi \subset \cup_{\ell\in\Gamma}(\Delta+\ell)$. The equation (B.3) for ϕ can be rewritten by defining the $|\Gamma|$-dimensional vector $v(x)$ by

$$v_j(x) \;=\; \phi(x+j)\,, \qquad j\in\Gamma,\ x\in\Delta\ , \tag{B.4}$$

we have

$$v_j(x) \;=\; \sum_k c_{Rj+d_1(x)-k}\ v_k(\tau x)$$

where $d_1(x)$ is the first digit in the dragonic expansion of x, and τx is the point obtained by dropping $d_1(x)$ from the same dragonic expansion of x,

$$\tau x \;=\; \sum_{j=1}^{\infty} d_{j+1}(x)\left(\frac{1+i}{2}\right)^j\ .$$

Equation (B.4) can be recast as

$$v(x) \;=\; T_{d_1(x)}\ v(\tau x)\ , \tag{B.5}$$

where $\qquad (T_0)_{jk} \;=\; c_{Rj-k}\ , \qquad (T_1)_{jk} \;=\; c_{Rj-k+(1,0)}\ .$

We have completed a setup analogous to that of [DL]. The question is now whether the proof techniques of [DL] still work in this case. The answer is basically yes. For instance, we still have

Theorem B.1. *Assume that the c_k in* (B.3) *satisfy*

$$\sum_n c_{Rn} = \sum_n c_{Rn+(1,0)} = 1 .$$

Then $e_1 = (1, 1, \ldots, 1)$ is a common lefteigenvector of T_0, T_1 with eigenvalue 1 for both matrices. Define E_1 to be the one-dimensional subspace orthogonal to e_1. If there exist $\lambda < 1$, $C > 0$ so that

$$\|T_{d_1} \cdots T_{d_m}|_{E_1}\| \leq C\,\lambda^m \tag{B.6}$$

for all possible $d_j = 0$ or 1, all $m \in \mathbb{N}$, then the L^1-solution ϕ to (B.3) *is Hölder continuous with exponent $\alpha = |\log\lambda|/\log\sqrt{2}$.*

This is the analog of Theorem 2.3 in [DL]. Two different strategies of proof are given in [DL]. The first one involves piecewise linear spline approximants; this technique would be hard to generalize here because of the fractal boundary of our domain building blocks $\Delta + k$. A second strategy, which does not use splines at all, but leads to longer proofs, is explained in the Appendix in [DL]; this strategy generalizes to the present case. The main point we have to check to make sure the proof carries over is whether elements that are close necessarily have dragonic expansions with the same starting digits. In the one-dimensional, binary case, if two dyadic rationals x, y are closer than 2^{-m}, $|x-y| < 2^{-m}$, then x and y have binary expansions with coinciding first m digits. (If *e.g.* $x \leq y < x + 2^{-m}$, then the expansion "from above" of x – ending in all zeros – has the same first m digits as the expansion "from below" of y – ending in all ones.) This is crucial in the proof, and allows to extract Hölder continuity from the condition (B.6). We therefore have to check whether a similar property holds in the "dragonic" case.

By analogy we shall call *dragonic rationals* all the points in Δ for which a terminating dragonic expansion can be written. Typically dragonic rationals also have other, non-terminating dragonic expansions. For each dragonic rational x the terminating expansion is unique; we denote its digits by $d_j^0(x)$, $j \in \mathbb{N}$.

Let us also introduce the notations R_0, R_1,

$$R_0 y = Ry, \qquad R_1 y = Ry + (1,0) ,$$

or $$R_d y = Ry + d(1,0), \qquad \text{with } d = 0 \text{ or } 1 .$$

Take now a fixed dragonic rational x, and assume that $d_j^0(x) = 0$ for $j > J$. All the $y \in \Delta$ that have the same first J digits $d_j^0(x)$, $j \le J$, constitute a little dragon $\Delta_J(x)$ themselves,

$$\Delta_J(x) = R^{-1}_{d_J(x)} \cdots R^{-1}_{d_1(x)} \Delta ;$$

x itself is the image of $(0,0)$ under the same map $R^{-1}_{d_J(x)} \cdots R^{-1}_{d_1(x)}$. The set Δ is tiled by 2^J little dragons of the same size as Δ_J, all translates of Δ_J. For every such little dragon, we call the point corresponding to $(0,0)$ the "bottom", and the point corresponding to $(0,1)$ (the only other point in $\mathbb{Z}^2 \cap \Delta$) the "top". If x is a dragonic rational with at most N nonzero digits, then x is the bottom of $\Delta_J(x)$ for all $J > N$. (But note that the "orientation" of $\Delta_J(x)$, as indicated by the line connecting bottom and top, changes with J!). It follows that x is on the border of these $\Delta_J(x)$. If x is not at the edge of Δ itself, then there must exist another little dragon $\Delta_J(y)$ so that x is the top of $\Delta_J(y)$ (since Δ is the union of all the 2^J possible dragons Δ_J). Since the top $(0,1)$ of Δ is given by the expansion $.111111\ldots$, we can therefore find another dragonic expansion for x, ending in all ones, and with the same J first digits as y,

$$d_j^1(x) = d_j^0(y) \quad \text{for } j \le J , \quad d_j^1(x) = 1 \quad \text{for } j > J .$$

We have seen how to obtain the two expansions for a dragonic rational x. We now want to show that if another dragonic rational y is "close" to x, then at least one of its expansions starts with the same digits as one of the expansions for x. Define

$$\rho = \max \{r : B((0,0); r) \subset \Delta \cup (\Delta - (0,1))\} ,$$

where $B(y;\lambda)$ is the open Euclidean ball centered at y with radius λ. Suppose x is a dragonic rational with $d_j^0(x) = 0$ for $j > J$. Take $m > J$, and consider the set

$$B_m = \{y \in \Delta : |y - x| \le \rho\, 2^{-m/2}\} .$$

There are two possibilities: either x is on the border $\partial\Delta$ of Δ, or it isn't. If $x \in \partial\Delta$, then

$$R^{-1}_{d_m^0(x)} \cdots R^{-1}_{d_1^0(x)}(\Delta - (0,1))$$

has no common interior points with Δ, so that $B_m \subset \Delta_m(x)$, and the terminating expansions of all $y \in B_m$ have the same first m digits $d_j^0(x)$, $j = 1, \ldots, m$. If $x \notin \partial\Delta$, then $R^{-1}_{d_m^0(x)} \cdots R^{-1}_{d_1^0(x)} (\Delta - (0,1)) \subset \Delta$; this set is then a little dragon $\Delta_m(z)$ of which x is the top. In this case $B_m \subset \Delta_m(x) \cup \Delta_m(z)$, so that every point $y \in B_m$ has a dragonic expansion with either the same first m digits as $d^0(x)$ (if $y \in \Delta_m(x)$) or as $d^1(x)$ (if $y \in \Delta_m(z)$). This is the main ingredient needed to make the proof of Theorem 2.3, as sketched in the Appendix in [DL], work in the present case.

One other point that needs checking is whether the existence of two different dragonic expansions for x does not lead to inconsistencies for the definition of $v(x)$. If $d_j^0(x) = 0$ for $j > J$, then $d^0(x)$, $d^1(x)$ are linked by

$$x = \sum_{j=1}^{N} d_j^0(x)\, R^{-j}(1,0) = \sum_{j=1}^{N} d_j^1(x)\, R^{-j}(1,0) + R^{-N}(0,1)$$

for $N \geq J$ arbitrary. One can then compute $v(x)$ in two ways, using the two expansions. The following computation shows that they lead to the same result: for $k \in \Gamma$,

$$\begin{aligned}
&\left[v\left(\sum_{j=1}^{N} d_j^0(x) R^{-j}(1,0) \right) \right]_k = \left[T_{d_1^0(x)} \cdots T_{d_n^0(x)} v(0,0) \right]_k \\
&= \sum_{j_1 \ldots j_N} c_{Rk + d_1^0(x)(1,0) - j_1} c_{Rj_1 + d_2^0(x)(1,0) - j_2} \cdots c_{Rj_N + d_N^0(x)(1,0) - j_N} \\
&\qquad \cdot [v(0,0)]_{j_N} \\
&= \sum_{m_1 \ldots m_N} c_{Rk + d_1^1(x)(1,0) - m_1} \cdots c_{Rm_{N-1} + d_n^1(x)(1,0) - m_N} \\
&\qquad \cdot [v(0,0)]_{m_N + \sum_{k=1}^{N} d_k^0(x) R^{N-k}(1,0) - \sum_{k=1}^{N} d_k^1(x) R^{N-k}(1,0)} \\
&= \sum_{m_1 \ldots m_N} c_{Rk + d_1^1(x)(1,0) - m_1} \cdots c_{Rm_{N-1} + d_n^1(x)(1,0) - m_N} \, [v(0,0)]_{m_N + (0,1)} \\
&= \left[T_{d_1^1(x)} \cdots T_{d_N^1(x)} \, v(0,1) \right]_k .
\end{aligned}$$

The reader can now check that the proof in [DL] indeed carries over to prove Theorem B.1. Similarly, one can prove differentiability of ϕ under stronger conditions on T_0, T_1, similar to Theorem 3.1

in [DL]. Finally, the same techniques can also be used for local regularity estimates, but these are a bit more tricky, and require further study of the properties of dragonic expansions. In practice, the matrices $T_0|_{E_1}$, $T_1|_{E_1}$ are often too large to permit a rigorous estimate of λ in (B.6). However, λ is bounded below by the quantities $\rho(T_{d_1} \cdots T_{d_m}|_{E_1})^{1/m}$, and this leads to upper bounds for the Hölder exponent α.

EXAMPLES.

1. $g(x) = \frac{1}{2}\, g(Rx + (1,0)) + g(Rx) + \frac{1}{2}\, g(Rx - (1,0))$
   The solution to this equation is the convolution $\chi_\Delta * \chi_\Delta$, where χ_Δ is the indicator function of the dragon set Δ (see also the second remark following Proposition 5.2). In this case Γ has 10 elements. The largest spectral radius of $T_d|_{E_1}$ is obtained for $d = 0$, $\rho(T_0|_{E_1}) = .847810\ldots$, corresponding to a lower bound $\lambda \geq \rho(T_0|_{E_1})$ in (B.6), or a Hölder exponent $\alpha \leq .47637\ldots$ Via other methods (using the transition operator T of (5.19)) one also derives that this value is a lower bound. This global Hölder exponent is attained in dragonic rationals, in particular in $(0,0)$.
   Note that when M_0 is positive, as in this case, the transition operator T is already known to give optimal results. One easily checks that the matrix representing T is in fact a submatrix of T_0, so that it is not surprising that they have a common eigenvalue!
2. $\phi(x) = h_0\phi(Rx) + h_1\phi(Rx - (1,0)) + h_2\phi(Rx - (-1,1)) + h_3\phi(Rx - (0,1))$, with $h_0 = .506970418225$, $h_1 = -.207072424345$, $h_2 = .493029581775$, $h_3 = 1.20707242435$. This is an example from the family described at the very end of Section III.3.a. It leads to an orthonormal wavelet basis. In this case $|\Gamma| = 14$; the parameters have been chosen so that $\rho(T_0|_{E_1}) \simeq \rho(T_1|_{E_1}) \simeq .714$. Plots of approximations to ϕ seem to suggest that ϕ might be continuous, but we have no proof. If it is, then its Hölder exponent is bounded above by $\log[\rho(T_0T_1|_{E_1})^{1/2}]/\log\sqrt{2} \simeq \log(.90649)/\log\sqrt{2} \simeq .28327$.

Appendix C. Proof of the inequalities (5.52) for $G(\omega)$.

The function G is defined as

$$G(\omega) = \left[\cos^2\frac{\omega_1}{2} + \cos^2\frac{\omega_2}{2}\right]\left[\sin^2\frac{\omega_1}{2} + \sin^2\frac{\omega_2}{2}\right]$$

$$\cdot \left[\sin^2 \frac{\omega_1 + \omega_2}{2} + \sin^2 \frac{\omega_1 - \omega_2}{2} \right]^{-1} H(\omega) ,$$

with $$H(\omega) = h\left(\frac{1}{2} \left(\sin^2 \frac{\omega_1}{2} + \sin^2 \frac{\omega_2}{2} \right) \right) ,$$

and $$h(t) = \begin{cases} \dfrac{1}{1-t} & 0 \le t \le 1/2 , \\ 4t & 1/2 \le t \le 1 . \end{cases}$$

We want to prove inequalities for $G(\omega)$, $G(\omega)G(D\omega)$ and $G(\omega)G(D\omega)$ $G(D^2\omega)$, where $D(\omega_1, \omega_2)$ is either $(\omega_1 + \omega_2, \omega_1 - \omega_2)$ or $(\omega_1 - \omega_2, \omega_1 + \omega_2)$. (Since G is invariant for the interchange of ω_1, ω_2, it does not matter which definition of D is taken, $D = R$ or $D = S$.) To prove these inequalities it is convenient to use different variables,

$$s = s(\omega) = \frac{1}{2} \left(\sin^2 \frac{\omega_1}{2} + \sin^2 \frac{\omega_2}{2} \right) , \quad p = p(\omega) = \sin^2 \frac{\omega_1}{2} \sin^2 \frac{\omega_2}{2} .$$

We then have

$$G(\omega) = \frac{s(1-s)}{s-p} h(s) = 2\eta(s,p) .$$

Moreover,

$$s(D\omega) = 2(s-p) , \qquad p(D\omega) = 4(s^2 - p) .$$

As ω ranges over $[-\pi, \pi]$, (s,p) fill out the domain Δ defined by

$$\Delta = \{(s,p) : \ 0 \le s \le 1, \ \max\{0, 2s-1\} \le p \le s^2\} .$$

In terms of these new variables, we therefore want to study $\eta(s,p)$, $\eta(s,p)\ \eta(\tilde{D}(s,p))$ and $\eta(s,p)\ \eta(\tilde{D}(s,p))\ \eta(\tilde{D}^2(s,p))$, for all $(s,p) \in \Delta$, where $\tilde{D}$ is defined by

$$\tilde{D}(s,p) = (\tilde{s}, \tilde{p}) = (2(s-p), \ 4(s^2 - p)) .$$

Note that $\tilde{D}$ maps Δ twice onto itself (both $\Delta \cap \{s \le 1/2\}$ and $\Delta \cap \{s \ge 1/2\}$ get mapped to all of Δ). Moreover $\tilde{D}$ has one fixed point, $(s_0, p_0) = (5/8, \ 5/16)$, corresponding to $\eta(s_0, p_0) = 15/16$.

We shall prove that $\Delta = \Delta_1 \cup \Delta_2 \cup \Delta_3$, where

$$\eta(s,p) \le \zeta \quad \text{on } \Delta_1 , \tag{C.1}$$

$$\eta(s,p)\, \eta(\tilde{D}(s,p)) \le \zeta^2 \quad \text{on } \Delta_2 , \tag{C.2}$$

$$\eta(s,p)\, \eta(\tilde{D}(s,p))\, \eta(\tilde{D}^2(s,p)) \le \zeta^3 \quad \text{on } \Delta_3 . \tag{C.3}$$

The value of ζ will be fixed by our estimates below; our goal is to obtain $\zeta < 1$.

Choose $\alpha = \sqrt{.9}$, and define the region Δ_1 by

$$\Delta_1 = \{(s,p) \in \Delta : p \le \left(1 - \frac{1}{2\alpha}\right) s \quad \text{if } s \le 1/2 ,$$
$$p \le s - \frac{2}{\alpha}\, s^2(1-s) \quad \text{if } s \ge 1/2 \} .$$

Since

$$\eta(s,p) = \frac{s}{2(s-p)} \quad \text{if } s \le 1/2 , \quad \frac{2s^2(1-s)}{s-p} \quad \text{if } s \ge 1/2 ,$$

we automatically have

$$\eta(s,p) \le \alpha \text{ on } \Delta_1 . \tag{C.4}$$

By the definition of η and $\tilde{D}$, we have to distinguish four different regions when studying $\eta_2(s,p) = \eta(s,p)\eta(\tilde{D}(s,p))$:

$$\eta_2(s,p) =$$
$$\begin{cases} \dfrac{s}{2(\tilde{s}-\tilde{p})} = \dfrac{s}{4(s-2s^2+p)} & \text{if } s \le 1/2, \quad p \le s - 1/4 , \\ \dfrac{2s\tilde{s}(1-\tilde{s})}{\tilde{s}-\tilde{p}} & \text{if } s \le 1/2, \quad p \ge s - 1/4 , \\ \dfrac{s^2(1-s)}{\tilde{s}-\tilde{p}} = \dfrac{s^2(1-s)}{s-2s^2+p} & \text{if } s \ge 1/2, \quad p \ge s - 1/4 , \\ \dfrac{4s^2(1-s)(1-\tilde{s})}{\tilde{s}-\tilde{p}} = \dfrac{8s^2(1-s)(s-p)(1-2(s-p))}{s+p-2s^2} & \\ \qquad\qquad \text{if } s \ge 1/2, \quad p \le s - 1/4 . \end{cases}$$

We define Δ_2 by

$$\Delta_2 = \left\{(s,p) \in \Delta : \ s \le \frac{1}{2}, \ p \ge \left(1 - \frac{1}{2\alpha}\right) s\right\}$$
$$\bigcup \left\{(s,p) \in \Delta : \ s \ge \frac{1}{2}, \ p \ge 1.8\, s - .81\right\}$$
$$= \Delta_{2,1} \cup \Delta_{2,2} \ .$$

Since $\tilde{s}(1-\tilde{s}) \le 1/4$ for all $\tilde{s} \in [0,1]$, we have

$$\eta_2(s,p) \le \frac{s}{2(\tilde{s} - \tilde{p})} \ = \ \frac{s}{4(s - 2s^2 + p)}$$

on all of $\Delta_{2,1}$. Since moreover $p \ge (1 - 1/2\alpha)\, s$, we have

$$\eta_2(s,p) \le \left[4\left(2 - \frac{1}{2\alpha} - 2s\right)\right]^{-1} \le \left[4\left(1 - \frac{1}{2\alpha}\right)\right]^{-1} < \alpha^2$$

on $\Delta_{2,1}$.

On $\Delta_{2,2} \cap \{(s,p) \in \Delta : \ p \ge s - 1/4\}$, one easily checks that

$$\eta_2(s,p) \ = \ \frac{s^2(1-s)}{s - 2s^2 + p}$$

satisfies $\partial_p \eta_2 \neq 0$ everywhere. It follows that η_2 achieves its maximum on the boundary of this domain, given by the three pieces $p = s^2$, with $1/2 \le s \le .9$, $p \ = \ s - 1/4$ with $1/2 \le s \le .7$, and $p \ = \ 2\alpha s - \alpha^2$ with $.7 \le s \le .9$. One easily checks that the maximum value of η_2 on this boundary is .9.

Similarly one checks that η_2 achieves its maximum on $\Delta_{2,2} \cap \{(s,p) \in \Delta : \ p \le s - 1/4\}$ on the boundary of this set; again this leads to $\eta_2 \le .9$.

It follows that

$$\eta_2(s,p) \le .9 \ = \ \alpha^2 \text{ on all of } \Delta_2 \ . \tag{C.5}$$

It remains to determine an upper bound on $\eta_3(s,p) \ = \ \eta(s,p)\eta(\tilde{D}(s,p))$ $\eta(\tilde{D}^2(s,p))$ on $\Delta \setminus (\Delta_1 \cup \Delta_2) \ = \ \{(s,p) : \ 2s - 1 \le p \le s^2, \ p \ge s - 2\, s^2)(1-s)/\alpha, \ p \le 1.8\, s - .81\}$ Since $s - 2\, s^2(1-s)/\alpha$ is strictly increasing, we have $\Delta \setminus (\Delta_1 \cup \Delta_2) \subset \Delta_3 \ = \ \{(s,p) : \ 2s - 1 \le p \le s^2, \ p_1 \ = \ 1.8\, s_1 - .81 \le p \le 1.8\, s - .81\}$, where s_1 is the solution

to $s - 2s^2(1-s)/\alpha = 1.8s - .81$. In Δ_3 one has to distinguish 4 subdomains, corresponding to different expressions for η_3, namely $\Delta_{3,1} = \Delta_3 \cap \{p \geq p_1,\ p \geq 2s-1,\ p \leq 2(s-1/4)^2\}$, $\Delta_{3,2} = \Delta_3 \cap \{p \geq 2(s-1/4)^2,\ p \leq s-1/4\}$, $\Delta_{3,3} = \Delta_3 \cap \{p \geq s-1/4,\ p \geq 2(s-1/4)^2\}$ and $\Delta_{3,4} = \Delta_3 \cap \{p \geq s-1/4,\ p \leq s(s-1/4)^2\}$. On $\Delta_{3,1}$, $\Delta_{3,3}$ and $\Delta_{3,4}$ one checks explicitly that $\partial_p \eta_3 \neq 0$. On $\Delta_{3,2}$, the exact expression for η_3 is too complicated, but one can replace it by an upper bound,

$$\begin{aligned}\eta_3(s,p) &= \frac{2s^2(1-s)}{s-p}\ \frac{2\tilde{s}^2(1-\tilde{s})}{\tilde{s}-\tilde{p}}\ \frac{2\tilde{\tilde{s}}^2(1-\tilde{\tilde{s}})}{\tilde{\tilde{s}}-\tilde{p}} \\ &\leq \frac{2s^2(1-s)}{s-p}\ \frac{\tilde{s}}{\tilde{s}-\tilde{p}}\ \frac{2\tilde{\tilde{s}}^2(1-\tilde{\tilde{s}})}{\tilde{\tilde{s}}-\tilde{p}} \\ &= \frac{16\,s^2(1-s)\tilde{\tilde{s}}(1-\tilde{\tilde{s}})}{\tilde{\tilde{s}}-\tilde{p}} = \overline{\eta}_3(s,p)\ .\end{aligned}$$

This upper bound again satisfies $\partial_p \overline{\eta}_3 \neq 0$ on $\Delta_{3,2}$. It follows that η_3 on Δ_3 is bounded by the maximum of η_3 on the boundaries of $\Delta_{3,1}$, $\Delta_{3,3}$, $\Delta_{3,4}$ and of $\overline{\eta}_3$ on the boundary of $\Delta_{3,2}$. Explicitly, for all $(s,p) \in \Delta_3$,

$$\eta_3(s,p) \leq \overline{\eta}_3(s_1,p_1) = .88145650226\ldots \tag{C.6}$$

This numerical upper bound is larger than $(.9)^{3/2}$; it follows therefore from (C.4) and (C.5) that we have proved (C.1)-(C.2) for

$$\zeta = [\overline{\eta}_3(s_1,p_1)]^{1/3} = .958812370442\ldots$$

Acknowledgments. The authors are grateful to K. Gröchenig, W. Madych and W. Lawton for introducing them to fractal tilings and several related problems. They are also indebted to J. Kovačević and M. Vetterli for fruitful discussions and exchange of ideas.

References.

[AB] Adelson, E. and Burt, P., The Laplacian Pyramid as a compact image code. *IEEE Trans. Comm.* **31** (1983), 482-540.

[ASH] Adelson, E., Simoncelli E. and Hingorani R., Orthogonal pyramid transform for image coding. *SPIE* **845** (1987), 50-58.

[CC] Cohen, A. and Conze, J. P., Régularité des bases d'ondelettes et mesures ergodiques. *Revista Mat. Iberoamericana* **8** (1992), 351-366.

[CD] Cohen, A. and Daubechies, I., A stability criterion for biorthogonal wavelet bases and their related subband coding schemes. *Duke Math. J.* **68** (1992), 313-335.

[CDF] Cohen, A., Daubechies, I. and Feauveau, J. C., Biorthogonal bases of compactly supported wavelets. To appear in *Comm. Pure Appl. Math.* (1991).

[CDM] Caveretta, A., Dahmen W. and Micchelli, C., Stationary Subdivision. *Mem. Amer. Math. Soc.* **93** (1991), 1-186.

[CR] Conze, J. P. and Raugi, A., Fonction Harmonique pour un operateur de transition et application. *Bull. Soc. Math. France* **118** (1990), 273-310.

[Co1] Cohen, A., Ondelettes, analyses multiresolutions et filtres miroirs en quadrature. *Ann. Inst. H. Poincaré, Analyse non linéaire* **7** (1990), 439-459.

[Co2] Cohen, A., Construction de bases d'ondelettes α-Hölderiennes. *Revista Mat. Iberoamericana* **6** (1990), 91-108.

[Con] Conze, J. P., Sur le calcul de la norme de Sobolev des fonctions d'échelles. Preprint, Dept. de Math., Université de Rennes (1990).

[Dau1] Daubechies, I., Orthonormal bases of compactly supported wavelets. *Comm. Pure Appl. Math.* **41** (1989), 909-996.

[Dau2] Daubechies, I., *Ten Lectures on Wavelets.* CBMS Lecture notes **61** SIAM, 1992.

[Dau3] Daubechies, I., Orthonormal bases of compactly supported wavelets. Part II & III: variation on a theme. *SIAM J. Math. Anal.* **24** (1993), to appear.

[DD] Deslauriers, G. and Dubuc, S., *Interpolation dyadique. Fractals, dimensions non entieres et applications*, Masson (1987), 44-45.

[DL] Daubechies, I. and Lagarias, J., Two scale difference equations. Part I & II. *SIAM J. Math. Anal.* **22** (1991), 1388-1410 & **23** (1992), 1031-1079.

[DyL] Dyn, N. and Levin, D., Interpolating subdivision schemes for the generation of curves and surfaces, in *Multivariate Interpolation and Approximation*, W. Haussmann and K. Jeller, eds., Birkhaüser (1990), 91-106.

[Fea] Feauveau, J. C., Analyse multirésolution par ondelettes non orthogonale et benes de filtres numériques. PhD. Thesis, Université de Paris Sud (1990).

[FS] Fix, G., and Strang G., A Fourier analysis of the finite element method, in *Ritz-Galerkin theory, Stud. Appl. Math.* **48** (1969), 265-273.

[K] Knuth, D., *The art of computer programming,* II. Addison Wesley, 1968.

[KV] Kovačević, J. and Vetterli, M., Non separable multidimensional perfect

reconstruction filter banks and wavelet bases for $\mathbb{R}^n$. Preprint Columbia Univ. (1991).

[Le] Lemarié, P. G., Ondelettes à localisation exponentielle. *J. Math. Pures et Appl.* **67** (1988), 227-236.

[LR] Lawton, W. and Resnikoff, M., Multidimensional wavelet bases. Preprint AWARE (1990).

[M] Marr, D., *Vision.* Freeman & Co., 1982.

[Ma1] Mallat, S., Multiresolution approximation and wavelets orthonormal bases of $L^2(\mathbb{R})$. *Trans. Amer. Math. Soc.* **315** (1989), 69-87.

[Ma2] Mallat, S., A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. PAMI* **11** (1989), 674-693.

[Me1] Meyer, Y., *Ondelettes et Opérateurs.* Hermann, 1990.

[Me2] Meyer, Y., Ondelettes, fonctions splines et analyses graduées. CEREMADE Report 8703 (1987).

[MG] Madych, W. and Gröchenig, G., Multiresolution analysis, Haar bases and self similar tilings of $\mathbb{R}^n$. *IEEE Trans. Inf. Th.* **38** (1992), 556-568.

[Mo] Mongeau, J. P., Proprietés de l'interpolation iterative. PhD. Thesis, Université de Montréal (1990).

[Ri] Rioul, O., Simple regularity criteria for subdivision schemes. *SIAM J. Math. Anal.* **23** (1992), 1544-1576.

[SB1] Smith, M. J. and Barnwell, T. P., Exact reconstruction techniques for tree structured subband coders. *IEEE ASSP* **34** (1986), 434-441.

[SB2] Smith, M. J. and Barnwell, T. P., A new filter bank theory for time frequency representation. *IEEE ASSP* **35** (1987), 314-326.

[V] Volkner, M., On the regularity of wavelets. *IEEE Trans. Inf. Th.* **38** (1992), 872-876.

[Ve] Vetterli, M., Filter bank allowing perfect reconstruction. *Signal Processing* **10** (1986), 219-244.

Recibido: 12 de diciembre de 1.991

A. Cohen and I. Daubechies
AT&T Bell Laboratories
New Jersey 07974, U.S.A.

Revista Matemática Iberoamericana
Vol. 9, N.° 1, 1993

Exceptional modular form of weight 4 on an exceptional domain contained in $\mathbb{C}^{27}$

Henry H. Kim

Abstract. Resnikoff [12] proved that weights of a non trivial singular modular form should be integral multiples of 1/2, 1, 2, 4 for the Siegel, Hermitian, quaternion and exceptional cases, respectively. The θ-functions in the Siegel, Hermitian and quaternion cases provide examples of singular modular forms (Krieg [10]). Shimura [15] obtained a modular form of half-integral weight by analytically continuing an Eisenstein series. Bump and Baily suggested the possibility of applying an analogue of Shimura's method to obtain singular modular forms, *i.e.* modular forms of weight 4 and 8, on the exceptional domain of 3×3 hermitian matrices over Cayley numbers. The idea is to use Fourier expansion of a non-holomorphic Eisenstein series defined by using the factor of automorphy as in Karel [7]. The Fourier coefficients are the product of confluent hypergeometric functions as in Nagaoka [11] and certain singular series which we calculate by the method of Karel [6]. In this note we describe a modular form of weight 4 which may be viewed as an analogue of a θ zero-value and as an application, we consider its Mellin transform and prove a functional equation of the Eisenstein series which is a Nagaoka's conjecture (Nagaoka [11]).

Introduction.

As it is well-known, the classical θ-functions are modular forms of half-integral weight with respect to a congruence subgroup. They provide examples of modular forms of singular weights (or critical weights). Resnikoff [12], using differential operators, proved an interesting theorem that weights of a non trivial singular modular form should be some integral mutiples of $1/2, 1, 2, 4$ for the Siegel, Hermitian, quaternion and exceptional cases, respectively. Later he proved the converse. Krieg [10] considered θ-functions in the Hermitian and quaternion cases and showed that they are the modular forms in Resnikoff's Theorem. However, in the exceptional case, one does not know how to construct θ-functions. It has been a long-standing problem since Baily [1] initiated the study of automorphic forms on the exceptional domain, to construct a "θ-function" on the exceptional domain, that is, to construct modular forms of weight 4 and 8, if they exist.

Raghavan considered a non-holomorphic Eisenstein series of degree 3 in the Siegel case. He obtained a modular form of weight 4 by analytic continuation and showed that it is a θ-function. Shimura [15] actually obtained a modular form of half-integral weight by analytically continuing an Eisenstein series. Daniel Bump and Walter Baily suggested the possibility of applying an analogue of Shimura's method to obtain singular modular forms, *i.e.* modular forms of weight 4 and 8, on the exceptional domain. The idea is to use the Fourier expansion of a non-holomorphic Eisenstein series. We use a factor of automorphy similar to that defined by Karel [7] to define a non-holomorphic Eisenstein series; we thank him for his ideas which he shared in a private conversation. The Fourier coefficients are the product of confluent hypergeometric functions and certain singular series which are called Siegel series or Whittaker functions. The confluent hypergeometric functions were studied by Nagaoka [11]. The singular series for the full rank case were explicitly calculated by Karel [6] who used it to prove that there is a common denominator for the Fourier coefficients of a holomorphic Eisenstein series defined by Baily [1]. We calculate the singular series for the singular cases by modifying Karel's method. Thus we calculate the Fourier coefficients completely. Using an estimate on confluent hypergeometric functions, the Eisenstein series can be continued as meromorphic functions to the whole complex plane and especially by considering $s \longrightarrow 0$, we obtain an interesting result that we have a holomorphic modular form of weight k unless $k = 2, 6, 10$. Thus we obtain modular forms of weight 4 and 8. Nagaoka [11] made

a conjecture on a functional equation of an Eisenstein series similar to our Eisenstein series. The constant term of the Fourier expansion of the Eisenstein series contains an "Epstein zeta function". We consider the Mellin transform of the modular form of weight 4 just like a θ-function (Riemann's trick) to get a functional equation of the "Epstein zeta function" and of the Eisenstein series.

It is noted that there is no modular form of the lowest weight $1/2$, $1, 2$ with respect to the full modular group in the Siegel, Hermitian and quaternion cases, respectively. Non-zero modular forms of these weights exist only with respect to congruence subgroups. But in our exceptional case, the modular form of lowest weight is a modular form with respect to the full modular group.

1. An exceptional modular group and its Eisenstein series.

We recall several definitions of an exceptional group of type E_7 (of index -25) and a tube domain from Baily [1].

(i) *Integral Cayley numbers and exceptional Jordan algebras.* Let $\mathfrak{C}$ be the Cayley algebra and $\mathfrak{o}$ be the integral Cayley numbers. Let $\mathfrak{o}_p = \mathfrak{o} \otimes_{\mathbb{Z}} \mathbb{Z}_p$, $\mathfrak{C}_p = \mathfrak{o} \otimes_{\mathbb{Z}} \mathbb{Q}_p$.

We define an exceptional Jordan algebra $\mathfrak{J}$ as the vector space of matrices $X = (x_{ij})$, $x_{ij} \in \mathfrak{C}$ satisfying $x_{ij} = \bar{x}_{ji}$, suppied with the product

$$X \circ Y = \frac{1}{2}(XY + YX),$$

where XY is the ordinary matrix product. Let e_{ij} be the 3×3 matrix with a 1 in the intersection of the i-th row and j-th column and zeros elsewhere, and let $e_i = e_{ii}$, $i = 1, 2, 3$.

If $X \in \mathfrak{J}$, then we shall write

$$X = \begin{pmatrix} a & x & y \\ \bar{x} & b & z \\ \bar{y} & \bar{z} & c \end{pmatrix}.$$

Define

$$\mathrm{tr}(X) = a + b + c$$

and an inner product $(\cdot, \cdot)$ on $\mathfrak{J}$ by

$$(X, Y) = \mathrm{tr}(X \circ Y).$$

Moreover we define

$$\det(X) = abc - aN(z) - bN(y) - cN(x) + \mathrm{tr}((xz)\bar{y}),$$

and define a symmetric trilinear form $(\cdot,\cdot,\cdot)$ on $\mathfrak{J}\times\mathfrak{J}\times\mathfrak{J}$ by letting

$$(X,X,X) = \det(X).$$

Then we define a bilinear mapping $(X,Y)\longrightarrow X\times Y$ of $\mathfrak{J}\times\mathfrak{J}$ into $\mathfrak{J}$ by requiring the identity

$$(X\times Y, Z) = 3\,(X,Y,Z)$$

to hold. We have

$$X\times X = \begin{pmatrix} bc-N(z) & y\bar{z}-cx & xz-by \\ z\bar{y}-c\bar{x} & ac-N(y) & \bar{x}y-az \\ \bar{z}\bar{x}-b\bar{y} & \bar{y}x-a\bar{z} & ab-N(x) \end{pmatrix}. \tag{1.1}$$

Since $X\circ(X\times X) = (\det X)\,\varepsilon$ for any $X\in\mathfrak{J}$, where $\varepsilon = D(1,1,1)$, $X\times X$ plays the role of a matrix adjoint for $X\in\mathfrak{J}$.

We define

$$\begin{aligned} \mathfrak{K}_3 &= \{X\in\mathfrak{J}:\ \det X\neq 0\}, \\ \mathfrak{K}_2 &= \{X\in\mathfrak{J}:\ \det X = 0,\ X\times X\neq 0\}, \\ \mathfrak{K}_1 &= \{X\in\mathfrak{J}:\ X\times X = 0,\ X\neq 0\}, \\ \mathfrak{K}_0 &= \{0\}. \end{aligned}$$

Then $\mathfrak{J}$ is the disjoint union of these four sets. Finally we denote by $\mathfrak{K}$ the set of squares of elements of $\mathfrak{J}$ and put $\mathfrak{K}_i^+ = \mathfrak{K}_i\cap\mathfrak{K}$, $i=1,2,3$. Then $\mathfrak{K}_i^+$ is a cone and $\mathfrak{K}_3^+$ is open and convex in $\mathfrak{J}$. We let

$$\mathfrak{J}_{\mathfrak{o}} = \{X\in\mathfrak{J}:\ x_{ij}\in\mathfrak{o},\ i,j=1,2,3\};$$

this lattice is self-dual with respect to the inner product $(\cdot,\cdot)$. Let $\mathfrak{J}(j) = \mathfrak{J}^{(j)} = \{X=(x_{ii'})\in\mathfrak{J}:\ x_{ii'}=0$ unless both $i,i'\leq j\}$, and $\Lambda(j) = \Lambda_j = \mathfrak{J}(j)\cap\mathfrak{J}_{\mathfrak{o}}$. We identify $\mathfrak{J}^{(j)}$ with $j\times j$ hermitian matrices over Cayley numbers.

(ii) *An exceptional group of type E_7.* Define two subgroups of $GL(\mathfrak{J})$ as follows:

$$\begin{aligned} \mathcal{S} &= \{g:\quad g\in GL(\mathfrak{J}),\ \det(gX)\equiv\nu(g)\det(X),\ \nu(g)\neq 0\}, \\ \mathcal{J} &= \{g\in\mathcal{S}:\quad \nu(g)=1\}. \end{aligned} \tag{1.2}$$

We define $g^* \in \mathfrak{S}$ so that $(gX, g^*Y) \equiv (X, Y)$. Let

$$\mathfrak{T} = \{Z = X + iY \in \mathfrak{J}_{\mathbb{C}} : \ X \in \mathfrak{J}_{\mathbb{R}}, Y \in \mathfrak{K}_3^+\}.$$

The group of holomorphic automorphisms of the domain $\mathfrak{T}$ has been described by Freudenthal [5] as follows (*cf.* Baily [1]). Let V and V' be two real vector spaces, each isomorphic to $\mathfrak{J}$, and let Ξ and Ξ' be copies of $\mathbb{R}$. Let $\mathbb{W} = V \oplus \Xi \oplus V \oplus \Xi'$. Define a quartic form Q on $\mathbb{W}_{\mathbb{C}}$ by

$$Q(w) = (X \times X, X' \times X') - \xi \det X - \xi' \det X' - \frac{1}{4}((X, X') - \xi\xi')^2,$$

and an alternating form $\{\cdot, \cdot\}$ by

$$\{w_1, w_2\} = (X_1, X_2') - (X_2, X_1') + \xi_1\xi_2' - \xi_2\xi_1',$$

for $w = (X, \xi, X', \xi')$. Then

$$\mathfrak{G} = \{g \in GL(\mathbb{W}) : \ Qg = Q, \ g\{\cdot, \cdot\} = \{\cdot, \cdot\}\}$$

defines a connected algebraic $\mathbb{Q}$-group of type E_7. The group $\mathfrak{S}$ is embedded in $\mathfrak{G}$ by operating on $\mathbb{W}$

$$k(X, \xi, X', \xi') = (kX, \nu(k)^{-1}\xi, k^*X', \nu(k)\xi'), \quad k \in \mathfrak{S}. \tag{1.3}$$

A copy, $\mathbb{P}^+$, of the additive group $\mathfrak{J}$ is embedded in $\mathfrak{G}$ by letting p'_B, for $B \in \mathfrak{J}$, be defined by

$$p'_B(X, \xi, X', \xi') = (X_1, \xi_1, X_1', \xi_1'),$$

where

$$\begin{aligned} X_1 &= X + 2B \times X' + \xi B \times B, \quad X_1' = X' + \xi B, \\ \xi_1' &= \xi' + (B, X) + (B \times B, X') + \xi \det B, \quad \xi_1 = \xi. \end{aligned} \tag{1.4}$$

Define $\iota \in \mathfrak{G}$ by

$$\iota(X, \xi, X', \xi') = (-X', -\xi', X, \xi), \tag{1.5}$$

and let $p_B = \iota^{-1} p'_{-B} \iota$.

Let $N_0 = \{g \in \mathfrak{G} : \ g(0,0,0,\xi') = (0,0,0,\xi''), \quad \xi', \xi'' \in \mathbb{R}\}$. Then N_0 is a maximal $\mathbb{Q}$-parabolic subgroup of $\mathfrak{G}$. Define $\iota_{e_i} = p'_{e_i} \cdot p_{-e_i} \cdot p'_{e_i}$,

and ι_J be the product of all ι_j for $j \in J$ if J is a subset of $\{1,2,3\}$ and let $(j) = \{1,\dots,j\}$. Then we have Bruhat decomposition

$$\mathfrak{G}_{\mathbb{Q}} = \bigcup_{j=0}^{3} N_{0\mathbb{Q}} \iota_{(j)} N_{0\mathbb{Q}} \,. \tag{1.6}$$

Let $\mathbb{W}_0$ be the lattice in $\mathbb{W}_{\mathbb{R}}$ given by $\mathbb{W}_0 = V_0 \oplus \mathbb{Z} \oplus V_0' \oplus \mathbb{Z}$, where V_0 and V_0' are identified with the lattice $\mathfrak{J}_0$. Then let

$$\Gamma = \{g \in \mathfrak{G} : \quad g\mathbb{W}_0 = \mathbb{W}_0\} \,.$$

Γ is an arithmetic subgroup of $\mathfrak{G}_{\mathbb{Q}}$ and Γ shares two most important properties with $Sp_n(\mathbb{Z})$, that is, Γ is a unicuspidal and maximal discrete subgroup of Γ. Let $\Gamma_0 = \Gamma \cap N_{0\mathbb{Q}}$. Then by Baily [1, p. 531], Γ_0 is the semi-direct product of $\mathfrak{S}_0$ and $\mathbb{P}_0^+$. (Here $\mathfrak{S}_0 = \{\pm I\}\mathfrak{J}_0$ by Baily [1, p. 524]). We note that $\mathfrak{G}$ has a center $Z_2 = \{\pm I\}$, and $\mathfrak{G}/Z_2 \cong \mathrm{Hol}(\mathfrak{T})$.

(iii) *The group action and Eisenstein series.* The exceptional group $\mathfrak{G}_{\mathbb{R}}$ of type of E_7 (of index -25) acts on the tube domain $\mathfrak{T} = \{Z = X + iY \in \mathfrak{J}_{\mathbb{C}} : \ Y \in \mathfrak{K}_3^+\}$. In Baily [1], the explicit form of the action (defined by Harish-Chandra) is given by showing that for each $g \in \mathfrak{G}_{\mathbb{R}}$ and $Z \in \mathfrak{T}$ there is a unique solution $Z_1 \in \mathfrak{T}$, $A \in \mathfrak{J}_{\mathbb{C}}$ and $k \in \mathfrak{S}$ such that

$$p_Z' \cdot g = p_A k p_{Z_1}' \,. \tag{1.7}$$

Now we put

$$Z \cdot g = Z_1 \,, \tag{1}$$

$$\mathfrak{z}(Z,g) = k \in \mathfrak{S} \,, \quad j(Z,g) = \nu(\mathfrak{z}(Z,g)) \,, \tag{2}$$

where ν is as in (1.2). Then (1) define the action of $\mathfrak{G}_{\mathbb{R}}$ on $\mathfrak{T}$ and $j(Z,g)$ is the canonical factor of automorphy as in Karel [7](*cf.* Borel [3]). Then we have the following properties of $j(Z,\gamma)$:

$$j(Z,p_B') = 1 \qquad \text{for all } B \in \mathfrak{J}_{\mathbb{R}} \,, \tag{i}$$

$$j(Z,\gamma) = 1 \qquad \text{for } \gamma \in \mathfrak{J}_0 \,, \tag{ii}$$

$$j(Z,\iota) = \det(-Z) \,, \tag{iii}$$

$$j(Z,g_1g_2) = j(Z,g_1)j(Z\cdot g_1,g_2) \,. \tag{iv}$$

Moreover, if $J(Z,g)$ is the functional determinant of g at Z, then we have,

(v) $$J(Z,g) = j(Z,g)^{-18}.$$

(i)-(v) can be proved by (1.3), (1.4), (1.5) and (1.6) and Tsao [16, Theorem 5.3].

Let f be a holomorphic function on $\mathfrak{T}$ which for some integer $k > 0$ satisfies

(1.8) $$f(Z \cdot \gamma) = f(Z) j(Z,\gamma)^k, \qquad Z \in \mathfrak{T}, \quad \gamma \in \Gamma.$$

Then f is called a modular form of weight k on $\mathfrak{T}$ with respect to Γ. If we take in particular $\gamma = -I$, then we can see easily that $Z \cdot \gamma = Z$, $j(Z,\gamma) = -1$. Therefore if k is an odd integer, then $f \equiv 0$. So we consider only a modular form of even weight. From (1.8), $f(Z+B) = f(Z)$, $B \in \mathfrak{J}_0$. This implies that f has a Fourier expansion, and since the lattice $\mathfrak{J}_0$ is self-dual, this has the form

$$f(Z) = \sum_{T \in \mathfrak{J}_0} a(T) e^{2\pi i (T,Z)}.$$

Now we define an Eisenstein series as follows

$$E_{k,s}(Z) = \sum_{\gamma \in \Gamma/\Gamma_0} j(Z,\gamma)^{-k} |j(Z,\gamma)|^{-s},$$

where k is even positive integer and $s \in \mathbb{C}$ and $Z = X + iY$, $Y \in \mathfrak{R}_3^+$. Since Γ_0 is a semi-direct product of $\pm\{I\}\mathfrak{J}_0$ and $\mathbb{P}_0^+$, this is well-defined and converges absolutely and uniformly on compact subsets of $\mathfrak{T}$ if $k + \operatorname{Re} s > 18$.

The purpose of this note is to prove the following two theorems.

Theorem A. *$E_{k,s}(Z)$ can be continued as a meromorphic function in s to a whole complex plane and*

1) *$E_{k,s}(Z)$ is finite at $s = 0$ for all k,*
2) *$E_{k,0}(Z)$ is holomorphic in Z unless $k = 2, 6, 10$,*
3) *$E_{k,0}(Z)$ is a modular form of weight k with rational Fourier coefficients unless $k = 2, 6, 10$,*

4) *$E_{4,0}(Z)$ and $E_{8,0}(Z)$ are singular modular forms,*

$$E_{4,0}(Z) = 1 + 240 \sum_{T \in \mathfrak{J}_o^+,\, \text{rank } T=1} \sigma_3(\Delta(T)) e^{2\pi i (T,Z)},$$

where $\sigma_3(t) = \sum_{a|t} a^3$ and $\Delta(T)$ is as in Karel [6, p. 186].

Theorem B. *Let*

$$\Psi(s) = (\det Y)^{s/2} \rho(s)\rho(s-4)\rho(s-8)(s-2)(s-4)E_{0,s}(Z),$$

where $\rho(s) = \pi^{-s/2}\Gamma(s/2)\zeta(s)$. Then $\Psi(s)$ can be continued as a meromorphic function in s to a whole complex plane with a simple pole at $s = 0, 1, 5, 8, 10, 13, 17, 18$ and satisfies a functional equation

$$\Psi(18 - s) = \Psi(s).$$

2. Fourier expansion of the Eisenstein series.

By (1.6), we have a relative Bruhat decomposition

$$\mathcal{G}_{\mathbb{Q}} = \bigcup_{j=0}^{3} N_{0\mathbb{Q}} \iota_{(j)} N_{0\mathbb{Q}}.$$

By Tsao [16], every element of $N_{0\mathbb{Q}}\iota_{(j)}N_{0\mathbb{Q}}$ can be represented by

$$\mu p'_X \iota_{(j)} N_{0\mathbb{Q}}/N_{0\mathbb{Q}}, \qquad \mu \in \mathfrak{I}_o, X \in \mathfrak{J}_{\mathbb{Q}}^{(j)}.$$

Now we show that, for $g \in \mu p'_X \iota_{(j)} N_{0\mathbb{Q}}/N_{0\mathbb{Q}}$,

$$j(g, Z) = \pm \det((Z \cdot \mu)_j + X)\kappa(X), \tag{2.1}$$

where $(Z \cdot \mu)_j$ is the left upper corner $j \times j$ matrix of $Z \cdot \mu$ and $\kappa(X)$ is as in Baily [1, p. 522]. For $g \in \mu p'_X \iota_{(j)} N_{0\mathbb{Q}}/N_{0\mathbb{Q}}$, $g = \mu p'_X \iota_{(j)} p$ for $p \in N_{0\mathbb{Q}}$. Then

$$p'_Z \cdot g = (p'_Z \mu p'_X \iota(j))p = (p_A k p'_{Z_1})k_1 p'_B,$$

where $p = k_1 p'_B$. Therefore we get

$$p'_Z \cdot g = p_A k k_1 p'_{\tilde{Z}_1} .$$

Now by using the formula for $\iota_{(j)}$, we can see that $\nu(k) = \det((Z\cdot\mu)_j + X)$. Also by Tsao [16, p. 266], we have $\nu(k_1) = \kappa(X)$. Therefore we get (2.1).

Therefore we can decompose the Eisenstein series as follows

$$E_{k,s}(Z) = 1 + E^{(1)}_{k,s}(Z) + E^{(2)}_{k,s}(Z) + E^{(3)}_{k,s}(Z),$$

where

$$E^{(j)}_{k,s}(Z) = \sum_{\mu\iota_{(j)} \in \mathfrak{I}_0 \iota_{(j)} N_{0\mathbb{Q}}/N_{0\mathbb{Q}}} \sum_{X \in \mathfrak{I}^{(j)}_{\mathbb{Q}}/\Lambda_j} \kappa(X)^{-(k+s)} S_j((Z\cdot\mu)_j + X; k + \frac{s}{2}, \frac{s}{2}),$$

$$S_j(z;\alpha,\beta) = \sum_{a\in\Lambda_j} \det(z+a)^{-\alpha} \det(\bar{z}+a)^{-\beta}.$$

As we will see in Section 3, $S_j(z;\alpha,\beta)$ has a Fourier expansion

$$\mu(\mathfrak{I}^{(j)}_{\mathbb{R}}/\Lambda_j) S_j(z;\alpha,\beta) = \sum_{h\in\Lambda_j} e^{2\pi i(h,x)} \xi_j(y,h;\alpha,\beta).$$

Therefore we have the Fourier expansion of $E^{(j)}_{k,s}$:

$$\begin{aligned} E^{(j)}_{k,s}(Z) &= \sum_{\mu\iota_{(j)} \in \mathfrak{I}_0 \iota_{(j)} N_{0\mathbb{Q}}/N_{0\mathbb{Q}}} \sum_{T\in\Lambda_j} \frac{1}{\mu_j} \xi_j((\mu^{*-1}Y)_j, T; k + \frac{s}{2}, \frac{s}{2}) \\ &\qquad \cdot S(T, k+s) e^{2\pi i (T, (\mu^{*-1}X)_j)} \\ &= \sum_{T\in\Lambda_j} \sum_{\mu\iota_{(j)} \in \mathfrak{I}_0 \iota(j) N_{0\mathbb{Q}}/N_{0\mathbb{Q}}} a(T,Y,s) e^{2\pi i (T, (\mu^{*-1}X)_j)}, \end{aligned} \tag{2.2}$$

where $Z = X + iY$ and

$$a(T,Y,s) = \frac{1}{\mu_j} \xi_j((\mu^{*-1}Y)_j, T; k + \frac{s}{2}, \frac{s}{2}) S(T, k+s),$$

$$\mu_j = \mu(\mathfrak{I}^{(j)}_{\mathbb{R}}/\Lambda_j),$$

$$S(T, k+s) = \sum_{X \in \mathfrak{I}^{(j)}_{\mathbb{Q}}/\Lambda_j} \kappa(X)^{-(k+s)} e^{2\pi i (T,X)}.$$

3. Confluent hypergeometric function.

Consider the infinite series

$$S_m(z, L_m : \alpha, \beta) = \sum_{a \in L_m} \det(z + a)^{-\alpha} \det(\bar{z} + a)^{-\beta},$$

where $z \in H_m$, $z = x + iy$, $y \in \mathfrak{K}_m^+$, L_m is a lattice in $\mathfrak{J}_{\mathbb{R}}^{(m)}$. This has a Fourier expansion

$$\mu(\mathfrak{J}_{\mathbb{R}}^{(m)}/L_m) S_m(z, L_m : \alpha, \beta) = \sum_{h \in L_m'} e^{2\pi i (h,x)} \xi_m(y, h : \alpha, \beta),$$

where L_m' is the dual lattice of L_m with respect to $(\cdot, \cdot)$,

$$(x, y) = \frac{1}{2} \operatorname{tr}(xy + yx)$$

and $\mu(\mathfrak{J}_{\mathbb{R}}^{(m)}/L_m)$ is the measure of $\mathfrak{J}_{\mathbb{R}}^{(m)}/L_m$ and

$$\xi_m(g, h : \alpha, \beta) = \int_{\mathfrak{J}_{\mathbb{R}}^{(m)}} e^{-2\pi i (h,x)} \det(x + ig)^{-\alpha} \det(x - ig)^{-\beta} \, dx,$$

where $g \in \mathfrak{K}_m^+$. Consider the function

$$\eta_m(g, h : \alpha, \beta) = \int_{Q(h)} e^{-(g,x)} \det(x + h)^{\alpha - \kappa(m)} \det(x - h)^{\beta - \kappa(m)} \, dx,$$

where $Q(h) = \{x \in \mathfrak{J}_{\mathbb{R}}^{(m)} : \ x \pm h > 0\}$, $\kappa(m) = 4m - 3$.

S. Nagaoka [11] defined the following function ω_m and gave a theorem on its analytic continuation and functional equation but did not publish a proof. In this section we prove his assertion and get the additional results which are necessary for the analytic continuation of the Eisenstein series.

We denote by $V(p, q, r)$ the subset of $\mathfrak{J}_{\mathbb{R}}^{(m)}$ consisting of the elements with p positive, q negative, and r zero eigenvalues ($p + q + r = m$). The precise definition of eigenvalues is as follows. When $m = 3$, the eigenvalues of an element $h \in \mathfrak{J}_{\mathbb{R}}^{(3)}$ are defined as the roots of a cubic equation

$$t^3 - \operatorname{tr}(h) t^2 + \operatorname{tr}(h \times h) t - \det(h) = (t\varepsilon - h, t\varepsilon - h, t\varepsilon - h) = 0,$$

where $x \times y$ denotes the crossed product of $x, y \in \mathfrak{J}_{\mathbb{R}}^{(3)}$ (*cf.* Baily [1, p. 516]). In case $m = 2$, as in Shimura [13], we define the eigenvalues of $h \in \mathfrak{J}_{\mathbb{R}}^{(2)}$ to be the roots of a quadratic equation $t^2 - \mathrm{tr}(h)t + \det(h) = 0$. Moreover, as in Shimura [13], we introduce the notion of the eigenvalues of h *relative to* g for $h \in \mathfrak{J}_{\mathbb{R}}^{(m)}$ and $g \in \mathfrak{R}_m^+$. In case of $m = 3$, we define them to be the roots of an equation

$$t^3 - (g,h)t^2 + (g \times g, h \times h)t - \det(g)\det(h) = 0\,.$$

When $m = 2$, they are defined as the roots of an equation $t^2 - (g,h)t + \det(g)\det(h) = 0$. Now we denote by $\delta_+(hg)$ (respectively, $\tau_+(hg)$) the product (respectively, the sum) of all positive eigenvalues of h relative to g. Moreover, we put

$$\delta_-(hg) = \delta_+((-h)g)\,, \quad \tau_-(hg) = \tau_+((-h)g)$$

and

$$\tau(hg) = \tau_+(hg) + \tau_-(hg)\,.$$

We also denote by $\mu(hg)$ the smallest absolute value of non zero eigenvalues of h relative to g if $h \neq 0$; $\mu(hg) = 1$ if $h = 0$. We define, as in Baily [1, p. 520], $a^* \in \mathfrak{S}$ so that $(ax, a^*y) = (x,y)$ for all x and y. Then from the definitions , we can see easily that the above quantities are invariant under the map $(g,h) \mapsto (a^*g, ah)$ for all $a \in \mathfrak{S}$.

Now we can show the following facts as in Shimura [13, sections 1 and 2].

1) $$\eta_m(g, 0 : \alpha, \beta) = \Gamma_m(\alpha + \beta - \kappa(m))\det g^{\kappa(m)-\alpha-\beta}\,,$$

where $$\Gamma_m(s) = \pi^{2m(m-1)} \prod_{n=0}^{m-1} \Gamma(s - 4n) = \int_{\mathfrak{R}_m^+} e^{-\mathrm{tr}x} \det x^{s-\kappa(m)}\, dx\,.$$

2) $$\eta_m(g, \varepsilon : \alpha, \beta) = e^{-(g,\varepsilon)} 2^{m(\alpha+\beta-\kappa(m))} \zeta_m(2g : \alpha, \beta)\,,$$

where $$\zeta_m(g : \alpha, \beta) = \int_{\mathfrak{R}_m^+} e^{-(g,u)} \det(u + \varepsilon)^{\alpha-\kappa(m)} \det u^{\beta-\kappa(m)}\, du\,.$$

3)

$$\xi_m(g, h : \alpha, \beta) = |\sigma_m|^{-1}\, i^{m(\beta-\alpha)}\, 2^{m\kappa(m)}\, \pi^{m(\alpha+\beta)} \cdot \Gamma_m(\alpha)^{-1}\Gamma_m(\beta)^{-1}\eta_m(2\pi g, h : \alpha, \beta)\,.$$

4) If $g \in \mathfrak{R}_m^+$, $h \in V(p,q,r)$, $p+q+r=m$, then there exists $a \in \mathfrak{S}$ such that a^*g is diagonal and $a\,h = \mathrm{diag}(1_p, -1_q, 0_r)$.

5) $$d(ax) = \nu(a)^{\kappa(m)}dx \qquad \text{for } a \in \mathfrak{S}, \quad \nu(a^*) = \nu(a)^{-1},$$

6) $$\eta_m(g, -h : \alpha, \beta) = \eta_m(g, h : \alpha, \beta).$$

7) Let $\eta_m^*(g,h:\alpha,\beta) = \det g^{\alpha+\beta-\kappa(m)}\eta_m(g,h:\alpha,\beta)$. Then for all $a \in \mathfrak{S}$,

$$\eta_m^*(g,h:\alpha,\beta) = \eta_m^*(a^*g, ah : \alpha, \beta).$$

Define for $g \in \mathfrak{R}_m^+$, $h \in V(p,q,r)$,

$$\begin{aligned}\omega_m(g,h:\alpha,\beta) = {} & 2^{-p\alpha-q\beta}\,\Gamma_p(\beta-4(m-p))^{-1} \\ & \cdot \Gamma_q(\alpha-4(m-q))^{-1}\Gamma_r(\alpha+\beta-\kappa(m))^{-1} \\ & \cdot \delta_+(hg)^{\kappa(m)-\alpha-2q}\delta_-(hg)^{\kappa(m)-\beta-2p} \\ & \cdot \eta_m^*(g,h:\alpha,\beta).\end{aligned} \tag{3.1}$$

Theorem. *The function ω_m can be continued as a holomorphic function in (α,β) to the whole $\mathbb{C}^2$ and satisfies the functional equation*

$$\omega_m(g,h:\alpha,\beta) = \omega_m(g,h:\kappa(m)+4r-\beta, \kappa(m)+4r-\alpha). \tag{3.2}$$

Moreover, for every compact set $T \subset \mathbb{C}^2$, there exist two positive constants A,B depending only on T such that

$$|\omega_m(g,h:\alpha,\beta)| \le A\,e^{-\tau(hg)}(1+\mu(hg)^{-B}) \tag{3.3}$$

for all $(g,h) \in \mathfrak{R}_m^+ \times \mathfrak{J}_{\mathbb{R}}^{(m)}$ and $(\alpha,\beta) \in T$.

PROOF. *Case* 1. $m=2$.

We note that H_2 is a domain of type IV as in Shimura [13]. Because of (4), (6) and (7), it is enough to consider the cases

$$h = \begin{pmatrix}1&0\\0&0\end{pmatrix}, \begin{pmatrix}1&0\\0&1\end{pmatrix}, \begin{pmatrix}1&0\\0&-1\end{pmatrix}, \quad g = \begin{pmatrix}\lambda_1&0\\0&\lambda_2\end{pmatrix}$$

where $0 < \lambda_1 \le \lambda_2$.

(i) $h = \begin{pmatrix}1&0\\0&0\end{pmatrix}$. Then

$$\eta_2(g,h:\alpha,\beta) = \int_{x\pm h \in \mathfrak{R}_2^+} e^{-(g,x)}\det(x+h)^{\alpha-\kappa(2)}\det(x-h)^{\beta-\kappa(2)}\,dx.$$

By Shimura [13], the equation preceding (4.29) with $2\lambda_1 = a$, $2\lambda_2 = b$ and $\kappa(2) = 5$,

$$\eta_2(g, h : \alpha, \beta) = \pi^4 \, 2^{2\alpha+2\beta-10} e^{-\lambda_1} (2\lambda_1)^{-4} (2\lambda_2)^{5-\alpha-\beta} \cdot \Gamma(\alpha+\beta-5) \, \zeta_1(2\lambda_1 : \alpha-4, \beta-4) \, .$$

So we get

$$\omega_2(g, h : \alpha, \beta) = 2^{-5} \, \pi^4 e^{-\lambda_1} \, \omega_1(2\lambda_1 : \alpha-4, \beta-4) \, , \tag{3.4}$$

where $\omega_1(g; \alpha, \beta) = \Gamma(\beta)^{-1} g^{\beta} \zeta(g; \alpha, \beta)$ for $g > 0$. Our assertion follows from (3.4) by Shimura [13, Theorem 3.1].

(ii) $h = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Then

$$\begin{aligned} \eta_2(g, h : \alpha, \beta) &= \int_{x \pm h \in \mathfrak{R}_2^+} e^{-(g,x)} \det(x+h)^{\alpha-5} \det(x-h)^{\beta-5} \, dx \\ &= e^{-(\lambda_1+\lambda_2)} \, 2^{2(\alpha+\beta-5)} \, \zeta_2(2g : \alpha, \beta) \, . \end{aligned}$$

By Theorem 3.1 in Shimura [13] with $a = 2\lambda_1$, $b = 2\lambda_2$, we get

$$\begin{aligned} \zeta_2(g : \alpha, \beta) = \int_{\mathbb{R}^8} & \zeta_1(\lambda_1 + \lambda_2 W : \alpha, \beta) \, \zeta_1(\lambda_2(1+W) : \alpha-4, \beta-4) \\ & \cdot e^{-\lambda_2 W} (1+W)^{\alpha+\beta-9} \, dw \, , \end{aligned}$$

where $W = \|w\|^2$. Therefore we get

$$\begin{aligned} \omega_2(g, h : \alpha, \beta) = & e^{-(\lambda_1+\lambda_2)} \, 2^{-6} \, \pi^{-4} \, \lambda_2^4 \int_{\mathbb{R}^8} e^{-2\lambda_2 W} (1 + \lambda_1^{-1} \lambda_2 W)^{-\beta} \\ & \cdot (1+W)^{\alpha-5} \omega_1(2(\lambda_1 + \lambda_2 W) : \alpha, \beta) \\ & \cdot \omega_1(2\lambda_2(1+W) : \alpha-4, \beta-4) \, dw \, . \end{aligned}$$

Then our assertion follows from this expression as in Shimura [13, Theorem 3.1, Case IV].

(iii) $h = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. Then

$$\eta_2(g, h : \alpha, \beta) = \int_{x \pm h \in \mathfrak{R}_2^+} e^{-(g,x)} \det(x+h)^{\alpha-5} \det(x-h)^{\beta-5} \, dx \, .$$

By (4.28) in Shimura [13] with $a = 2\lambda_1$, $b = 2\lambda_2$, we get

$$\eta_2(g,h:\alpha,\beta) = 2^{2\alpha+2\beta-10} e^{-(\lambda_1+\lambda_2)} \int_{\mathbb{R}^8} e^{-2(\lambda_1+\lambda_2)W} (1+W)^{\alpha+\beta-5} \cdot \zeta_1(2\lambda_1(1+W):\alpha,\beta-4) \cdot \zeta_1(2\lambda_2(1+W):\beta,\alpha-4)\, dw\,.$$

Therefore

$$\omega_2(g,h:\alpha,\beta) = 2^{-2}(\lambda_1\lambda_2)^2 e^{-(\lambda_1+\lambda_2)} \cdot \int_{\mathbb{R}^8} e^{-2(\lambda_1+\lambda_2)W} (1+W)^3 \omega_1(2\lambda_1(1+W):\alpha,\beta-4) \cdot \omega_1(2\lambda_2(1+W):\beta,\alpha-4)\, dw.$$

Our assertion follows from this expression as in Shimura [13, Theorem 4.2, Case IV].

Case 2. $m = 3$.

Because of (4), (6) and (7), it is enough to consider the cases

$$h = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$
$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix},$$
$$g = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda_3 \end{pmatrix},$$
$$\lambda = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}.$$

(i) $h = \begin{pmatrix} k & 0 \\ 0 & 0 \end{pmatrix}$, where $k = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$.

Let $x = \begin{pmatrix} z & y \\ y^* & c \end{pmatrix}$, where z is a 2×2 Hermitian matrix over the Cayley numbers. Then

$\eta_3(g,h:\alpha,\beta) =$

$$= \int_{Q(h)} e^{-(\lambda,z)-\lambda_3 c} \det \begin{pmatrix} z+k & y \\ y^* & c \end{pmatrix}^{\alpha-9} \det \begin{pmatrix} z-k & y \\ y^* & c \end{pmatrix}^{\beta-9} dzdydc.$$

Put $u = z - c^{-1}yy^*$. If $y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$, then

$$yy^* = \begin{pmatrix} N(y_1) & y_1\bar{y}_2 \\ y_2\bar{y}_1 & N(y_2) \end{pmatrix}.$$

Therefore

$$\begin{pmatrix} z & y \\ y^* & c \end{pmatrix} \in \mathfrak{R}_3^+ \quad \text{if and only if} \quad z - c^{-1}yy^* \in \mathfrak{R}_2^+, \quad c > 0,$$

and

$$\det \begin{pmatrix} z & y \\ y^* & c \end{pmatrix} = c \det(z - c^{-1}yy^*).$$

Then $Q(h)$ is mapped into $Q' = \{(u, y, c) : u \pm k \in \mathfrak{R}_2^+\}$ by $(z, y, c) \longrightarrow (u, y, c)$. So

$$\eta_3(g, h : \alpha, \beta) =$$

$$\int_{Q'} e^{-(\lambda,u)-\lambda_3 c-(\lambda,c^{-1}yy^*)} \det(u+k)^{\alpha-9} \det(u-k)^{\beta-9} c^{\alpha+\beta-18}\, dudydc$$

$$= \frac{\pi^8\, \Gamma(\alpha+\beta-9)\, \lambda_3^{-(\alpha+\beta-9)}}{\det \lambda^4}$$

$$\cdot \int_{u \in \mathfrak{R}_2^+} e^{-(\lambda,u)} \det(u+k)^{\alpha-9} \det(u-k)^{\beta-9}\, du$$

$$= \frac{\pi^8 \Gamma(\alpha+\beta-9) \lambda_3^{-(\alpha+\beta-9)}}{\det \lambda^4}\, \eta_2(\lambda, k : \alpha-4, \beta-4).$$

Therefore we get

$$\omega_3(g, h : \alpha, \beta) = 2^{-4p-4q}\, \pi^8\, \omega_2(\lambda, k : \alpha-4, \beta-4) \tag{3.5}$$

for $k = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, and

$$\omega_3(g, h : \alpha, \beta) = \pi^4\, 2^{-4}\, \omega_2(\lambda, k : \alpha-4, \beta-4) \tag{3.6}$$

for $k = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$. Our assertion follows from these relations by Case 1.

(ii) $h = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. By (2),

$$\eta_3(g, \varepsilon : \alpha, \beta) = e^{-(g,\varepsilon)}\, 2^{3(\alpha+\beta-9)} \zeta_3(2g : \alpha, \beta)\,, \tag{3.7}$$

$$\zeta_3(g : \alpha, \beta) = \int_{\mathfrak{R}_3^+} e^{-(g,x)} \det(x+\varepsilon)^{\alpha-9} \det x^{\beta-9}\, dx. \tag{3.8}$$

As in Shimura [13, p. 283], we make the following substitution

$$x = \begin{pmatrix} z & y \\ y^* & c \end{pmatrix} \longrightarrow \begin{pmatrix} u & v \\ v^* & w \end{pmatrix},$$

$$u = z, \quad y = (u^2 + u)^{1/2} v, \quad r\, w\, r = c - (z+1)[v], \quad r = (1 + v\, v^*)^{1/2}.$$

NOTE. Here we define

$$v^* v = N(v_1) + N(v_2), \quad v\, v^* = \begin{pmatrix} N(v_1) & v_1 \bar{v}_2 \\ v_2 \bar{v}_1 & N(v_2) \end{pmatrix} \quad \text{if } v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix},$$

$$Y[\xi] = y_1 N(\xi_1) + y_2 N(\xi_2) + (y_{12}\, \xi_2, \xi_1)$$

if $Y = \begin{pmatrix} y_1 & y_{12} \\ \bar{y}_{12} & y_2 \end{pmatrix}$, $\xi = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}$.

Then we can prove that

$$A[B\xi] = B^* A B\,[\xi],$$

where A, B are 2×2 Hermitian matrices over Cayley numbers. (Here B^*AB is well-defined since $B(AB) = (BA)B$ which follows from the identity $(ax)a = a(xa)$ for a, x Cayley numbers.) And also

$$\operatorname{tr}((xy)z) = \operatorname{tr}(x(yz)) = \operatorname{tr}((yz)x) = \operatorname{tr}(y(zx))\,.$$

Therefore we can show that

$$(z+1)[v] = z^{-1}[y]\,, \quad (z+1)^{-1}[y] = z[v]\,.$$

Here z^{-1} represents Jordan algebra inverse of z and we can show by direct calculation that $\det(z(z+1)) = \det z \det(z+1)$. Hence

$$r\,w\,r = c - z^{-1}[y]\,,$$
$$c+1-(z+1)^{-1}[y] = c+1-z[v] = r\,w\,r + r^2 = r(1+w)r\,.$$

Therefore

$$\begin{aligned}\det(x+\varepsilon) &= \det(z+1)(c+1-(z+1)^{-1}[y]) \\ &= \det(1+u)(1+w)(1+v^*v)\,, \\ \det x &= \det z(c-z^{-1}[y]) = \det u(1+v^*v)w\,.\end{aligned}$$

By the above substitution, $\mathfrak{K}_3^+$ is mapped bijectively into $\{(u,v,w) : u>0, w>0, v\in\mathfrak{C}^2\} = R$. Then

$$\frac{\partial(z,y,c)}{\partial(u,v,w)} = r^2\det(u^2+u)^4 = (1+v^*v)\det u^4\det(1+u)^4\,.$$

Therefore

$$\begin{aligned}\zeta_3(g:\alpha,\beta) &= \int_R e^{-(\lambda,u)-\lambda_3(rwr+(u+1)[v])}\det(u+1)^{\alpha-5} \\ &\qquad\cdot\det u^{\beta-5}(1+w)^{\alpha-9}w^{\beta-9}(1+v^*v)^{\alpha+\beta-17}\,du dv dw \\ &= \int_{v\in F}\left(\int_{u>0} e^{-(\lambda+\lambda_3 vv^*,u)}\det(u+1)^{\alpha-5}\det u^{\beta-5}\,du\right) \\ &\qquad\cdot\left(\int_{w>0} e^{-\lambda_3 r^2 w}(1+w)^{\alpha-9}w^{\beta-9}\,dw\right) \\ &\qquad\cdot e^{-\lambda_3(N(v_1)+N(v_2))}(1+v^*v)^{\alpha+\beta-17}\,dv \\ &= \int_{v\in F}\zeta_2(\lambda+\lambda_3 vv^*:\alpha,\beta)\,\zeta_1(\lambda_3(1+v^*v):\alpha-8,\beta-8) \\ &\qquad\cdot e^{-\lambda_3(N(v_1)+N(v_2))}(1+v^*v)^{\alpha+\beta-17}\,dv.\end{aligned}$$

where $F = \mathfrak{C}^2$. Therefore we get

$$\begin{aligned}\omega_3(g,\varepsilon:\alpha,\beta) &= 2^{-9}\pi^{-8} \\ &\cdot\int_F \omega_2(\lambda+\lambda_3 vv^*,\varepsilon:\alpha,\beta)\,\omega_1(2\lambda_3(1+v^*v):\alpha-8,\beta-8) \\ &\quad\cdot e^{-\lambda_3(1+v^*v)}(\lambda_1\lambda_2)^\beta\lambda_3^8(1+v^*v)^{\alpha-9}\det(\lambda+\lambda_3 vv^*)^{-\beta}\,dv\,.\end{aligned}$$

This integral expression provides the analytic continuation of ω_3. Now we can assume $\lambda_1, \lambda_2 \geq \lambda_3$. Then we have

$$\det(\lambda + \lambda_3 vv^*) \leq \lambda_1\lambda_2(1 + v^*v)\,.$$

By induction, we get

$$|\,\omega_3(g, \varepsilon : \alpha, \beta)| \leq A\, e^{-\tau(hg)}(1 + \mu(hg)^{-B})\,.$$

(Use Lemma 2.8 in Shimura [13, p. 278]).
To prove the functional equation, consider

$$\begin{aligned}
&\Gamma_3(\beta)\zeta_3(g : 9-\beta, \alpha) \\
&\quad = \int_{\mathfrak{K}_3^+} e^{-(g,x)}\Gamma_3(\beta)\det(x+\varepsilon)^{-\beta}\det x^{\alpha-9}\,dx \\
&\quad = \int_{\mathfrak{K}_3^+} e^{-(g,x)} \int_{\mathfrak{K}_3^+} e^{-(u,x+\varepsilon)}\delta(u)^{\beta-9}\,du\,\det x^{\alpha-9}\,dx \\
&\quad = \int_{\mathfrak{K}_3^+} e^{-(u,\varepsilon)}\delta(u)^{\beta-9} \int_{\mathfrak{K}_3^+} e^{-(g+u,x)}\det x^{\alpha-9}\,dxdu \\
&\quad = \Gamma_3(\alpha) \int_{\mathfrak{K}_3^+} e^{-(ay,\varepsilon)}\det a(y+\varepsilon)^{-\alpha}\det a y^{\beta-9}\nu(a)^9\,dy \\
&\quad = \Gamma_3(\alpha) \int_{\mathfrak{K}_3^+} e^{-(y,a^{*-1}\varepsilon)}\det(y+\varepsilon)^{-\alpha}\det y^{\beta-9}\nu(a)^{\beta-\alpha}\,dy \\
&\quad = \Gamma_3(\alpha)\,\nu(a)^{\beta-\alpha}\zeta_3(a^{*-1}\varepsilon : 9-\alpha, \beta)\,.
\end{aligned}$$

(Here we have taken $a \in \mathfrak{S}$ such that $g = a\,\varepsilon$ and let $ay = u$.)
Here $\nu(a) = \det g$, a^{*-1} and g have the same eigenvalues, *i.e.* there exists $a, b \in GL(\mathfrak{J}_{\mathbb{R}}^{(3)})$ such that $bg = a^{*-1}\varepsilon$. So we have $\zeta_3(a^{*-1}\varepsilon : 9-\alpha, \beta) = \zeta_3(g : 9-\alpha, \beta)$. Therefore

$$\Gamma_3(\beta)\zeta_3(g : 9-\beta, \alpha) = \Gamma_3(\alpha)\det g^{\beta-\alpha}\zeta_3(g : 9-\alpha, \beta)\,.$$

This proves the functional equation (3.2) from (3.1) and (3.8).

(iii) $h = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$. Then

$$\eta_3(g, h : \alpha, \beta) = \int_{Q(h)} e^{-(g,x)}\det(x+h)^{\alpha-9}\det(x-h)^{\beta-9}\,dx\,.$$

Let $x = \begin{pmatrix} z & y \\ y^* & c \end{pmatrix}$. Then

$$x + h = \begin{pmatrix} z+1 & y \\ y^* & c-1 \end{pmatrix}, \quad x - h = \begin{pmatrix} z-1 & y \\ y^* & c+1 \end{pmatrix}.$$

As in Shimura [13, p. 289], we make the following substitution

$$\begin{gathered} (z, y, c) \longrightarrow (u, v, w), \\ u = z - 1 - 2\,v\,v^*, \qquad w = c - 1 - 2\,v^* v, \\ v = (1 - r\,r^*)^{-1/2} r, \qquad r = (z+1)^{-1/2}(c+1)^{-1/2} y. \end{gathered}$$

Here

$$\begin{aligned} 1 - r\,r^* &= 1 - (c+1)^{-1}(z+1)^{-1/2} y y^* (z+1)^{-1/2} \\ &= (z+1)^{-1/2}(z + 1 - (c+1)^{-1} y y^*)(z+1)^{-1/2} \in \mathfrak{K}_2^+ . \end{aligned}$$

(Here we use the fact that $(ax)(ya) = a(xy)a$, for $a, x, y \in \mathfrak{C}$ in order to justify the associativity.) And

$$\begin{aligned} 1 + v\,v^* &= (1 - r\,r^*)^{-1}, \\ r &= (1 + v\,v^*)^{-1/2} v, \\ \det(1 + v\,v^*) &= 1 + v^* v, \\ c - 1 - (z+1)^{-1}[y] &= c - 1 - (c+1) r^* r \\ &= (c+1)(1 - r^* r) - 2 \\ &= (c+1)(1 + v^* v)^{-1} \left(1 - \frac{2(1 + v^* v)}{c+1}\right) \\ &= (1 + v^* v)^{-1} w, \\ z - 1 - (c+1)^{-1} y\,y^* &= z - 1 - (z+1)^{1/2} r\,r^* (z+1)^{1/2} \\ &= (z+1)^{1/2}(1 - r\,r^*)(z+1)^{1/2} - 2 \\ &= (z+1)^{1/2}(1 + v\,v^*)^{-1} \\ &\quad \cdot ((z + 1 - 2(1 + vv^*))(z+1)^{-1/2}) \\ &= (z+1)^{1/2}(1 + v\,v^*)^{-1}(u(z+1)^{-1/2}). \end{aligned}$$

(Here we can show by direct calculation that $zz^{-1/2} = z^{1/2}$ for $z \in \mathfrak{K}_2^+$.) Now we can check by direct calculation that

$$\det(zwz - 2) = \det w \, \det(z^2 - 2w^{-1}).$$

(Both are well-defined.) Therefore

$$\begin{aligned}\det(z-1-(c+1)^{-1}y\,y^*) &= \det((z+1)^{1/2}(1-r\,r^*)(z+1)^{1/2}-2)\\ &= \det(1+v\,v^*)^{-1}\det u\,.\end{aligned}$$

The substitution maps $Q(h)$ into $Q' = \{(u,v,w):\ u>0, w>0, v\in \mathfrak{C}^2\}$. Here the inverse map is given by

$$\begin{aligned} z &= u+1+2\,v\,v^*\,,\\ c &= w+1+2\,v^*v\,,\\ y &= (z+1)^{1/2}(c+1)^{1/2}r\,,\\ r &= (1+v\,v^*)^{-1/2}v\,,\end{aligned}$$

and the Jacobian determinant is

$$\begin{aligned}\frac{\partial(u,v,w)}{\partial(z,r,c)} &= \det\left(\frac{\partial v}{\partial r}\right),\\ \frac{\partial(z,r,c)}{\partial(z,y,c)} &= \frac{\partial r}{\partial y}\\ &= \det((z+1)^{-1/2}(c+1)^{-1/2})^8\\ &= (c+1)^{-8}\det(z+1)^{-4}\,.\end{aligned}$$

As in Shimura [13, p. 278], we can show

$$\det\left(\frac{\partial v}{\partial r}\right) = \det(1-r\,r^*)^{-9} = \det(1+v\,v^*)^9\,.$$

Therefore

$$\frac{\partial(z,y,c)}{\partial(u,v,w)} = \det(z+1)^4(c+1)^8\det(1+v\,v^*)^{-9}\,.$$

So we have

$$\begin{aligned}&\eta_3(g,h:\alpha,\beta)\\ &\quad=\int_{Q'} e^{-(\lambda,u+1+2v\,v^*)-\lambda_3(w+1+2v^*v)}\det(u+2(1+v\,v^*))^{\alpha-9}\\ &\qquad\cdot(1+v^*v)^{-(\alpha-9)}w^{\alpha-9}(w+2(1+v^*v))^{\beta-9}\det(1+v\,v^*)^{-(\beta-9)}\\ &\qquad\cdot\det u^{\beta-9}\det(u+2(1+v\,v^*)^4(w+2(1+v^*v))^8\end{aligned}$$

$$\cdot \det(1 + v\,v^*)^{-9}\,du dv dw\,.$$

Let $K = v\,v^*$, $K' = v^*v$. And we note that $1 + v^*v = \det(1 + v\,v^*)$. Therefore we get

$$\begin{aligned}\eta_3(g,h:\alpha,\beta) = e^{-\operatorname{tr} g}\int_{Q'} & e^{-(\lambda,u)}\det(u+2(1+K))^{\alpha-5}\det u^{\beta-9}\\ & \cdot (w+2(1+K'))^{\beta-1}w^{\alpha-9}e^{-\lambda_3 w}\\ & \cdot e^{-2(\lambda,K)-2\lambda_3 K'}(1+K')^{9-\alpha-\beta}\,du dv dw\,.\end{aligned}$$

Here

$$\begin{aligned}\int_{u>0} & e^{-(\lambda,u)}\det(u+2(1+K))^{\alpha-5}\det u^{\beta-9}\\ & = (\det(2(1+K)))^{\alpha+\beta-9}\zeta_2(a^{*-1}\lambda:\alpha,\beta-4)\,.\end{aligned}$$

(Take $a \in \mathfrak{S}$ such that $2(1+K) = a\,\varepsilon$. Actually, $a : x \longrightarrow (2(1+K))^{1/2}x(2(1+K))^{1/2}$.)

$$\begin{aligned}\int_{w>0} & e^{-\lambda_3 w}(w+2(1+K'))^{\beta-1}w^{\alpha-9}\,dw\\ & = \int_{w>0} e^{-2\lambda_3(1+K')w}(1+w)^{\beta-1}w^{\alpha-9}\,2^{\alpha+\beta-9}(1+K')^{\alpha+\beta-9}\,dw\\ & = 2^{\alpha+\beta-9}(1+K')^{\alpha+\beta-9}\zeta_1(2\lambda_3(1+K'):\beta,\alpha-8)\,.\end{aligned}$$

Therefore

$$\begin{aligned}\eta_3(g,h:\alpha,\beta) = & e^{-\operatorname{tr} g}\,2^{3(\alpha+\beta-9)}\\ & \cdot \int_F \zeta_2(a^{*-1}\lambda:\alpha,\beta-4)\,\zeta_1(2\lambda_3(1+K'):\beta,\alpha-8)\\ & \cdot e^{-2(\lambda,K)-2\lambda_3 K'}(1+K')^{\alpha+\beta-9}\,dv\,,\end{aligned}$$

where $F = \mathfrak{C}^2$. So

$$\begin{aligned}\omega_3(g,h:\alpha,\beta) = 2^{-1}e^{-\operatorname{tr} g}\int_F & e^{-2(\lambda,K)-2\lambda_3 K'}e^{(\lambda,1+K)}(1+K')^3\\ & \cdot (\det\lambda)^2\,\lambda_3^4\,\omega_2(\frac{1}{2}a^{*-1}\lambda,\varepsilon:\alpha,\beta-4)\\ & \cdot \omega_1(2\lambda_3(1+K'):\beta,\alpha-8)\,dv\,.\end{aligned}$$

Now we have

$$|\omega_2(\frac{1}{2}a^{*-1}\lambda, \varepsilon : \alpha, \beta - 4)| \leq A(1 + \mu(hg)^{-B})e^{-(\lambda, 1+K)},$$
$$|\omega_1(2\lambda_3(1 + K') : \beta, \alpha - 8)| \leq A(1 + \mu(hg)^{-B'}).$$

(Use $\mu(\lambda) \leq \mu(\frac{1}{2}a^{*-1}\lambda)$.) Therefore

$$|\omega_3(g, h : \alpha, \beta)| \leq A(1 + \mu(hg)^{-B})(\det \lambda)^2 \lambda_3^4 e^{-\tau(hg)} \cdot \int_F e^{-2(\lambda, K) - 2\lambda_3 K'}(1 + K')^3 \, dv,$$

if (α, β) stays in a compact set T in $\mathbb{C}^2$ and $A, B > 0$ are constants depending only on T.
By Holder's inequality,

$$\begin{aligned}
&\int_F e^{-2(\lambda, K) - 2\lambda_3 K'}(1 + K')^3 \, dv \\
&\quad \leq \left(\int_F e^{-4(\lambda, K)} \det(1 + K)^3 \, dv \int_F e^{-4\lambda_3 K'}(1 + K')^3 \, dv \right)^{1/2} \\
&\quad \leq A(\det \lambda)^{-2} \lambda_3^{-4}(1 + \mu(hg)^{-B}).
\end{aligned}$$

(Here we use Lemma 2.8 in Shimura [13, p. 278] in the second inequality.) Therefore this proves the analytic continuation as well as the inequality. On the other hand,

$$\begin{aligned}
&\omega_3(g, h : 9 - \beta, 9 - \alpha) \\
&= 2^{-1} e^{-\operatorname{tr} g} \int_F e^{-2(\lambda, K) - 2\lambda_3 K'} e^{(\lambda, 1+K)}(1 + K')^3 (\det \lambda)^2 \lambda_3^4 \\
&\qquad \cdot \omega_2(\frac{1}{2}a^{*-1}\lambda, \varepsilon : 9 - \beta, 5 - \alpha)\omega_1(2\lambda_3(1 + K') : 9 - \alpha, 1 - \beta) \, dv \\
&= \omega_3(g, h : \alpha, \beta).
\end{aligned}$$

This completes the proof of the Theorem.

From (3) and (3.1), we have

$$\begin{aligned}
\xi_m(g, h; \alpha, \beta) = |\sigma_m|^{-1} i^{m(\beta - \alpha)} 2^{\varphi} \pi^{\psi} \Gamma_r(\alpha + \beta - \kappa(m)) \\
\cdot \Gamma_{m-q}(\alpha)^{-1} \Gamma_{m-p}(\beta)^{-1} \\
\cdot \det g^{\kappa(m) - \alpha - \beta} \delta_+(hg)^{\alpha + 2q - \kappa(m)} \\
\cdot \delta_-(hg)^{\beta + 2p - \kappa(m)} \omega_m(2\pi g, h; \alpha, \beta),
\end{aligned} \tag{3.9}$$

if $h \in V(p,q,r)$, where $|\sigma_m| = 2^{m(\kappa(m)-1)}$ and

$$\varphi = (2p-m)\alpha + (2q-m)\beta + (m+r)\kappa(m) + 4pq\,,$$
$$\psi = p\alpha + q\beta + r + 4(r(r-1) - pq)\,.$$

Furthermore, we have

Corollary. *If* $h \in V(p,0,r)$,

$$\omega_m(g,h;\alpha,4r) = \omega_m(g,h;\kappa(m),\beta) = 2^{-p\kappa(m)}\pi^{4pr}e^{-(g,h)}. \tag{3.10}$$

PROOF. The case when $r=0$, *i.e.* $h \in \mathfrak{K}_m^+$, follows from (2) and (3.1) by observing that

$$\zeta_m(g;\kappa(m),\beta) = \int_{\mathfrak{K}_m^+} e^{-(g,u)}\det u^{\beta-\kappa(m)}\,du = \Gamma_m(\beta)\det g^{-\beta}\,.$$

If $h \in V(p,0,r)$ with $r>0$, the formula follows from (3.4), (3.5) and (3.6) by noting that $\omega_1(g;1,\beta)=1$.

4. The singular series $S(T,s)$.

$$S(T,s) = \sum_{X \in \mathfrak{Z}_{\mathbb{Q}}^j/\Lambda_j} e^{2\pi i(T,X)}\kappa(X)^{-s}. \tag{4.1}$$

By Baily [1], we can assume $T = \begin{pmatrix} T_1 & 0 \\ 0 & 0 \end{pmatrix}$ where T_1 is a $j \times j$ nonsingular matrix if T has rank equal to j. In this section we calculate $S(T,s)$ for T a singular matrix by Karel's method (*cf.* Karel [6]). For T nonsingular, Karel [6] gave explicit formulas. We modify his method to get the results for the case when T is singular. The idea is to represent $S(T,s)$ in terms of $S(T_1, s-8(3-j))$.

Theorem. (i) $j=1$,

$$S(t,s) = \begin{cases} \dfrac{\zeta(s-1)}{\zeta(s)}\,, & \text{if } t=0\,, \\ \dfrac{1}{\zeta(s)}\displaystyle\sum_{a|t,a>0} a^{1-s}\,, & \text{if } t \neq 0\,. \end{cases} \tag{4.2}$$

(ii) $j = 2$,

$$(4.3)\quad S(T,s) = \begin{cases} \dfrac{\zeta(s-5)\zeta(s-9)}{\zeta(s)\zeta(s-4)}, & \text{if } T = 0, \\ \dfrac{\zeta(s-5)\zeta(s-8)}{\zeta(s)\zeta(s-4)} S(t, s-8), & \text{if } T = \begin{pmatrix} t & 0 \\ 0 & 0 \end{pmatrix}, \\ \dfrac{1}{\zeta(s)\zeta(s-4)} \displaystyle\prod_{p|\det T} f_T^p(p^{5-s}), & \text{if } \det T \neq 0. \end{cases}$$

(iii) $j = 3$,

$$(4.4.a)\qquad S(T,s) = \frac{\zeta(s-9)\zeta(s-13)\zeta(s-17)}{\zeta(s)\zeta(s-4)\zeta(s-8)}, \qquad \text{if } T = 0.$$

$$(4.4.b)\qquad S(T,s) = \frac{\zeta(s-9)\zeta(s-13)\zeta(s-16)}{\zeta(s)\zeta(s-4)\zeta(s-8)} S(t, s-16),$$

$$\text{if } T = \begin{pmatrix} t & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

$$(4.4.c)\qquad S(T,s) = \frac{\zeta(s-9)\zeta(s-12)}{\zeta(s)\zeta(s-4)} S(T_1, s-8),$$

$$\text{if } T = \begin{pmatrix} T_1^{2\times 2} & 0 \\ 0 & 0 \end{pmatrix}.$$

$$(4.4.d)\qquad S(T,s) = \frac{1}{\zeta(s)\zeta(s-4)\zeta(s-8)} \prod_{p|\det T} f_T^p(p^{9-s}),$$

if $\det T \neq 0$, *where* f_T^p *is a polynomial.* (See Karel [6])

By Karel [6, pp. 186-187], we have

$$S(T,s) = \prod_p S_p(T,s),$$

$$(4.5)\qquad S_p(T,s) = \sum_{X \in \mathfrak{Z}(j)_p / \Lambda(j)_p} \varepsilon_p((T,X)) \kappa_p(X)^{-s},$$

$$(4.6)\qquad (1-p^{-s})^{-1} S_p(T,s) = \sum_{m=0}^{\infty} \alpha_m(T) p^{-ms},$$

$$\alpha_m(T) = \sum_X \omega_m^{(T,X)},$$

where $X \in \Lambda(j)_p/p^m\Lambda(j)_p$ and $\tau_i(X) \equiv 0 \mod p^{m(i-1)}$ for $2 \leq i \leq j$.

When $j = 1$, the results are well-known (See Karel [6]). So we calculate the cases $j = 2$ and $j = 3$.

1) *The case* $j = 2$. By Karel [6], we can assume $T = D(t, t')$, where $t|t'$ in $\mathbb{Z}_p$. Let $q = p^m$, $\omega = \omega_m$. By (4.6),

$$S_p(T, s) = (1 - p^{-s}) \sum_{m=0}^{\infty} \alpha_m(T) p^{-ms}$$

$$\alpha_m(T) = \sum_{Z} \omega_m^{(T,Z)}$$

where $Z \in \Lambda(2)_p/p^m\Lambda(2)_p$, $Z^* \equiv 0 \mod p^m$. By Karel [6, (6.2)],

$$\alpha_m(T) = q^4 \sum_{b=1}^{q} \omega^{bt'} \sum_{h=1}^{q} [h]_m^4 , \quad t + hb \equiv 0 \mod q . \tag{4.7}$$

(i) $T = 0$. Here $t = t' = 0$. So we have, by using the fact that

$$\sum_{a,b=1}^{q} \omega^{hab} = [h]_m ,$$

$$\alpha_m(0) = q^4 \sum_{b=1}^{q} \sum_{h=1}^{q} [h]_m^4 , \quad (hb \equiv 0 \mod q)$$

$$= q^4 \sum_{h=1}^{q} [h]_m^4 \sum_{a,b=1}^{q} \omega^{hab} = q^4 \sum_{h=1}^{q} [h]_m^5 = q^4 \sum_{k=0}^{m} p^{5k} M_k ,$$

where M_k is the number of $h \mod p^m$ with $v^m(h) = k$;

$$M_k = \begin{cases} (p-1)p^{m-k-1} , & \text{if } k < m , \\ p^{m-m} = 1 , & \text{if } k = m . \end{cases}$$

Therefore

$$\alpha_m(0) = p^{4m} \frac{(p^m - p^{m-1})(1 - p^{4m}) + p^{5m}(1 - p^4)}{1 - p^4} .$$

So we have

$$(1-p^{-s})^{-1}S_p(0,s) = \frac{1-p^{4-s}}{(1-p^{5-s})(1-p^{9-s})}$$
$$S_p(0,s) = \frac{(1-p^{-s})(1-p^{4-s})}{(1-p^{5-s})(1-p^{9-s})}.$$

(ii) $T = \begin{pmatrix} t & 0 \\ 0 & 0 \end{pmatrix}$, $t \neq 0$. By (4.7),

$$\alpha_m(T) = q^4 \sum_{b=1}^{q}\sum_{h=1}^{q}[h]_m^4\,, \quad t+hb \equiv 0 \mod q\,.$$

By Karel [6, (6.5)], we have

$$S_p(T,s) = (1-p^{-s})(1-p^{4-s})r_p(T,s)$$
$$r_p(T,s) = \sum_{m=0}^{\infty}\alpha'_m(T)p^{m(4-s)}$$
$$\alpha_m(T) = p^{4m}(\alpha'_m(T) - \alpha'_{m-1}(T))$$
$$\alpha'_m(T) = [t]_m^4 \sum_{b=1}^{p^m}\sum_{k=0}^{v^m(t;b)} p^{-3k} = [t]_m^4 \sum_{k=0}^{v^m(t)} p^{-3k} \sum_{\substack{b \bmod p^m \\ b\equiv 0 \bmod p^k}} 1\,.$$

Let $v = v^m(t) = \min\{m, v(t)\}$, $\tau = v(t)$. Then we have

$$\alpha'_m(T) = p^{4v}\sum_{k=0}^{v} p^{-3k}p^{m-k} = p^m \sum_{k=0}^{v} p^{4k}.$$

So

$$r_p(T,s) = \sum_{m=0}^{\infty} p^{(4-s)m}p^m \sum_{k=0}^{v} p^{4k} = \sum_{m=0}^{\infty} p^{(5-s)m}\sum_{k=0}^{v} p^{4k}$$
$$= \sum_{k=0}^{\tau} p^{4k}\sum_{m=k}^{\infty} p^{m(5-s)} = \sum_{k=0}^{\tau} p^{4k}\frac{(p^{5-s})^k}{1-p^{5-s}}$$
$$= \frac{1}{1-p^{5-s}}\sum_{k=0}^{\tau}(p^{9-s})^k\,.$$

Therefore

$$S_p(T,s) = \frac{(1-p^{-s})(1-p^{4-s})}{1-p^{5-s}} \sum_{k=0}^{\tau} (p^{9-s})^k.$$

Since $S_p(t,s) = (1-p^{-s}) \sum_{k=0}^{\tau} p^{k(1-s)}$, we have

$$S_p(T,s) = \frac{(1-p^{-s})(1-p^{4-s})}{(1-p^{5-s})(1-p^{8-s})} S_p(t, s-8) \quad \text{for} \quad T = \begin{pmatrix} t & 0 \\ 0 & 0 \end{pmatrix}.$$

2) $j = 3$. By Karel [6], we can assume $T = D(t, t', t'')$, where $t|t'$, $t'|t''$ in $\mathbb{Z}_p$. As in Karel [6, p. 191], let $U = D(0,0,1)$ and we use the letters j, k, l, m, n to denote rational integers; a, b, c, h, ξ denote p-adic integers; u, v, w, x, y, z denote elements of $\mathfrak{o}_p$. The letters W, Z will be used for elements of Λ_p of the respective forms

$$W = \begin{pmatrix} 0 & 0 & y \\ 0 & 0 & z \\ \bar{y} & \bar{z} & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} a & x & 0 \\ \bar{x} & b & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

We let $X = \begin{pmatrix} a & x & y \\ \bar{x} & b & z \\ \bar{y} & \bar{z} & c \end{pmatrix} = cU + W + Z$. The letters H will denote any elements in $p^m \Lambda_p$ with

$$H = \begin{pmatrix} h' & u & 0 \\ \bar{u} & h'' & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

and the letter K will denote the sum $H + hU$. Primed and subscripted variables will always have the same generic meaning as the corresponding variables without primes and subscripts. Let

$$\tilde{Z} = 2U \times Z = \begin{pmatrix} b & -x & 0 \\ -\bar{x} & a & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Then $X^* = c'U + W' + Z'$, where $c' = ab - N(x) = Q(Z)$, $W' = -2W \circ \tilde{Z}$, $Z' = W^* + c\tilde{Z}$, *i.e.*

$$W' = \begin{pmatrix} 0 & 0 & -by + xz \\ 0 & 0 & \bar{x}y - az \\ * & * & 0 \end{pmatrix}, \quad Z' = \begin{pmatrix} bc - N(z) & y\bar{z} - cx & 0 \\ z\bar{y} - c\bar{x} & ac - N(y) & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

We have $\det X = cQ(Z) + (W^*, Z)$ and by (4.6)

$$\alpha_m(T) = \sum_{X \bmod q} \omega^{(T,X)}, \tag{4.8}$$
$$(X^* \equiv 0 \mod q,\ \det X \equiv 0 \mod q^2).$$

Using the block decomposition $X = cU + W + Z$, we get

$$\alpha_m(T) = \sum_{Z(q)} \sum_{c,W(q)} \omega^{(T,cU+W+Z)}, \tag{4.9}$$

where Z, c, W satisfy: (i) $Q(Z) \equiv 0 \mod q$, (ii) $W^* + c\tilde{Z} \equiv 0 \mod q$, (iii) $W\tilde{Z} \equiv 0 \mod q$, (iv) $cQ(Z) + (W^*, Z) \equiv 0 \mod q^2$. This may be rewritten

$$\alpha_m(T) = \sum_{Z(q)} \omega^{(T,Z)} \beta_m(T; Z), \quad (Q(Z) \equiv 0 \mod q),$$

with

$$\beta_m(T; Z) = \sum_{c,W(q)} \omega^{(T,cU+W)},$$

where c, W are summed under the restrictions (ii), (iii), (iv) of (4.9). By Karel [6, p. 192], we can assume $Z = D(a, b, 0)$ in order to calculate $\beta_m(T; Z)$ and $a|b$ in $\mathbb{Z}_p$, *i.e.* $v_p(a) \leq v_p(b) \leq m$. Then by Karel [6, (8.4), p. 193], we get

$$q^{27} \beta_m(T; Z) = \sum_K \sum_W \omega_{2m}^{(H+hZ,W^*)}, \tag{4.10}$$

where K, W are summed (mod-q^2) with the restrictions $W\tilde{Z} \equiv 0 \mod q$ and

$$qt'' + (H, \tilde{Z}) + hQ(Z) \equiv 0 \mod q^2.$$

Now we calculate $S_p(T, s)$ for $T = D(t, t', 0)$, $t|t'$ (t, t' might be zero.) in terms of $S_p(T', s)$ for $T' = D(t, t')$. From (4.10), we have

$$q^{27} \beta_m(T; Z) = \sum_K \sum_W \omega_{2m}^{(H+hZ,W^*)}, \tag{4.11}$$

where $K = H + hU$, $H \in p^m \Lambda_p$, K and W are summed (mod-q^2) with the restrictions

$$W\tilde{Z} \equiv 0 \mod q, \quad (H, \tilde{Z}) + hQ(Z) \equiv 0 \mod q^2.$$

Rearranging (4.11),

$$q^{27}\beta_m(T;Z) = \sum_W \sum_K \omega_{2m}^{(H+hZ,W^*)},$$

where W, K are summed (mod-q^2) with

$$W\tilde{Z} \equiv 0 \mod q, \quad (H,\tilde{Z}) + hQ(Z) \equiv 0 \mod q^2. \tag{4.12}$$

CLAIM: we may add the restriction on W that $N(y) \equiv 0 \mod p^\zeta$, where $\zeta = v_p(a)$.

PROOF OF CLAIM. Suppose h_0 and $H_0 = \begin{pmatrix} h_0' & u_0 & 0 \\ \bar{u}_0 & h_0'' & 0 \\ 0 & 0 & 0 \end{pmatrix}$ satisfy (4.12), *i.e.*

$$h_0 ab + bh_0' + ah_0'' \equiv 0 \mod q, \tag{4.13}$$

Then h_0 and $H = \begin{pmatrix} h_0' & u_0 & 0 \\ \bar{u}_0 & h_0'' + h'' & 0 \\ 0 & 0 & 0 \end{pmatrix}$ also satisfy (4.12) provided $h''a \equiv 0 \mod q^2$. Since $(H + hZ, W^*) = -h''N(y) + \sigma$, where σ does not depend on h'', so

$$\sum_K \cdots = \sum_{h,h',u} \cdots \left(\sum_{h''} \omega_{2m}^{-h''N(y)} \right).$$

But by setting $h'' = h_0'' + \alpha$, we see that

$$\begin{aligned} \sum_{h''} \omega_{2m}^{-h''N(y)} &= \left(\sum_{\substack{\alpha \bmod q^2 \\ \alpha a \equiv 0 \bmod q^2}} \omega_{2m}^{-\alpha N(y)} \right) \omega_{2m}^{-h_0'' N(y)} \\ &= 0, \qquad \text{unless } N(y) \equiv 0 \mod a. \end{aligned}$$

In this case, multiplying (4.12) by $N(y)$ gives

$$-(h'' + hb)aN(y) \equiv h'N(y) \mod p^{2m+\zeta}.$$

Given $h \in \mathbb{Z}_p$, $h' \equiv 0 \mod q$, the number of $h'' \mod q^2$ satisfying (4.13) and $h'' \equiv 0 \mod q$ is p^ζ if $h \equiv 0 \mod p^{\zeta-k}$, *i.e.* $h \equiv 0 \mod p^{m-v_p(b)}$

and is zero otherwise. ($k = v^m(p^{-m}Q(Z)) = \min\{m, v_p(p^{-m}Q(Z))\} = -m + v_p(a) + v_p(b)$. So $k \le \zeta = v_p(a)$.) Since $Z = D(a,b,0)$,

$$a(H + hZ, W^*) \equiv h'(bN(y) - aN(z)) - ha^2N(z) + a\,\mathrm{tr}(uz\bar{y}) \mod p^{2m+\zeta}. \tag{4.14}$$

Note that h'' does not appear on the right side of (4.14). Therefore

$$\begin{aligned}\sum_K \omega_{2m}^{(H+hZ,W^*)} &= \sum_{h,H} \omega_{2m+\zeta}^{a(H+hZ,W^*)} \\ &= \sum_{h \bmod q^2} \omega_{2m+\zeta}^{-ha^2N(z)} \sum_{h' \bmod q^2} \omega_{2m+\zeta}^{h'(bN(y)-aN(z))} \\ &\quad \cdot \sum_{u \bmod q^2} \omega_{2m+\zeta}^{a\,\mathrm{tr}(uz\bar{y})} \sum_{h''} 1,\end{aligned}$$

where $h \equiv 0 \mod p^{m-v_p(b)}$, $h' \equiv 0 \mod q$ and $u \equiv 0 \mod q$. Here $\sum_{h''} 1 = p^\zeta$ by the above argument. So we have

$$q^{27}\beta_m(T;Z) = p^{10m+\zeta+v_p(b)} \sum_{y,z} 1, \tag{4.15}$$

where y and z are summed (mod-q^2) and satisfy $by \equiv az \equiv 0 \mod q$, $a^2N(z) \equiv 0 \mod p^{m+\zeta+v_p(b)}$, $bN(y) \equiv aN(z) \mod p^{m+\zeta}$ and $z\bar{y} \equiv 0 \mod q$. Equivalently, y and z satisfy that $q^{-1}by$, $q^{-1}az$ are in $\mathfrak{o}_p$, $aN(z) \equiv 0 \mod p^{m+v_p(b)}$, $bN(y) \equiv 0 \mod p^{m+\zeta}$, and $z\bar{y} \equiv 0 \mod p^m$. Thus, $N(q^{-1}by) \equiv N(q^{-1}az) \equiv 0 \mod p^k$. Replacing y by $q^{-1}by$ and z by $q^{-1}az$, (4.15) becomes

$$\beta_m(T;Z) = p^{-16m+k} \sum_y A(y), \tag{4.16}$$

where $y \in \mathfrak{o}_p/p^{2m+k-\zeta}\mathfrak{o}_p$ satisfies $N(y) \equiv 0 \mod p^k$ and where $A(y)$ is the number of $z \in \mathfrak{o}_p/p^{m+\zeta}\mathfrak{o}_p$ satisfying $z\bar{y} \equiv N(z) \equiv 0 \mod p^k$. (Here $z\bar{y} \equiv 0 \mod p^k$ implies $\bar{y}z \equiv 0 \mod p^k$.) By Karel [6, p. 181, Lemma 2.4],

$$A(y) = p^{8(m+\zeta-k)}p^{4k}\left(\sum_{\nu=0}^{f} p^{3\nu} - \sum_{\nu=0}^{f-1} p^{3\nu-1}\right),$$

where $f = f(y) = v^k(y; p^{-k}N(y))$. So

$$A(y) = p^{8m+8\zeta-4k}\left(\sum_{\nu=0}^{f} p^{3\nu} - \sum_{\nu=1}^{f} p^{3\nu-4}\right).$$

Then from (4.16),

$$p^{8m-8\zeta+3k}\beta_m(T;Z) = \sum_{\substack{y \bmod p^{2m+k-\zeta} \\ N(y)\equiv 0 \bmod p^k}} \Big(\sum_{\nu=0}^{f} p^{3\nu} - \sum_{\nu=1}^{f} p^{3\nu-4}\Big)$$

$$= \sum_{f=0}^{k} \Big(\sum_{\nu=0}^{f} p^{3\nu} - \sum_{\nu=1}^{f} p^{3\nu-4}\Big)\Big(\sum_{\substack{y \bmod p^{2m+k-\zeta} \\ f(y)=f}} 1\Big) \tag{4.17}$$

$$= \sum_{f=0}^{k}\sum_{\nu=0}^{f} p^{3\nu}\sigma_f - \sum_{f=0}^{k}\sum_{\nu=1}^{f} p^{3\nu-4}\sigma_f$$

$$= \sum_{\nu=0}^{k} p^{3\nu}\sigma'_\nu - \sum_{\nu=1}^{k} p^{3\nu-4}\sigma'_\nu ,$$

where

$$\sigma_f = \sum_{\substack{y \bmod p^{2m+k-\zeta} \\ f(y)=f}} 1 \quad \text{and} \quad \sigma'_\nu = \sum_{f=\nu}^{k} \sigma_f .$$

Here $\sigma'_\nu = \sum_y 1$, where $y \in \mathfrak{o}_p/p^{2m+k-\zeta}\mathfrak{o}_p$ satisfies $v^k(y;p^{-k}N(y)) \geq \nu$, *i.e.* $y \equiv 0 \bmod p^\nu$ and $N(y) \equiv 0 \bmod p^{k+\nu}$. If we write $y = p^\nu y_0$, then $N(y) \equiv 0 \bmod p^{k+\nu}$ is equivalent to $N(y_0) \equiv 0 \bmod p^{k-\nu}$; hence,

$$\sigma'_\nu = \sum_{\substack{y \bmod p^{2m+k-\zeta-\nu} \\ N(y)\equiv 0 \bmod p^{k-\nu}}} 1 .$$

By Karel [6, p. 182, Corollary of Lemma 2.4], if $\nu < k$, (the case $\nu = k$ is obvious),

$$\sigma'_\nu = p^{8(2m-\zeta)}p^{4(k-\nu)}\Big(\sum_{i=0}^{k-\nu} p^{3i} - \sum_{i=0}^{k-\nu-1} p^{3i-1}\Big) .$$

Substituting this into (4.17) yields

$$p^{-8m-k}\beta_m(T;Z) = \sum_{\nu=0}^{k} p^{3\nu}p^{-4\nu}\Big(\sum_{i=0}^{k-\nu} p^{3i} - \sum_{i=0}^{k-\nu-1} p^{3i-1}\Big)$$

$$
\begin{aligned}
&\quad - \sum_{\nu=1}^{k} p^{3\nu-4} p^{-4\nu} \Big(\sum_{i=0}^{k-\nu} p^{3i} - \sum_{i=0}^{k-\nu-1} p^{3i-1} \Big) \\
&= \sum_{\nu=0}^{k} p^{-\nu} \sum_{i=0}^{k-\nu} p^{3i} - \sum_{\nu=0}^{k-1} p^{-\nu} \sum_{i=0}^{k-\nu-1} p^{3i-1} \\
&\quad - \sum_{\nu=1}^{k} p^{-\nu-4} \sum_{i=0}^{k-\nu} p^{3i} + \sum_{\nu=1}^{k-1} p^{-\nu-4} \sum_{i=0}^{k-\nu-1} p^{3i-1} .
\end{aligned}
$$

Here

$$
\sum_{\nu=0}^{k-1} p^{-\nu} \sum_{i=0}^{k-\nu-1} p^{3i-1} = \sum_{\mu=1}^{k} p^{-\mu+1} \sum_{i=0}^{k-\mu} p^{3i-1} = \sum_{\mu=1}^{k} p^{-\mu} \sum_{i=0}^{k-\mu} p^{3i} ,
$$

$$
\sum_{\nu=1}^{k-1} p^{-\nu-4} \sum_{i=0}^{k-\nu-1} p^{3i-1} = \sum_{\nu=2}^{k} p^{-\nu-4} \sum_{i=0}^{k-\nu} p^{3i} .
$$

Therefore we have

$$
p^{-8m-k} \beta_m(T; Z) = \sum_{i=0}^{k} p^{3i} - \sum_{i=0}^{k-1} p^{3i-5} .
$$

Because

$$
v^{m-1}(p^{-1}Z; p^{-(m-1)}Q(p^{-1}Z)) = v^m(Z; p^{-m}Q(Z)) - 1 = k - 1 ,
$$

we can write

$$
\beta_m(T; Z) = p^{4m}(\beta'_m(T; Z) - \beta'_{m-1}(T; p^{-1}Z)) ,
$$

where

$$
\beta'_m(T; Z) = \begin{cases} p^{4m+4k} \sum_{i=0}^{k} p^{-3i} , & \text{if } Z \in \Lambda_p , \\ 0 , & \text{otherwise} . \end{cases}
$$

So

$$
\begin{aligned}
\alpha_m(T) &= \sum_{Z \bmod p^m} \omega_m^{(T,Z)} \beta_m(T; Z) \\
&= \sum_{Z \bmod p^m} \omega_m^{(T,Z)} p^{4m} \beta'_m(T; Z)
\end{aligned}
$$

$$- \sum_{Z \bmod p^m} \omega_m^{(T,Z)} p^{4m} \beta'_{m-1}(T; p^{-1}Z) .$$

Here $\beta'_m(T; p^{-1}Z) = 0$ unless $Z \equiv 0 \mod p$. So

$$\sum_{Z \bmod p^m} \omega_m^{(T,Z)} p^{4m} \beta'_{m-1}(T; p^{-1}Z) = \sum_{Z \bmod p^{m-1}} \omega_{m-1}^{(T,Z)} p^{4m} \beta'_{m-1}(T; Z) .$$

Therefore we write

$$\alpha_m(T) = p^{4m}(\alpha'_m(T) - \alpha'_{m-1}(T)) ,$$

where

$$\alpha'_m(T) = \sum_{Z \bmod p^m} \omega_m^{(T,Z)} \beta'_m(T; Z) .$$

Hence

$$S_p(T,s) = (1-p^{-s})(1-p^{4-s}) \sum_{m=0}^{\infty} \alpha'_m(T) p^{(4-s)m} \tag{4.18}$$

$$\alpha'_m(T) = \sum_{Z \bmod p^m} \omega_m^{(T,Z)} \beta'_m(T; Z) = p^{4m} \sum_{k=0}^{m} c_k \psi_k ,$$

where $c_k = p^k \sum_{i=0}^{k} p^{3i}$ and

$$\psi_k = \sum_{\substack{Z \bmod p^m \\ v^m(Z; p^{-m}Q(Z)) = k}} \omega_m^{(T,Z)} .$$

In particular, any such Z satisfies $v^m(Z) \geq k$, so we may replace Z by $p^k Z$. Since $v^m(p^k Z; p^{2k-m}Q(Z)) = k$ is equivalent to

$$v^{m-k}(Z; p^{k-m}Q(Z)) = 0 ,$$

$\psi_k = \sum_Z \omega_{m-k}^{(T,Z)}$, where $Z \bmod p^{m-k}$ satisfies $Q(Z) \equiv 0 \mod p^{m-k}$ and either $Z \not\equiv 0 \mod p$ or $Q(Z) \not\equiv 0 \mod p^{m+1-k}$. Thus, $\psi_k = \psi'_{m-k} - \psi'_{m-1-k}$ $(k < m)$, $\psi_m = \psi'_0$, where

$$\psi_\nu = \sum_{\substack{Z \bmod p^\nu \\ Q(Z) \equiv 0 \bmod p^\nu}} \omega_\nu^{(T,Z)} .$$

Hence

$$\alpha'_m(T) = p^{4m}(\alpha''_m(T) - \alpha''_{m-1}(T)), \tag{4.19}$$

where

$$\alpha''_m(T) = \sum_{k=0}^{m} c_k \psi'_{m-k}.$$

Let $T = \begin{pmatrix} T' & 0 \\ 0 & 0 \end{pmatrix}$, $T' = \begin{pmatrix} t' & 0 \\ 0 & t'' \end{pmatrix}$. Then $\psi'_\nu(T) = \alpha_\nu(T')$, ($2 \times 2$ case). By Karel [6, p. 189, (6.4) (2×2 case)],

$$\alpha_\nu(T') = p^{4\nu}(\alpha'_\nu(T') - \alpha'_{\nu-1}(T')).$$

So

$$\alpha''_m(T) = \sum_{k=0}^{m} c_k\, p^{4(m-k)}(\alpha'_{m-k}(T') - \alpha'_{m-k-1}(T')).$$

Let

$$c_k = p^{4k}\, b_k\,, \quad b_k = \sum_{i=0}^{k} p^{-3i} \quad \text{and} \quad b_{-1} = 0.$$

Then

$$\begin{aligned} \alpha''_m(T) &= \sum_{k=0}^{m} p^{4m}\, b_k\, \alpha'_{m-k}(T') - \sum_{k=0}^{m-1} p^{4m}\, b_k\, \alpha'_{m-k-1}(T') \\ &= p^{4m} \sum_{j=0}^{m} \alpha'_{m-j}(T')(b_j - b_{j-1}) \\ &= p^{4m} \sum_{j=0}^{m} \alpha'_{m-j}(T')\, p^{-3j}\,. \end{aligned}$$

Therefore by (4.18), (4.19) and the fact that

$$\sum_{\mu=0}^{\infty} p^{(12-s)\mu}\, \alpha'_\mu(T') = r_p(T', s-8)$$

by Karel [6, p. 189, (6.5)],

$$(1-p^{-s})^{-1}(1-p^{4-s})^{-1} S_p(T,s) = \sum_{m=0}^{\infty} \alpha'_m(T)\, p^{(4-s)m}$$

$$= \sum_{m=0}^{\infty} p^{(8-s)m}(\alpha''_m(T) - \alpha''_{m-1}(T))$$

$$= (1-p^{8-s}) \sum_{m=0}^{\infty} p^{(8-s)m} \alpha''_m(T)$$

$$= (1-p^{8-s}) \sum_{m=0}^{\infty} p^{(8-s)m} p^{4m} \sum_{j=0}^{m} p^{-3j} \alpha'_{m-j}(T')$$

$$= (1-p^{8-s}) \sum_{j=0}^{\infty} p^{-3j} \sum_{m=j}^{\infty} p^{(12-s)m} \alpha'_{m-j}(T')$$

(set $m-j=\mu$)

$$= (1-p^{8-s}) \sum_{j=0}^{\infty} p^{-3j} p^{(12-s)j} \sum_{\mu=0}^{\infty} p^{(12-s)\mu} \alpha'_{\mu}(T')$$

$$= (1-p^{8-s})\, r_p(T', s-8) \sum_{j=0}^{\infty} (p^{9-s})^j$$

$$= \frac{1-p^{8-s}}{1-p^{9-s}} \frac{S_p(T', s-8)}{(1-p^{8-s})(1-p^{12-s})}\,.$$

Therefore

$$S_p(T,s) = \frac{(1-p^{-s})(1-p^{4-s})}{(1-p^{9-s})(1-p^{12-s})}\, S_p(T', s-8)\,.$$

(i) $T = 0$.

$$S_p(0,s) = \frac{(1-p^{-s})(1-p^{4-s})}{(1-p^{9-s})(1-p^{12-s})}\, S_p(0^{2\times 2}, s-8)$$

$$= \frac{(1-p^{-s})(1-p^{4-s})(1-p^{8-s})}{(1-p^{9-s})(1-p^{13-s})(1-p^{17-s})}\,.$$

(ii) $T = \begin{pmatrix} t & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad t \neq 0\,.$

$$S_p(T,s) = \frac{(1-p^{-s})(1-p^{4-s})}{(1-p^{9-s})(1-p^{12-s})}\, S_p(T', s-8)$$

$$= \frac{(1-p^{-s})(1-p^{4-s})(1-p^{8-s})}{(1-p^{9-s})(1-p^{13-s})(1-p^{16-s})}\, S_p(t, s-16)\,.$$

(iii) $T = \begin{pmatrix} t & 0 & 0 \\ 0 & t' & 0 \\ 0 & 0 & 0 \end{pmatrix}$.

$$S_p(T,s) = \frac{(1-p^{-s})(1-p^{4-s})}{(1-p^{9-s})(1-p^{12-s})} S_p(T', s-8)$$
$$= \frac{(1-p^{-s})(1-p^{4-s})(1-p^{8-s})}{(1-p^{9-s})} f_T^p(p^{13-s}),$$

where $f_T^p(p^{13-s})$ is a polynomial. This completes the proof of the theorem.

5. Proof of Theorem A.

Because of the inequality (3.3), the Fourier expansion (2.2) converges uniformly on compact subsets of $\mathbb{C}$. Since each term in the series can be continued as a meromorphic function in s, it follows that $E_{k,s}(Z)$ can be continued as a meromorphic function in s to the whole s-plane. And in particular we can take the limit $s \longrightarrow 0$ term by term.

1) $E_{k,s}^{(1)}(Z)$. By (2.2), (3.9) and (4.2) and the fact that

$$\mu(\mathfrak{J}_{\mathbb{R}}^{(1)}/\Lambda_1) = \mu(\mathbb{R}/\mathbb{Z}) = 1,$$

we have

$$a(t,y,s) = i^{-k}\, 2^{k+1}\, \pi^{k+s/2}\, \Gamma(k+\frac{s}{2})^{-1}\, t^{k+s/2-1}\, y^{-s/2}$$
(5.1.a)
$$\cdot\, \omega_1(2\pi y, t; k+\frac{s}{2}, \frac{s}{2}) \frac{1}{\zeta(k+s)} \Big(\sum_{a|t} a^{1-k-s} \Big),$$

if $t > 0$,

$$a(t,y,s) = i^{-k}\, 2^{1-k}\, \pi^{s/2}\, \Gamma(\frac{s}{2})^{-1}\, y^{-k-s/2}\, |t|^{s/2-1}$$
(5.1.b)
$$\cdot\, \omega_1(2\pi y, t; k+\frac{s}{2}, \frac{s}{2}) \frac{1}{\zeta(k+s)} \Big(\sum_{a||t|} a^{1-k-s} \Big),$$

if $t < 0$, and

$$a(t,y,s) = i^{-k}\, 2^{2-k-s}\, \pi\, y^{-1}\, \Gamma(k+s-1)\, \Gamma(k+\frac{s}{2})^{-1}$$
(5.1.c)
$$\cdot\, \Gamma(\frac{s}{2})^{-1} \frac{\zeta(k+s-1)}{\zeta(k+s)},$$

if $t = 0$. Here we use the fact that $\Gamma(s)$ function has simple poles only at $s = 0, -1, -2, \ldots$ and the residue at $s = -k$ is $1/((-1)^k k!)$ and we have

$$\zeta(2k) = (2\pi)^{2k} B_{2k}/(2(2k)!), \qquad \zeta(-2k) = 0,$$
$$\zeta(-(2k-1)) = (-1)^k B_{2k}/(2k), \qquad \zeta(0) = -1/2,$$

where B_k are Bernoulli numbers. Also from (3.10), we have

$$\omega_1(2\pi y, t; k, 0) = 2^{-1} e^{-2\pi y t} \qquad \text{if } t > 0.$$

Therefore, letting $s \longrightarrow 0$, we have

$$a(t,y,s) = \begin{cases} \dfrac{i^{-k}\, 2k}{B_k} t^{k-1} e^{-2\pi y t} \displaystyle\sum_{a|t} a^{1-k}, & \text{if } t > 0, \\ 0, & \text{if } t < 0, \\ 0, & \text{if } t = 0 \text{ and } k > 2, \\ -\dfrac{1}{\pi B_2 y}, & \text{if } t = 0 \text{ and } k = 2. \end{cases}$$

2) $E^{(2)}_{k,s}(Z)$. Because of Γ-factors in ξ_2, we can easily see that if $T \in V(p,q,r)$, $q > 0$, then $a(T,Y,s) \longrightarrow 0$ as $s \longrightarrow 0$ for all k. Therefore it suffices to consider the cases $q = 0$. By (2.2), (3.9) and (4.3) and the fact that $\mu(\mathfrak{J}^{(2)}_{\mathbb{R}}/\Lambda_2) = \mu(\mathfrak{C}/\mathfrak{o}) = 2^{-4}$, we have the following three cases:

$$\begin{aligned} a(T,Y,s) = {} & i^{-2k}\, 2^{2+2k}\, \pi^{2k+s-4}\, \Gamma(k + \frac{s}{2})^{-1}\, \Gamma(k + \frac{s}{2} - 4)^{-1} \\ & \cdot \det T^{k+s/2-5} \det Y^{-s/2}\, \omega_2(2\pi Y, T; k + \frac{s}{2}, \frac{s}{2}) \\ & \cdot \frac{1}{\zeta(k+s)\zeta(k+s-4)} \prod_{p|\det T} f^p_T(p^{5-k-s}), \end{aligned} \tag{5.2.a}$$

if $T > 0$;

$$\begin{aligned} a(T,Y,s) = {} & i^{-2k}\, 2^{7-s}\, \pi^{k+\frac{s}{2}-3}\, \Gamma(k+s-5) \\ & \cdot \Gamma(k+s/2)^{-1}\, \Gamma(k + \frac{s}{2} - 4)^{-1}\, \Gamma(\frac{s}{2})^{-1} \end{aligned}$$

$$\begin{aligned}(5.2.b)\qquad &\cdot \det Y^{5-k-s}\,\delta_+(TY)^{k+s/2-5}\\ &\cdot \omega_2(2\pi Y, T; k+\frac{s}{2}, \frac{s}{2})\\ &\cdot \frac{\zeta(k+s-5)}{\zeta(k+s)\zeta(k+s-4)}\Big(\sum_{a|t} a^{9-k-s}\Big),\end{aligned}$$

if $T = \begin{pmatrix} t & 0 \\ 0 & 0 \end{pmatrix}$, $t > 0$; and

$$\begin{aligned}(5.2.c)\qquad a(T,Y,s) &= i^{-2k}\,2^{12-2k-2s}\,\pi^6\,\Gamma(k+s-5)\,\Gamma(k+s-9)\\ &\cdot \Gamma(k+\frac{s}{2})^{-1}\Gamma(k+\frac{s}{2}-4)^{-1}\Gamma(\frac{s}{2})^{-1}\Gamma(\frac{s}{2}-4)^{-1}\\ &\cdot \det Y^{5-k-s}\,\frac{\zeta(k+s-5)\zeta(k+s-9)}{\zeta(k+s)\zeta(k+s-4)},\end{aligned}$$

if $T = 0$.

Now let $s \longrightarrow 0$. Then we have:

If $T = 0$,

$$a(T,Y,s) = \begin{cases} \dfrac{(*)}{\pi^2}(\det Y)^{-1}, & \text{if } k = 6,\\ 0, & \text{otherwise.}\end{cases}$$

If $T > 0$,

$$a(T,Y,s) = \begin{cases} (*)(\det T)^{k-5}\displaystyle\prod_{p|\det T} f_T^p(p^{5-k})\,e^{-2\pi(T,Y)}, & \text{if } k \geq 6,\\ 0, & \text{if } k = 2,4.\end{cases}$$

If $T = \begin{pmatrix} t & 0 \\ 0 & 0 \end{pmatrix}$, $t > 0$,

$$a(T,Y,s) = \begin{cases} \dfrac{(*)}{\pi^5}(\det Y)^{-1}\,\delta_+(TY)\,\omega_2(2\pi Y, T; 6, 0)\\ \quad \cdot \Big(\sum_{a|t} a^3\Big), & \text{if } k = 6,\\ 0, & \text{otherwise.}\end{cases}$$

(Here $(*)$ means that it is a rational number.)

3) $E^{(3)}_{k,s}(Z)$. Again because of Γ-factors in ξ_3, we can easily show that if $T \in V(p,q,r)$, $q > 0$, then $a(T,Y,s) \longrightarrow 0$ as $s \longrightarrow 0$ for all k. Therefore it suffices to consider the cases $q = 0$. By (2.2), (3.9), (4.4) and the fact that

$$\mu(\mathfrak{J}^{(3)}_{\mathbb{R}}/\Lambda_3) = \mu(\mathfrak{C}^3/\mathfrak{o}^3) = 2^{-12},$$

we have the following four cases

$$\begin{aligned} a(T,Y,s) = {} & i^{-3k}\, 2^{3+3k}\, \pi^{3k+s/2-12}\, \Gamma(k+\tfrac{s}{2})^{-1} \\ & \cdot \Gamma(k+\tfrac{s}{2}-4)^{-1}\Gamma(k+\tfrac{s}{2}-8)^{-1}(\det Y)^{-s/2} \\ & \cdot (\det T)^{k+s/2-9}\, \omega_3(2\pi Y, T; k+\tfrac{s}{2}, \tfrac{s}{2}) \\ & \cdot \frac{1}{\zeta(k+s)\zeta(k+s-4)\zeta(k+s-8)} \\ & \cdot \prod_{p|\det T} f^p_T(p^{9-k-s}), \end{aligned} \tag{5.3.a}$$

if $T > 0$;

$$\begin{aligned} a(T,Y,s) = {} & i^{-3k}\, 2^{k-s+12}\, \pi^{2k+s-11}\, \Gamma(k+\tfrac{s}{2})^{-1}\, \Gamma(k+\tfrac{s}{2}-4)^{-1} \\ & \cdot \Gamma(k+\tfrac{s}{2}-8)^{-1}\Gamma(\tfrac{s}{2})^{-1}\Gamma(k+s-9)(\det Y)^{9-k-s} \\ & \cdot \delta_+(TY)^{k+s/2-9}\, \omega_3(2\pi Y, T; k+\tfrac{s}{2}, \tfrac{s}{2}) \\ & \cdot \frac{\zeta(k+s-9)}{\zeta(k+s)\zeta(k+s-4)\zeta(k+s-8)} \\ & \cdot \prod_{p|\det T_1} f^p_{T_1}(p^{13-k-s}), \end{aligned} \tag{5.3.b}$$

if $T = \begin{pmatrix} T_1 & 0 \\ 0 & 0 \end{pmatrix}$, $T_1 > 0$;

$$\begin{aligned} a(T,Y,s) = {} & i^{-3k}\, 2^{-k-2s+21}\, \pi^{k+s/2-2}\, \Gamma(k+\tfrac{s}{2})^{-1} \\ & \cdot \Gamma(k+\tfrac{s}{2}-4)^{-1}\Gamma(k+\tfrac{s}{2}-8)^{-1}\Gamma(\tfrac{s}{2})^{-1} \\ & \cdot \Gamma(\tfrac{s}{2}-4)^{-1}\Gamma(k+s-9)\Gamma(k+s-13) \end{aligned}$$

$$\cdot (\det Y)^{9-k-s}\,\delta_+(TY)^{k+s/2-9}\,\omega_3(2\pi Y, T; k+\frac{s}{2}, \frac{s}{2})$$
$$\cdot \frac{\zeta(k+s-9)\zeta(k+s-13)}{\zeta(k+s)\zeta(k+s-4)\zeta(k+s-8)}$$
$$\cdot \Big(\sum_{a|t} a^{17-k-s}\Big), \tag{5.3.c}$$

if $T = \begin{pmatrix} t & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$, $t > 0$; and also

$$\begin{aligned} a(T,Y,s) = {} & i^{-3k}\,2^{-3k-3s+30}\,\pi^{15}\,\Gamma(k+\frac{s}{2})^{-1} \\ & \cdot \Gamma(k+\frac{s}{2}-4)^{-1}\Gamma(k+\frac{s}{2}-8)^{-1}\Gamma(\frac{s}{2})^{-1} \\ & \cdot \Gamma(\frac{s}{2}-4)^{-1}\Gamma(\frac{s}{2}-8)^{-1}\Gamma(k+s-9) \\ & \cdot \Gamma(k+s-13)\Gamma(k+s-17)\,(\det Y)^{9-k-s} \\ & \cdot \frac{\zeta(k+s-9)\zeta(k+s-13)\zeta(k+s-17)}{\zeta(k+s)\zeta(k+s-4)\zeta(k+s-8)}, \end{aligned} \tag{5.3.d}$$

if $T = 0$.

Now let $s \longrightarrow 0$. Then we have:
If $T = 0$,

$$a(T,Y,s) = \begin{cases} \dfrac{(*)}{\pi^3}(\det Y)^{-1}, & \text{if } k = 10, \\ 0, & \text{otherwise.} \end{cases}$$

If $T = \begin{pmatrix} t & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$,

$$a(T,Y,s) = \begin{cases} \dfrac{(*)}{\pi^{10}}(\det Y)^{-1}\,\delta_+(TY)\,\omega_3(2\pi Y, T; 10, 0) \\ \quad \cdot \Big(\sum_{a|t} a^7\Big), & \text{if } k = 10, \\ 0, & \text{otherwise.} \end{cases}$$

If $T = \begin{pmatrix} T_1 & 0 \\ 0 & 0 \end{pmatrix}$, $T_1 > 0$,

$$a(T,Y,s) = \begin{cases} \dfrac{(*)}{\pi^9}(\det Y)^{-1}\,\delta_+(TY)\,\omega_3(2\pi Y, T; 10, 0) \\ \quad \cdot \displaystyle\prod_{p|\det T_1} f^p_{T_1}(p^3), \qquad \text{if } k = 10, \\ \\ 0, \qquad\qquad\qquad\qquad \text{otherwise}. \end{cases}$$

If $T > 0$,

$$a(T,Y,s) = \begin{cases} (*)(\det T)^{k-9} \displaystyle\prod_{p|\det T} f^p_T(p^{9-k})\, e^{-2\pi(T,Y)}, & \text{if } k \geq 10, \\ \\ 0, & \text{if } k < 10. \end{cases}$$

(Here $(*)$ means that it is a rational number.)

Therefore we can summarize our results as follows:

1) $E_{k,s}(Z)$ is finite at $s = 0$ for all k,
2) $E_{k,0}(Z)$ is holomorphic in Z unless $k = 2, 6, 10$,
3) $E_{k,0}(Z)$ is a modular form of weight k with rational Fourier coefficients unless $k = 2, 6, 10$,
4) $E_{4,0}(Z)$ and $E_{8,0}(Z)$ are singular modular forms,

$$E_{4,0}(Z) = 1 + 240 \sum_{\mu\iota_{(1)} \in \mathfrak{I}_o \iota_{(1)} N_{0\mathbb{Q}}/N_{0\mathbb{Q}}} \sum_{\substack{t \in \mathbb{Z} \\ t>0}} \sigma_3(t)\, e^{2\pi i t (Z\cdot\mu)_1},$$

where $\sigma_3(t) = \sum_{a|t} a^3$. In Section 6, we show that the summation $\mu\iota_{(1)} \in \mathfrak{I}_o \iota(1) N_{0\mathbb{Q}}/N_{0\mathbb{Q}}$ is equivalent to $\mu \in \mathfrak{I}_o/(\mathcal{P}_1)_o$. Therefore

$$E_{4,0}(Z) = 1 + 240 \sum_{T \in \mathfrak{J}_o^+,\, \operatorname{rank} T = 1} \sigma_3(\Delta(T))\, e^{2\pi i (T,Z)},$$

where $\Delta(T)$ is as in Karel [6, p. 186]. We also consider the Mellin transform (see Section 6) of $E_{4,0}(Z)$ just like θ-function in order to obtain a functional equation of "Epstein zeta function".

6. Proof of Theorem B.

In this section we prove Theorem B which is a Nagaoka's conjecture on the functional equation of the Eisenstein series. But we have a slightly different functional equation. In the case of the group $Sp_2(\mathbb{Z})$ acting on the Siegel upper half-space of degree 2, Kaufhold [9] obtained a functional equation of an Eisenstein series. We follow his procedure.

From the Fourier expansion of $E_{0,s}(Z)$, we can decompose $E_{0,s}(Z)$ as follows:

$$E_{0,s}(Z) = \Phi_0(s,Z) + \Phi_1(s,Z) + \Phi_2(s,Z) + \Phi_3(s,Z)\,,$$

where

$$\begin{aligned}\Phi_0(s,Z) = 1 \;&+ \sum_{\mu\iota_{(1)}\in\mathfrak{I}_\circ\iota_{(1)}N_{0\mathbb{Q}}/N_{0\mathbb{Q}}} \frac{1}{\mu_1}\xi_1((\mu^{*-1}Y)_1,0;\frac{s}{2},\frac{s}{2})\,S(0^{1\times 1},s)\\ &+ \sum_{\mu\iota_{(2)}\in\mathfrak{I}_\circ\iota_{(2)}N_{0\mathbb{Q}}/N_{0\mathbb{Q}}} \frac{1}{\mu_2}\xi_2((\mu^{*-1}Y)_2,0;\frac{s}{2},\frac{s}{2})\,S(0^{2\times 2},s)\\ &+ \frac{1}{\mu_3}\xi_3(Y,0;\frac{s}{2},\frac{s}{2})\,S(0^{3\times 3},s)\,,\end{aligned}$$

$$\begin{aligned}\Phi_1(s,Z) = &\sum_{\mu\iota_{(1)}\in\mathfrak{I}_\circ\iota_{(1)}N_{0\mathbb{Q}}/N_{0\mathbb{Q}}}\ \sum_{t\in\mathbb{Z}-0} \frac{1}{\mu_1}\xi_1((\mu^{*-1}Y)_1,t;\frac{s}{2},\frac{s}{2})\\ &\qquad\cdot S(t,s)\,e^{2\pi it(\mu^{*-1}X)_1}\\ &+ \sum_{\mu\iota_{(2)}\in\mathfrak{I}_\circ\iota_{(2)}N_{0\mathbb{Q}}/N_{0\mathbb{Q}}}\ \sum_{\substack{T\in\Lambda_2\\ \operatorname{rank}T=1}} \frac{1}{\mu_2}\xi((\mu^{*-1}Y)_2,T;\frac{s}{2},\frac{s}{2})\\ &\qquad\cdot S(T,s)\,e^{2\pi i((\mu^{*-1}X)_2,T)}\\ &+ \sum_{\substack{T\in\Lambda_3\\ \operatorname{rank}T=1}} \frac{1}{\mu_3}\xi_3(Y,T;\frac{s}{2},\frac{s}{2})\,S(T,s)\,e^{2\pi i(X,T)}\,,\end{aligned}$$

$$\begin{aligned}\Phi_2(s,Z) = &\sum_{\mu\iota_{(2)}\in\mathfrak{I}_\circ\iota_{(2)}N_{0\mathbb{Q}}/N_{0\mathbb{Q}}}\ \sum_{\substack{T\in\Lambda_2\\ \det T\neq 0}} \frac{1}{\mu_2}\xi_2((\mu^{*-1}Y)_2,T;\frac{s}{2},\frac{s}{2})\\ &\qquad\cdot S(T,s)\,e^{2\pi i((\mu^{*-1}X)_2,T)}\end{aligned}$$

$$+ \sum_{\substack{T\in\Lambda_3\\ \operatorname{rank} T=2}} \frac{1}{\mu_3}\xi_3(Y,T;\frac{s}{2},\frac{s}{2})\,S(T,s)\,e^{2\pi i(X,T)}\,,$$

$$\Phi_3(s,Z) = \sum_{\substack{T\in\Lambda_3\\ \det T\neq 0}} \frac{1}{\mu_3}\xi_3(Y,T;\frac{s}{2},\frac{s}{2})\,S(T,s)\,e^{2\pi i(X,T)}\,.$$

Here $\mu_1 = 1$, $\mu_2 = 2^{-4}$, $\mu_3 = 2^{-12}$.

1) $\Phi_0(s,Z)$. By (3.9), (4.2), (4.3) and (4.4), we have

$$\begin{aligned}\Phi_0(s,Z) = 1 &+ 2^{2-s}\,\pi\,\Gamma(s-1)\Gamma(\frac{s}{2})^{-1}\Gamma(\frac{s}{2})^{-1}\frac{\zeta(s-1)}{\zeta(s)}\\ &\cdot \sum_{\mu\iota_{(1)}\in \mathfrak{I}_0\iota_{(1)}N_{0\mathbb{Q}}/N_{0\mathbb{Q}}} (\mu^{*-1}Y)_1^{1-s}\\ &+ 2^{16-2s}\,\pi^{10}\,\Gamma_2(\frac{s}{2})^{-1}\Gamma_2(s-5)\Gamma_2(\frac{s}{2})^{-1}\frac{\zeta(s-5)\zeta(s-9)}{\zeta(s)\zeta(s-4)}\\ &\cdot \sum_{\mu\iota_{(2)}\in \mathfrak{I}_0\iota_{(2)}N_{0\mathbb{Q}}/N_{0\mathbb{Q}}} \det(\mu^{*-1}Y)_2^{5-s}\\ &+ (\det Y)^{9-s}\,2^{42-3s}\,\pi^{27}\,\Gamma_3(\frac{s}{2})^{-1}\Gamma_3(\frac{s}{2})^{-1}\Gamma_3(s-9)\\ &\cdot \frac{\zeta(s-9)\zeta(s-13)\zeta(s-17)}{\zeta(s)\zeta(s-4)\zeta(s-8)}\,.\end{aligned}$$

Now we use the identities

$$\Gamma(s) = 2^{s-1}\,\pi^{-1/2}\,\Gamma(\frac{s}{2})\Gamma(\frac{s+1}{2})\,,\quad \rho(s) = \pi^{-s/2}\,\Gamma(\frac{s}{2})\,\zeta(s)\,.$$

Then we have

$$\begin{aligned}\Phi_0(s,Z) = 1 &+ \sum_{\mu\iota_{(1)}\in \mathfrak{I}_0\iota_{(1)}N_{0\mathbb{Q}}/N_{0\mathbb{Q}}} (\mu^{*-1}Y)_1^{1-s}\,\frac{\rho(s-1)}{\rho(s)}\\ &+ \sum_{\mu\iota_{(2)}\in \mathfrak{I}_0\iota_{(2)}N_{0\mathbb{Q}}/N_{0\mathbb{Q}}} \det(\mu^{*-1}Y)_2^{5-s}\,\frac{\rho(s-5)\rho(s-9)(s-6)(s-8)}{\rho(s)\rho(s-4)(s-2)(s-4)}\\ &+ (\det Y)^{9-s}\,\frac{\rho(s-9)\rho(s-13)\rho(s-17)(s-14)(s-16)}{\rho(s)\rho(s-4)\rho(s-8)(s-2)(s-4)}\,.\end{aligned}$$

Now let us look at the following series

$$\varphi_1(Y,s) = \sum_{\mu\iota_{(1)}\in\mathfrak{I}_0\iota_{(1)}N_{0\mathbb{Q}}/N_{0\mathbb{Q}}} (\mu^{*-1}Y)_1^{-s},$$
$$\varphi_2(Y,s) = \sum_{\mu\iota_{(2)}\in\mathfrak{I}_0\iota_{(2)}N_{0\mathbb{Q}}/N_{0\mathbb{Q}}} \det(\mu^{*-1}Y)_2^{-s}.$$

By Baily [1, p. 528], $N_{0\mathbb{Q}} = \{g \in \mathfrak{S}_{\mathbb{Q}} : g(0,0,0,\xi') = (0,0,0,\xi''),\ \xi',\xi'' \in \mathbb{Q}\}$. Since $\iota_{(1)} = \iota_{e_1}$, (Here we take $(j) = \{1,2,\ldots,j\}$ while in Baily $(j) = \{3-j+1,\ldots,3\}$.) we can show, by direct calculation,

$$\iota_{(1)}(0,0,0,\xi') = (-\xi' e_1, 0, 0, 0),$$

where $e_1 = \mathrm{diag}(1,0,0)$. Therefore if $\mu_1\iota_{(1)} = \mu_2\iota_{(1)}p$ for $\mu_1,\mu_2 \in \mathfrak{I}_0$ and $p \in N_{0\mathbb{Q}}$, then $\mu_1(e_1,0,0,0) = \mu_2(\xi e_1,0,0,0)$ for some $\xi \in \mathbb{Q}$. *i.e.* $\mu_1 e_1 = \xi\mu_2 e_1$. But $\mu_1,\mu_2 \in \mathfrak{I}_0$ and $\mathfrak{R}_1^+$ is stable under $\mathfrak{I}$, so $\xi = 1$. In the notation of Baily [1, p. 520], $\mu_1^{-1}\mu_2 \in (\mathcal{P}_1)_0$. Therefore

$$\varphi_1(Y,s) = \sum_{\substack{\mu\in\mathfrak{I}_0/\sim \\ \mu_1\sim\mu_2 \text{ iff } \mu_1 e_1 = \mu_2 e_1}} (\mu^{*-1}Y)_1^{-s}.$$

Here $(\mu^{*-1}Y)_1 = (\mu^{*-1}Y, e_1) = (Y, \mu e_1)$ and by Baily [1, Lemma 3.2], we have 1-1 correspondence

$$\begin{aligned} \mathfrak{I}_0/(\mathcal{P}_1)_0 &\longrightarrow \mathfrak{R}_1^+ \cap \mathfrak{J}_0, \quad \text{primitive}. \\ [\mu] &\longmapsto \mu e_1. \end{aligned}$$

Therefore

$$\varphi_1(Y,s) = \sum_{\mu\in\mathfrak{I}_0/\sim} (Y,\mu e_1)^{-s} = \sum_{\substack{X\in\mathfrak{R}_1^+\cap\mathfrak{J}_0 \\ \text{primitive}}} (Y,X)^{-s}.$$

On the other hand, $\iota_{(2)} = \iota_{e_1}\iota_{e_2}$, and so $\iota_{(2)}(0,0,0,\xi') = (0,0,\xi' e_3,0)$. (use the fact that $e_1 \times e_2 = \frac{1}{2}e_3$, $e_1 \times e_3 = \frac{1}{2}e_2$.) So if $\mu_1\iota_{(2)} = \mu_2\iota_{(2)}p$ for $\mu_1,\mu_2 \in \mathfrak{I}_0$ and $p \in N_{0\mathbb{Q}}$, then $\mu_1(0,0,e_3,0) = \mu_2(0,0,\xi e_3,0)$ for some $\xi \in \mathbb{Q}$. *i.e.* $\mu_1^* e_3 = \xi\mu_2^* e_3$. But $(\mu_1^{-1}\mu_2)^* \in \mathfrak{I}_0$, and so $\xi = 1$. In the notation of Baily [1, p. 520], $\mu_1^{-1}\mu_2 \in (\mathcal{P}_3^-)_0$. Therefore

$$\varphi_2(Y,s) = \sum_{\substack{\mu\in\mathfrak{I}_0/\sim \\ \mu_1\sim\mu_2 \text{ iff } \mu_1^* e_3 = \mu_2^* e_3}} \det(\mu^{*-1}Y)_2^{-s}.$$

But $\det(Y)_2 = (Y \times Y)_{33} = (Y \times Y, e_3)$, and so we have

$$\begin{aligned}\varphi_2(Y,s) &= \sum_{\mu \in \mathfrak{J}_0/\sim} (\mu^{*-1}Y \times \mu^{*-1}Y, e_3)^{-s} \\ &= \sum_{\mu \in \mathfrak{J}_0/\sim} (Y \times Y, \mu^* e_3)^{-s} \\ &= \sum_{\substack{X \in \mathfrak{K}_1^+ \cap \mathfrak{J}_0 \\ \text{primitive}}} (Y \times Y, X)^{-s} = \varphi_1(Y \times Y, s)\,.\end{aligned}$$

Here

$$Y^{-1} = \frac{1}{\det Y}\, Y \times Y\,.$$

Therefore

$$\varphi_2(Y,s) = \varphi_1((\det Y)Y^{-1}, s) = (\det Y)^{-s}\, \varphi_1(Y^{-1}, s)\,.$$

Now consider $E_{4,0}(Z)$

$$E_{4,0}(Z) = 1 + 240 \sum_{\mu \iota_{(1)} \in \mathfrak{J}_0 \iota_{(1)} N_{0\mathbb{Q}}/N_{0\mathbb{Q}}} \sum_{\substack{t \in \mathbb{Z} \\ t>0}} \sigma_3(t)\, e^{2\pi i t (Z \cdot \mu)_1}\,,$$

where $\sigma_3(t) = \sum_{a|t} a^3$. As we saw in Section 5, $E_{4,0}(Z)$ is a modular form of weight 4, and so

$$E_{4,0}(-Z^{-1}) = E_{4,0}(Z) \det(-Z)^4\,.$$

Take $Z = i\,rY$, $r > 0$. Then we have

$$E_{4,0}(\frac{i}{r}\, Y^{-1}) = r^{12} (\det Y)^4\, E_{4,0}(irY).$$

We consider the Mellin transform of $E_{4,0}(i\,rY)$: By Fubini's Theorem, we have

$$\begin{aligned}&\int_0^\infty (E_{4,0}(i\,rY) - 1) r^{s-1}\, dr \\ &\qquad = 240 \sum_{\mu \iota_{(1)} \in \mathfrak{J}_0 \iota_{(1)} N_{0\mathbb{Q}}/N_{0\mathbb{Q}}} \sum_{\substack{t \in \mathbb{Z} \\ t>0}} \sigma_3(t) \int_0^\infty e^{-2\pi \operatorname{tr}(\mu^{*-1}Y)_1}\, r^{s-1}\, dr\end{aligned}$$

$$= 240\,(2\pi)^{-s}\,\Gamma(s)\Big(\sum_{\substack{t\in\mathbb{Z}\\ t>0}}\sigma_3(t)t^{-s}\Big)\varphi_1(Y,s) = Z(Y,s)\,.$$

Here by the well-known identity,

$$\sum_{\substack{t\in\mathbb{Z}\\ t>0}}\sigma_3(t)t^{-s} = \zeta(s)\zeta(s-3)\,, \quad \text{for } \mathrm{Re}(s) > 4\,.$$

Let $Z(Y,s) = \int_0^1 + \int_1^\infty$. Then

$$\begin{aligned}
&\int_0^1 (E_{4,0}(i\,rY) - 1)\,r^{s-1}\,dr \\
&\qquad = \int_0^1 (r^{-12}(\det Y)^{-4}E_{4,0}(\frac{i}{r}Y^{-1}) - 1)r^{s-1}\,dr \\
&\qquad = \int_1^\infty ((\det Y)^{-4}u^{12}E_{4,0}(i\,uY^{-1}) - 1)\,u^{-s-1}\,du \quad (\text{set } \frac{1}{r} = u) \\
&\qquad = \int_1^\infty \Big[(\det Y)^{-4}u^{12}(E_{4,0}(i\,uY^{-1}) - 1) \\
&\qquad\qquad + (\det Y)^{-4}u^{12} - 1\Big]u^{-s-1}\,du \\
&\qquad = \int_1^\infty (\det Y)^{-4}(E_{4,0}(i\,uY^{-1}) - 1)\,u^{11-s}\,du \\
&\qquad\quad + \int_1^\infty \Big[(\det Y)^{-4}u^{11-s} - u^{-s-1}\Big]\,du\,.
\end{aligned}$$

Here

$$\int_1^\infty \Big[(\det Y)^{-4}u^{11-s} - u^{-s-1}\Big]\,du = -\frac{1}{s} - \frac{(\det Y)^{-4}}{12-s}\,,$$

for $\mathrm{Re}(s) > 12$. Therefore

$$\begin{aligned}
Z(Y,s) &= -\frac{1}{s} - \frac{(\det Y)^{-4}}{12-s} \\
&+ \int_1^\infty \Big[(E_{4,0}(i\,rY) - 1)r^{s-1} + (\det Y)^{-4}(E_{4,0}(i\,rY^{-1}) - 1)r^{11-s}\Big]\,dr\,.
\end{aligned}$$

Here the integral inside $Z(Y,s)$ is holomorphic in s, and so $Z(Y,s)$ is continued as a meromorphic function in s and satisfies the functional equation:

$$Z(Y^{-1}, 12-s) = Z(Y,s)\,(\det Y)^4.$$

Therefore

$$
\begin{aligned}
&\Psi_0(s,Z)\\
&=(\det Y)^{s/2}\rho(s)\rho(s-4)\rho(s-8)\,(s-2)(s-4)\Phi_0(s,Z)\\
&=(\det Y)^{s/2}\rho(s)\rho(s-4)\rho(s-8)\,(s-2)(s-4)\\
&\quad+(\det Y)^{9-s/2}\rho(s-9)\rho(s-13)\rho(s-17)\,(s-14)(s-16)\\
&\quad+\frac{2^3\pi^2}{240}\,Z(Y,s-1)(\det Y)^{s/2}\rho(s-8)\\
&\quad+Z(Y^{-1},s-5)(\det Y)^{5-s/2}\rho(s-9).
\end{aligned}
\tag{6.1}
$$

So $\Psi_0(s,Z)$ is continued as a meromorphic function in s which has pole of order 1 at $s=0,1,5,8,10,13,17,18$ and by the well-known identity, $\rho(s)=\rho(1-s)$, we have

$$\Psi_0(18-s,Z)=\Psi_0(s,Z)\,.$$

REMARK. $\Psi_0(s,Z)$ has at most a pole of order 1 at $s=9$, but because of the functional equation, $\Psi_0(s,Z)$ cannot have a pole of order 1. So $\Psi_0(s,Z)$ has no pole at $s=9$.

2) $\Phi_3(s,Z)$.

$$\Phi_3(s,Z)=\sum_{\substack{T\in\Lambda_3\\ \det T\neq 0}}\frac{1}{\mu_3}\xi_3(Y,T;\frac{s}{2},\frac{s}{2})\,S(T,s)\,e^{2\pi i(X,T)}\,.$$

We prove that for each T, $\det T\neq 0$,

$$
\begin{aligned}
\chi(s)=(\det Y)^{s/2}\rho(s)\rho(s-4)\rho(s-8)(s-2)(s-4)&\\
\cdot\frac{1}{\mu_3}\xi_3(Y,T;\frac{s}{2},\frac{s}{2})\,S(T,s)\,,&
\end{aligned}
$$

satisfies the functional equation

$$\chi(18-s)=\chi(s)\,.$$

First, we prove that $f_T^p(X)$ satisfies a functional equation

$$X^d f_T^p(X^{-1})=f_T^p(X)\,,$$

where

$$d = \deg(f_T^p) = v_p(\det T)$$

and

$$S_p(T,s) = (1-p^{-s})(1-p^{4-s})(1-p^{8-s})\, f_T^p(p^{9-s}).$$

It suffices to consider the case $T = \mathrm{diag}(t_1,t_2,t_3)$, $t_1|t_2$, $t_2|t_3$ in $\mathbb{Z}_p$. Let $\tau_i = v_p(t_i)$. Then $\tau_1 \le \tau_2 \le \tau_3$.

If $\tau_1 = 0$, then by Karel [8, p. 553],

$$f_T^p(X) = \sum_{k=0}^{\tau_2} (p^4X)^k \frac{1-X^{\tau_3+\tau_2+1-2k}}{1-X}\,.$$

From this expression, we get

$$X^{\tau_2+\tau_3} f_T^p(X^{-1}) = f_T^p(X).$$

Again by Karel [8, p. 553], if $\tau_1 = 0$, we have the formula

$$\begin{aligned}\frac{S(p^mT,s)}{S(T,s)} &= C_0(X^{-1})\,X^{3m} + C_1(X^{-1})\,q^{2m}X^{2m} \\ &\quad + C_1(X)\,q^{2m}X^m + C_0(X),\end{aligned}$$

where $q = p^4$, $X = p^{9-s}$.(see Karel [8, p. 553] for the definition of C_0 and C_1.) Then

$$\begin{aligned}f_{p^mT}^p(X) = f_T^p(X)(&C_0(X^{-1})\,X^{3m} + C_1(X^{-1})\,q^{2m}X^{2m} \\ &+ C_1(X)\,q^{2m}X^m + C_0(X)).\end{aligned}$$

Therefore $X^{3m}X^{\tau_2+\tau_3} f_{p^mT}^p(X^{-1}) = f_{p^mT}^p(X)$ for T with $\tau_1 = 0$. So we proved that

$$X^d f_T^p(X^{-1}) = f_T^p(X) \qquad \text{for all } T.$$

From this functional equation, we have

$$\begin{aligned}\prod_{p|\det T} f_T^p(p^{s-9}) &= \prod_{p|\det T} f_T^p(p^{9-s})(p^{s-9})^{d_p} \\ &= |\det T|^{s-9} \prod_{p|\det T} f_T^p(p^{9-s}),\end{aligned}$$

where

$$|\det T| = \prod_{p|\det T} p^{d_p}.$$

Now in order to prove the functional equation of $\chi(s)$, it suffices to consider the cases $T \in V(3,0,0)$ and $T \in V(2,1,0)$ by Section 3, (6).

(i) $T \in V(3,0,0)$. By (3.9) and (4.4),

$$\begin{aligned}\frac{1}{\mu_3}\xi_3(Y,T;\frac{s}{2},\frac{s}{2})S(T,s) = 2^3\,\pi^{3s/2-12}\,\Gamma(\frac{s}{2})^{-1}\Gamma(\frac{s}{2}-4)^{-1}\Gamma(\frac{s}{2}-8)^{-1}\\ \cdot(\det Y)^{-s/2}(\det T)^{s/2-9}\,\omega_3(2\pi Y,T;\frac{s}{2},\frac{s}{2})\\ \cdot\frac{\prod_{p|\det T} f_T^p(p^{9-s})}{\zeta(s)\zeta(s-4)\zeta(s-8)}.\end{aligned}$$

So we have

$$\begin{aligned}\chi(s) = 2^{21}\,\pi^{-6}\,(\det T)^{s/2-9}\,\omega_3(2\pi Y,T;\frac{s}{2},\frac{s}{2})\\ \cdot\prod_{p|\det T} f_T^p(p^{9-s})(s-2)(s-4)\dots(s-16).\end{aligned}$$

Therefore from the functional equation of ω_3 and f_T^p, we have

$$\chi(18-s) = \chi(s).$$

(ii) $T \in V(2,1,0)$. By (3.9) and (4.4),

$$\begin{aligned}\frac{1}{\mu_3}\xi_3(Y,T;\frac{s}{2},\frac{s}{2})S(T,s) = 2^{35}\,\pi^{3s/2-12}\,\Gamma(\frac{s}{2})^{-1}\Gamma(\frac{s}{2}-4)^{-1}\Gamma(\frac{s}{2})^{-1}\\ \cdot(\det Y)^{9-s}\,\delta_+(TY)^{s/2-7}\,\delta_-(TY)^{s/2-5}\\ \cdot\omega_3(2\pi Y,T;\frac{s}{2},\frac{s}{2})\,S(T,s).\end{aligned}$$

So

$$\begin{aligned}\chi(s) = 2^{33}\,\pi^{-6}\,(\det Y)^{9-s/2}\,\delta_+(TY)^{s/2-7}\,\delta_-(TY)^{s/2-5}\\ \cdot\omega_3(2\pi Y,T;\frac{s}{2},\frac{s}{2})\prod_{p|\det T} f_T^p(p^{9-s}).\end{aligned}$$

Therefore from the functional equation of ω_3 and and the fact that

$$\delta_+(TY)\,\delta_-(TY) = \det Y\;|\det T|,$$

we have again $\chi(18-s)=\chi(s)$. Here $\chi(s)$ is holomorphic for all T and because of the inequality of ω_3 in the Theorem in Section 3,

$$(6.2)\quad \Psi_3(s,Z)=(\det Y)^{s/2}\rho(s)\rho(s-4)\rho(s-8)(s-2)(s-4)\,\Phi_3(s,Z),$$

converges and so is holomorphic in s.

3) $\Phi_2(s,Z)$.

$$\Phi_2(s,Z)=\sum_{\mu\iota_{(2)}\in\mathfrak{I}_0\iota_{(2)}N_{0\mathbb{Q}}/N_{0\mathbb{Q}}}\ \sum_{\substack{T\in\Lambda_2\\ \det T\neq 0}}\frac{1}{\mu_2}\xi_2((\mu^{*-1}Y)_2,T;\frac{s}{2},\frac{s}{2})\cdot S(T,s)\,e^{2\pi i((\mu^{*-1}X)_2,T)}$$
$$+\sum_{\substack{T\in\Lambda_3\\ \operatorname{rank}T=2}}\frac{1}{\mu_3}\xi_3(Y,T;\frac{s}{2},\frac{s}{2})\,S(T,s)\,e^{2\pi i(X,T)}.$$

By Baily [1, Lemma 3.2],

$$T=\mu\begin{pmatrix}T_1 & 0\\ 0 & 0\end{pmatrix}$$

runs over all $T\in\Lambda_3$, $\operatorname{rank}T=2$ if T_1 runs over all $T_1\in\Lambda_2$, $\det T\neq 0$ and $\mu\in\mathfrak{I}_0/(P_3^-)_0$. But as we saw in Section 6, (1),

$$\mu\iota_{(2)}\in\mathfrak{I}_0\iota_{(2)}N_{0\mathbb{Q}}/N_{0\mathbb{Q}}\ \text{ if and only if }\ \mu\in\mathfrak{I}_0/(P_3^-)_0\,.$$

Therefore

$$\sum_{\substack{T\in\Lambda_3\\ \operatorname{rank}T=2}}\frac{1}{\mu_3}\xi_3(Y,T;\frac{s}{2},\frac{s}{2})\,S(T,s)\,e^{2\pi i(X,T)}$$
$$=\sum_{\mu}\sum_{\substack{T\in\Lambda_2\\ \det T\neq 0}}\frac{1}{\mu_3}\xi_3\Big(\begin{pmatrix}(\mu^{*-1}Y)_2 & *\\ * & *\end{pmatrix},\begin{pmatrix}T & 0\\ 0 & 0\end{pmatrix};\frac{s}{2},\frac{s}{2}\Big)\cdot S\Big(\begin{pmatrix}T & 0\\ 0 & 0\end{pmatrix},s\Big)\,e^{2\pi i((\mu^{*-1}X)_2,T)},$$

where $\mu\iota_{(2)}\in\mathfrak{I}_0\iota_{(2)}N_{0\mathbb{Q}}/N_{0\mathbb{Q}}$. But there exists $\mu_0\in\mathfrak{I}_{\mathbb{R}}$ such that

$$\mu_0^*\begin{pmatrix}(\mu^{*-1}Y)_2 & *\\ * & *\end{pmatrix}=\begin{pmatrix}(\mu^{*-1}Y)_2 & 0\\ 0 & *\end{pmatrix},\quad \mu_0\begin{pmatrix}T & 0\\ 0 & 0\end{pmatrix}=\begin{pmatrix}T & 0\\ 0 & 0\end{pmatrix}.$$

So

$$\xi_3\left(\begin{pmatrix}(\mu^{*-1}Y)_2 & * \\ * & *\end{pmatrix}, \begin{pmatrix}T & 0\\ 0 & 0\end{pmatrix}; \frac{s}{2}, \frac{s}{2}\right) = \xi_3\left(\begin{pmatrix}(\mu^{*-1}Y)_2 & 0 \\ 0 & *\end{pmatrix}, \begin{pmatrix}T & 0\\ 0 & 0\end{pmatrix}; \frac{s}{2}, \frac{s}{2}\right).$$

Therefore we have

$$\Phi_2(s,Z) = \sum_{\mu\iota_{(2)}\in \mathfrak{I}_0\iota_{(2)}N_{0\mathbb{Q}}/N_{0\mathbb{Q}}} \sum_{\substack{T\in\Lambda_2\\ \det T\neq 0}} a(T,s)\, e^{2\pi i((\mu^{*-1}X)_2,T)},$$

where

$$a(T,s) = \frac{1}{\mu_2}\xi_2((\mu^{*-1}Y)_2, T; \frac{s}{2}, \frac{s}{2})\, S(T,s) + \frac{1}{\mu_3}\xi_3\left(\begin{pmatrix}(\mu^{*-1}Y)_2 & 0 \\ 0 & *\end{pmatrix}, \begin{pmatrix}T & 0\\ 0 & 0\end{pmatrix}; \frac{s}{2}, \frac{s}{2}\right) S\left(\begin{pmatrix}T & 0\\ 0 & 0\end{pmatrix}, s\right).$$

We show that, for all $T \in \Lambda_2$, $\det T \neq 0$,

$$\begin{aligned}\chi(s) &= (\det Y)^{s/2}\rho(s)\rho(s-4)\rho(s-8)\,(s-2)(s-4)\,a(T,s)\\ &= \chi(18-s).\end{aligned}$$

It suffices to consider the cases $T \in V(2,0)$ and $T \in V(1,1)$ by Section 3, (6).

(i) $T \in V(2,0)$. By (3.5), (3.9), (4.3) and (4.4), we have

$$\begin{aligned}a(T,s) = {} & 2^6\,\pi^{s-4}\,\Gamma(\tfrac{s}{2})^{-1}\Gamma(\tfrac{s}{2}-4)^{-1}\det(\mu^{*-1}Y)_2^{-s/2}(\det T)^{s/2-5}\\ & \cdot \omega_2(2\pi(\mu^{*-1}Y)_2, T; \tfrac{s}{2}, \tfrac{s}{2})\frac{1}{\zeta(s)\zeta(s-4)}\prod_{p|\det T} f_T^p(p^{5-s})\\ & + 2^{16-s}\,\pi^{s-3}\,\Gamma(\tfrac{s}{2})^{-1}\Gamma(\tfrac{s}{2})^{-1}\Gamma(\tfrac{s}{2}-4)^{-1}\Gamma(\tfrac{s}{2}-8)^{-1}\\ & \cdot \Gamma(s-9)\,(\det Y)^{9-s}\det(\mu^{*-1}Y)_2^{s/2-9}(\det T)^{s/2-9}\\ & \cdot \omega_2(2\pi(\mu^{*-1}Y)_2, T; \tfrac{s}{2}-4, \tfrac{s}{2}-4)\\ & \cdot \frac{\zeta(s-9)}{\zeta(s)\zeta(s-4)\zeta(s-8)}\prod_{p|\det T} f_T^p(p^{13-s}).\end{aligned}$$

By using the identities,

$$\Gamma(s) = 2^{s-1}\,\pi^{-1/2}\,\Gamma(\frac{s}{2})\Gamma(\frac{s+1}{2}),$$
$$\rho(s) = \pi^{-s/2}\,\Gamma(\frac{s}{2})\,\zeta(s),$$

we can write

$$\begin{aligned} a(T,s) = {} & 2^4\,\pi^{-2}\det(\mu^{*-1}Y)_2^{-s/2}(\det T)^{s/2-5}\,\omega_2(2\pi(\mu^{*-1}Y)_2, T; \frac{s}{2}, \frac{s}{2}) \\ & \cdot \frac{(s-6)(s-8)}{\rho(s)\rho(s-4)} \prod_{p|\det T} f_T^p(p^{5-s}) \\ & + 2^4\,\pi^{-4}\,(\det Y)^{9-s}(\det T)^{s/2-9}\det(\mu^{*-1}Y)_2^{s/2-9} \\ & \cdot \omega_2(2\pi(\mu^{*-1}Y)_2, T; \frac{s}{2}-4, \frac{s}{2}-4)\,\frac{\rho(s-9)}{\rho(s)\rho(s-4)\rho(s-8)} \\ & \cdot \frac{(s-10)(s-12)(s-14)(s-16)}{(s-2)(s-4)} \prod_{p|\det T} f_T^p(p^{13-s}). \end{aligned}$$

Therefore

$$\begin{aligned} \chi(s) = {} & 2^4\,\pi^{-2}\,(\det Y)^{s/2}\det(\mu^{*-1}Y)_2^{-s/2} \\ & \cdot \omega_2(2\pi(\mu^{*-1}Y)_2, T; \frac{s}{2}, \frac{s}{2})\,\rho(s-8) \\ & \cdot (s-2)(s-4)(s-6)(s-8) \\ & \cdot (\det T)^{s/2-5} \prod_{p|\det T} f_T^p(p^{5-s}) \\ & + 2^4\,\pi^{-2}\,(\det Y)^{9-s/2}\det(\mu^{*-1}Y)_2^{s/2-9} \\ & \cdot \omega_2(2\pi(\mu^{*-1}Y)_2, T; \frac{s}{2}-4, \frac{s}{2}-4)\,\rho(s-9) \\ & \cdot (s-10)(s-12)(s-14)(s-16) \\ & \cdot (\det T)^{s/2-9} \prod_{p|\det T} f_T^p(p^{13-s}). \end{aligned}$$

By using the functional equation of ω_2, ρ and

$$(\det T)^{s-5} \prod_{p|\det T} f_T^p(p^{5-s}) = \prod_{p|\det T} f_T^p(p^{s-5}),$$

we have $\chi(18-s)=\chi(s)$.

(ii) $T\in V(1,1)$. By (3.5), (3.9), (4.3) and (4.4), we have

$$\begin{aligned}
a(T,s) &= 2^{10}\,\pi^{s-4}\,\Gamma(\tfrac{s}{2})^{-1}\Gamma(\tfrac{s}{2})^{-1}\det(\mu^{*-1}Y)_2^{2-s/2}|\det T|^{s/2-3}\\
&\quad\cdot\omega_2(2\pi(\mu^{*-1}Y)_2,T;\tfrac{s}{2},\tfrac{s}{2})\,S(T,s)\\
&\quad+2^{20-2s}\,\pi^{s-3}\,\Gamma(\tfrac{s}{2})^{-1}\Gamma(\tfrac{s}{2}-4)^{-1}\Gamma(\tfrac{s}{2})^{-1}\Gamma(\tfrac{s}{2}-4)^{-1}\Gamma(s-9)\\
&\quad\cdot(\det Y)^{9-s}\det(\mu^{*-1}Y)_2^{s/2-7}\,|\det T|^{s/2-7}\\
&\quad\cdot\omega_2(2\pi(\mu^{*-1}Y)_2,T,\tfrac{s}{2}-4,\tfrac{s}{2}-4)\,S\left(\begin{pmatrix}T&0\\0&0\end{pmatrix},s\right)\\
&=2^{12}\,\pi^{-2}\det(\mu^{*-1}Y)_2^{2-s/2}\omega_2(2\pi(\mu^{*-1}Y)_2,T;\tfrac{s}{2},\tfrac{s}{2})\\
&\quad\cdot\frac{(\det T)^{s/2-3}\prod_{p|\det T}f_T^p(p^{5-s})}{\rho(s)\rho(s-4)(s-2)(s-4)}\\
&\quad+2^{12}\pi^{-2}(\det Y)^{9-s}\det(\mu^{*-1}Y)_2^{s/2-7}\\
&\quad\cdot\omega_2(2\pi(\mu^{*-1}Y)_2,T;\tfrac{s}{2},\tfrac{s}{2})\\
&\quad\cdot\frac{\rho(s-9)\,|\det T|^{s/2-7}\prod_{p|\det T}f_T^p(p^{13-s})}{\rho(s)\rho(s-4)\rho(s-8)(s-2)(s-4)}\,.
\end{aligned}$$

Therefore

$$\begin{aligned}
\chi(s) &= 2^{12}\,\pi^{-2}\,(\det Y)^{s/2}\det(\mu^{*-1}Y)_2^{2-s/2}\,\omega_2(2\pi(\mu^{*-1}Y)_2,T;\tfrac{s}{2},\tfrac{s}{2})\\
&\quad\cdot\rho(s-8)\,|\det T|^{s/2-3}\prod_{p|\det T}f_T^p(p^{5-s})\\
&\quad+2^{12}\,\pi^{-2}\,(\det Y)^{9-s/2}\det(\mu^{*-1}Y)_2^{s/2-7}\\
&\quad\cdot\omega_2(2\pi(\mu^{*-1}Y)_2,T;\tfrac{s}{2},\tfrac{s}{2})\,\rho(s-9)\\
&\quad\cdot|\det T|^{s/2-7}\prod_{p|\det T}f_T^p(p^{13-s})\,.
\end{aligned}$$

So again we have $\chi(18-s)=\chi(s)$. Here $\chi(s)$ is a meromorphic function for all T which has a pole of order 1 at $s=8,10$ and because of the inequality of ω_2 in the Theorem in Section 3,

$$(6.3)\quad \Psi_2(s,Z)=(\det Y)^{s/2}\rho(s)\rho(s-4)\rho(s-8)\,(s-2)(s-4)\,\Phi_2(s,Z)\,,$$

converges and so is meromorphic in s.

4) $\Phi_1(s,Z)$. We write

$$\Phi_1(s,Z) = \Phi_1'(s,Z) + \Phi_1''(s,Z),$$

where

$$\Phi_1'(s,Z) = \sum_{\mu\iota_{(1)} \in \mathfrak{I}_0 \iota_{(1)} N_{0\mathbb{Q}}/N_{0\mathbb{Q}}} \sum_{t \in \mathbb{Z}-0} \frac{1}{\mu_1} \xi_1((\mu^{*-1}Y)_1, t; \frac{s}{2}, \frac{s}{2}) \cdot S(t,s)\, e^{2\pi i t (\mu^{*-1}X)_1}$$
$$+ \sum_{\substack{T \in \Lambda_3 \\ \operatorname{rank} T = 1}} \frac{1}{\mu_3} \xi_3(Y, T; \frac{s}{2}, \frac{s}{2})\, S(T,s)\, e^{2\pi i (X,T)}$$

$$\Phi_1''(s,Z) = \sum_{\mu\iota_{(2)} \in \mathfrak{I}_0 \iota_{(2)} N_{0\mathbb{Q}}/N_{0\mathbb{Q}}} \sum_{T \in \Lambda_2 \ \operatorname{rank} T = 1} \frac{1}{\mu_2} \xi_2((\mu^{*-1}Y)_2, T; \frac{s}{2}, \frac{s}{2}) \cdot S(T,s)\, e^{2\pi i ((\mu^{*-1}X)_2, T)}.$$

(i) $\Phi_1'(s,Z)$. We have the 1-1 correspondence

$$\{T \in \Lambda_3 : \quad \operatorname{rank} T = 1\} \longleftrightarrow \{\mu\iota_{(1)} \in \mathfrak{I}_0 \iota_{(1)} N_{0\mathbb{Q}}/N_{0\mathbb{Q}}\} \times \{t \in \mathbb{Z} - 0\},$$

by $T = \mu \begin{pmatrix} t & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$. Also there exists $\mu_0 \in \mathfrak{I}_{\mathbb{R}}$ such that

$$\mu_0^* \begin{pmatrix} (\mu^{*-1}Y)_1 & * \\ * & * \end{pmatrix} = \begin{pmatrix} (\mu^{*-1}Y)_1 & 0 \\ 0 & * \end{pmatrix},$$
$$\mu_0 \begin{pmatrix} t & 0 \\ 0 & 0^{2\times 2} \end{pmatrix} = \begin{pmatrix} t & 0 \\ 0 & 0^{2\times 2} \end{pmatrix}.$$

By (3.4), (3.5),

$$\omega_3(\begin{pmatrix} y_1 & 0 \\ 0 & * \end{pmatrix}, \begin{pmatrix} t & 0 \\ 0 & 0^{2\times 2} \end{pmatrix}; \frac{s}{2}, \frac{s}{2}) = 2^{-9}\pi^8 e^{-|t|y_1}\, \omega_1(2y_1; \frac{s}{2} - 8, \frac{s}{2} - 8).$$

Therefore by (3.9), (4.2) and (4.4), we have

$$\Phi_1'(s,Z) = \sum_{\mu\iota_{(1)} \in \mathfrak{I}_0 \iota_{(1)} N_{0\mathbb{Q}}/N_{0\mathbb{Q}}} \sum_{t \in \mathbb{Z}-0} \left(\frac{1}{\mu_1} \xi_1((\mu^{*-1}Y)_1, t; \frac{s}{2}, \frac{s}{2}) S(t,s) \right.$$

$$+\frac{1}{\mu_3}\xi_3\Big(\begin{pmatrix}(\mu^{*-1}Y)_1 & 0\\ 0 & *\end{pmatrix},\begin{pmatrix}t & 0\\ 0 & 0^{2\times 2}\end{pmatrix};\frac{s}{2},\frac{s}{2}\Big)\,S\Big(\begin{pmatrix}t & 0\\ 0 & 0^{2\times 2}\end{pmatrix},s\Big)\Big)$$
$$\cdot e^{2\pi i t(\mu^{*-1}X)_1}$$

$$=\sum_{\mu}\sum_{t\in\mathbb{Z}-0} e^{2\pi i|t|(Z\cdot\mu)_1}\,|t|^{-1}\Big(\pi^{s/2}(\mu^{*-1}Y)_1^{-s/2}|t|^{s/2}S(t,s)$$
$$\cdot\omega_1(4\pi|t|(\mu^{*-1}Y)_1;\frac{s}{2},\frac{s}{2})$$
$$+2^{24-2s}\,\pi^{s/2+6}\,(\det Y)^{9-s}$$
$$\cdot(\mu^{*-1}Y)_1^{s/2-9}\,|t|^{s/2-8}$$
$$\cdot\omega_1(4\pi|t|(\mu^{*-1}Y)_1;\frac{s}{2}-8,\frac{s}{2}-8)$$
$$\cdot\Gamma(\frac{s}{2})^{-1}\Gamma(\frac{s}{2})^{-1}\Gamma(\frac{s}{2}-4)^{-1}$$
$$\cdot\Gamma(\frac{s}{2}-4)^{-1}\Gamma(\frac{s}{2}-8)^{-1}\Gamma(s-9)$$
$$\cdot\Gamma(s-13)\,S(t,s-16)$$
$$\cdot\frac{\zeta(s-9)\zeta(s-13)\zeta(s-16)}{\zeta(s)\zeta(s-4)\zeta(s-8)}\Big).$$

Here we use the identity

$$\beta(s)=|t|^{s/2}\,S(t,s)\,\zeta(s)=\beta(2-s).$$

(*cf.* Kaufhold [9]) Therefore we can write

$$\Phi_1'(s,Z)=\sum_{\mu}\sum_{t\in\mathbb{Z}-0} e^{-2\pi|t|(Z\cdot\mu)_1}\,|t|^{-1}\Big((\mu^{*-1}Y)^{-s/2}\frac{\beta(s)}{\rho(s)}$$
$$\cdot\omega_1(4\pi|t|(\mu^{*-1}Y)_1;\frac{s}{2},\frac{s}{2})$$
$$+(\mu^{*-1}Y)_1^{s/2-9}(\det Y)^{9-s}\omega_1(4\pi|t|(\mu^{*-1}Y)_1;\frac{s}{2}-8,\frac{s}{2}-8)$$
$$\cdot\frac{\beta(s-16)\rho(s-9)\rho(s-13)(s-14)(s-16)}{\rho(s)\rho(s-4)\rho(s-8)(s-2)(s-4)}\Big).$$

From the functional equation of ω_3, $\rho(s)$ and $\beta(s)$,

(6.4) $(\det Y)^{s/2}\rho(s)\rho(s-4)\rho(s-8)\,(s-2)(s-4)\,\Phi_1'(s,Z)=\Psi_1'(s,Z)$,

satisfies the functional equation

$$\Psi_1'(s,Z)=\Psi_1'(18-s,Z).$$

On the other hand, because of the inequality of ω_1 in the Theorem in Section 3, $\Psi_1'(s,Z)$ converges and defines a meromorphic function in s which has a pole of order 1 at $s = 5, 8, 10, 13$.

(ii) $\Phi_1''(s,Z)$. In order to prove the functional equation of $\Phi_1''(s,Z)$, we need to look at an Eisenstein series on the tube domain

$$\mathfrak{T}^{(2)} = \{Z = X + iY : \quad X \in \mathfrak{J}_{\mathbb{R}}^{(2)}, \quad Y > 0\},$$

which is the boundary component of $\mathfrak{T}$. For an element $X = \begin{pmatrix} a & x \\ \bar{x} & b \end{pmatrix} \in \mathfrak{J}^{(2)}$, we let $\det X = ab - N(x)$ and $\operatorname{tr} X = a + b$. We also define

$$\mathcal{I}^{(2)} = \{g \in \mathfrak{J}^{(2)} : \quad \det(gX) \equiv \det(X)\}.$$

Tsao studied an action of a subgroup $\mathcal{G}^{(2)}$ of $\mathcal{G}$ on the boundary component $\mathfrak{T}^{(2)}$ in (Tsao [16, p .254]). $\mathcal{G}^{(2)}$ is isogeneous to SO(10,2). We can define an Eisenstein series $E_{k,s}^{(2)}(Z)$ on $\mathfrak{T}^{(2)}$ in the exactly same way as $E_{k,s}(Z)$ and can get $E_{4,0}^{(2)}(Z)$ which is a holomorphic modular form of weight 4 on $\mathfrak{T}^{(2)}$ and which is given by

$$E_{4,0}^{(2)}(Z) = 1 + 240 \sum_{\mu \in \mathcal{I}_0^{(2)}/(\mathcal{P}_0)_0} \sum_{\substack{t \in \mathbb{Z} \\ t>0}} \sigma_3(t)\, e^{2\pi i t (Z\cdot\mu)_1},$$

where $\mathcal{P}_0$ is the minimal pabolic subgroup of $\mathcal{I}^{(2)}$ which is the stability group of the line $\mathbb{R}e_1'$ where $e_1' = \operatorname{diag}(1,0)$.

REMARK. Recently Eie and Krieg [4] studied modular forms on $\mathfrak{T}^{(2)}$ using Fourier-Jacobi expansion. Especially they obtained $E_{4,0}^{(2)}(Z)$ in terms of theta series

$$E_{4,0}^{(2)}(Z) = \sum_{h \in \mathfrak{o}^2} e^{2\pi i \tau(Z, h\bar{h}^t)}.$$

The equivalence of two expressions come from the well-known formula

$$\#\{a \in \mathfrak{o} : \quad n = N(a)\} = 240\ \sigma_3(a) \qquad \text{for all } n \geq 1.$$

Now consider the following series which is an "Epstein zeta function"

$$\varphi^{(2)}(Y,s) = \sum_{\mu \in \mathcal{I}_0^{(2)}/(\mathcal{P}_0)_0} (\mu^{*-1}Y)_1^{-s} = \sum_{\mu \in \mathcal{I}_0^{(2)}/(\mathcal{P}_0)_0} (Y, \mu e_1')^{-s},$$

for $Y > 0$ and $s \in \mathbb{C}$.
Now we apply the Mellin transform of $E_{4,0}^{(2)}(Z)$ in the exactly same way as $E_{4,0}(Z)$ to get the analytic continuation and a functional equation of $\varphi^{(2)}(Y,s)$: If we let

$$Z^{(2)}(Y,s) = \int_0^\infty r^{s-1}(E_{4,0}^{(2)}(i\,rY) - 1)\,dr\,,$$

then

$$Z^{(2)}(Y,s) = 240\,(2\pi)^{-s}\,\Gamma(s)\,\zeta(s)\,\zeta(s-3)\,\varphi^{(2)}(Y,s)\,,$$
$$Z^{(2)}(Y^{-1}, 8-s) = (\det Y)^4\,Z^{(2)}(Y,s)\,.$$

$Z^{(2)}(Y,s)$ has a pole of order 1 at $s = 0, 8$. Here if $\det Y = 1$, then $Y^{-1} = \begin{pmatrix} b & -x \\ -\bar{x} & a \end{pmatrix}$ for $Y = \begin{pmatrix} a & x \\ \bar{x} & b \end{pmatrix}$. The transformation

$$X = \begin{pmatrix} a & x \\ \bar{x} & b \end{pmatrix} \longmapsto X^* = \begin{pmatrix} b & -x \\ -\bar{x} & a \end{pmatrix},$$

is in $\mathfrak{I}_0^{(2)}$ since it preserves the determinant and the lattice $\mathfrak{J}_0$. Therefore

$$\varphi^{(2)}(Y^*, s) = \varphi^{(2)}(Y,s)\,.$$

So if $\det Y = 1$, we have

$$Z^{(2)}(Y, 8-s) = Z^{(2)}(Y,s)\,.$$

Now let us come back to $\Phi_1''(s,Z)$. We have 1-1 correspondence

$$\{T \in \Lambda_2 : \quad \operatorname{rank} T = 1\} \longleftrightarrow \{\nu \in \mathfrak{I}_0^{(2)}/(\mathcal{P}_0)_0\} \times \{t \in \mathbb{Z} - 0\}\,,$$

by $T = \nu \begin{pmatrix} t & 0 \\ 0 & 0 \end{pmatrix}$ and there exists $\nu' \in \mathfrak{I}_{\mathbb{R}}^{(2)}$ such that

$$\nu'^* \begin{pmatrix} y_1 & * \\ * & * \end{pmatrix} = \begin{pmatrix} y_1 & 0 \\ 0 & * \end{pmatrix}, \quad \nu' \begin{pmatrix} t & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} t & 0 \\ 0 & 0 \end{pmatrix}.$$

Therefore by (3.4) and (4.3), we have (set $(\mu^{*-1}Y)_2 = y$)

$$\frac{1}{\mu_2}\xi_2(\nu^{*-1}y, \begin{pmatrix} t & 0 \\ 0 & 0 \end{pmatrix}; \frac{s}{2}, \frac{s}{2})\; S(\begin{pmatrix} t & 0 \\ 0 & 0 \end{pmatrix}, s)$$

$$= 2^{6-s}\,\pi^{9+s/2}\,\Gamma_2(\frac{s}{2})^{-1}\Gamma_2(\frac{s}{2})^{-1}\Gamma(\frac{s}{2}-4)\Gamma(s-5)$$
$$\cdot (\det y)^{5-s}\,(\nu^{*-1}y)_1^{s/2-5}\,|t|^{s/2-5}\,e^{-2\pi|t|(\nu^{*-1}y)_1}$$
$$\cdot \omega_1(4\pi|t|(\nu^{*-1}y)_1;\frac{s}{2}-4,\frac{s}{2}-4)$$
$$\cdot \frac{\zeta(s-5)\zeta(s-8)}{\zeta(s)\zeta(s-4)}S(t,s-8)\,.$$

Therefore by using the identities

$$\beta(s) = |t|^{s/2}\,S(t,s)\,\zeta(s)\,, \quad \rho(s) = \pi^{-s/2}\,\Gamma(\frac{s}{2})\,\zeta(s)\,,$$

we get

$$\chi(s) = (\det Y)^{s/2}\,\rho(s)\rho(s-4)\rho(s-8)\,(s-2)(s-4)\,\Phi_1''(s,Z)$$
$$= \sum_{\mu\in\mathfrak{I}_0/(\mathcal{P}_3^-)_0}\ \sum_{\nu\in\mathfrak{I}_0^{(2)}/(\mathcal{P}_0)_0}\ \sum_{t\in\mathbb{Z}-0} (\det Y)^{s/2}\det(\mu^{*-1}Y)_2^{5-s}$$
$$\cdot (\nu^{*-1}(\mu^{*-1}Y)_2)_1^{s/2-5}\,e^{-2\pi t(\nu^{*-1}(\mu^{*-1}Y)_2)_1}\,|t|^{-1}$$
$$\cdot \omega_1(4\pi t(\nu^{*-1}(\mu^{*-1}Y)_2)_1;\frac{s}{2}-4,\frac{s}{2}-4)$$
$$\cdot \beta(s-8)\,\rho(s-5)\rho(s-8)\,(s-6)(s-8)\,.$$

Now any element of $\gamma \in \mathfrak{I}_0/(\mathcal{P}_*)_0$ can be written uniquely

$$\gamma = \mu\cdot\nu \quad \mu\in\mathfrak{I}_0/(\mathcal{P}_3^-)_0\,, \quad \nu\in(\mathcal{P}_3^-)_0/(\mathcal{P}_*)_0\,,$$

where $\mathcal{P}_*$ is the minimal parabolic subgroup of $\mathfrak{I}$ which is the stability group of the "flag" $(\mathfrak{J}^{\{1\}},\mathfrak{J}^{\{1,2\}})$ (see Baily [1, p. 520]). But $(\mathcal{P}_3^-)_0/(\mathcal{P}_*)_0$ is identified in a natural way with $\mathfrak{I}_0^{(2)}/(\mathcal{P}_0)_0$. Therefore we can write

$$\chi(s) = \sum_{\gamma\in\mathfrak{I}_0/(\mathcal{P}_*)_0}\ \sum_{t\in\mathbb{Z}-0} (\det Y)^{s/2}\det(\gamma^{*-1}Y)_2^{5-s}\,(\gamma^{*-1}Y)_1^{s/2-5}$$
$$\cdot e^{-2\pi|t|(\gamma^{*-1}Y)_1}\,|t|^{-1}\,\omega_1(4\pi|t|(\gamma^{*-1}Y)_1;\frac{s}{2}-4,\frac{s}{2}-4)$$
$$\cdot \beta(s-8)\,\rho(s-5)\rho(s-8)\,(s-6)(s-8)\,.$$

On the other hand, any element of $\gamma\in\mathfrak{I}_0/(\mathcal{P}_*)_0$ can be written uniquely

$$\gamma = \mu\cdot\nu\,, \quad \mu\in\mathfrak{I}_0/(\mathcal{P}_1)_0\,, \quad \nu\in(\mathcal{P}_1)_0/(\mathcal{P}_*)_0\,.$$

Then $(\gamma^{*-1}Y)_1 = (Y, \gamma e_1) = (Y, \mu\nu e_1) = (Y, \mu e_1)$ since $\nu \in (\mathcal{P}_1)_0$. Therefore

$$\chi(s) = \sum_{\mu \in \mathfrak{I}_0/(\mathcal{P}_1)_0} (\det Y)^{s/2} (\mu^{*-1}Y)^{s/2-5}\, e^{-2\pi|t|(\mu^{*-1}Y)_1}\, |t|^{-1}$$
$$\cdot \omega_1(4\pi|t|(\mu^{*-1}Y)_1; \frac{s}{2}-4, \frac{s}{2}-4)\, \beta(s-8)$$
$$\cdot \rho(s-5)\rho(s-8)\,(s-6)(s-8)$$
$$\cdot \left(\sum_{\nu \in (\mathcal{P}_1)_0/(\mathcal{P}_*)_0} \det((\mu\nu)^{*-1}Y)_2^{5-s} \right).$$

Here $\det((\mu\nu)^{*-1}Y)_2 = (\mu^{*-1}Y \times \mu^{*-1}Y, \nu^* e_3)$ and we have 1-1 correspondence

$$\nu \in (\mathcal{P}_1)_0/(\mathcal{P}_*)_0 \longleftrightarrow \nu^* \in (\mathcal{P}_1^-)_0/(\mathcal{P}_*)_0 .$$

But $(\mathcal{P}_1^-)_0/(\mathcal{P}_*)_0$ can be identified in a natually way with $\mathfrak{I}_0^{(2)}/(\mathcal{P}_0)_0$ and note that

$$\mu^{*-1}Y \times \mu^{*-1}Y = \det(\mu^{*-1}Y)(\mu^{*-1}Y)^{-1} = (\det Y)(\mu^{*-1}Y)^{-1} .$$

Therefore

$$\sum_{\nu \in (\mathcal{P}_1)_0/(\mathcal{P}_*)_0} \det((\mu\nu)^{*-1}Y)_2^{5-s}$$
$$= \sum_{\nu \in \mathfrak{I}_0^{(2)}/(\mathcal{P}_0)_0} \left((\det Y)(\mu^{*-1}Y)^{-1}_{\{2,3\}}, \nu e_3 \right)^{5-s}$$
$$= \varphi^{(2)} \left((\det Y)(\mu^{*-1}Y)^{-1}_{\{2,3\}}, s-5 \right) ,$$

where $(\mu^{*-1}Y)^{-1}_{\{2,3\}}$ is the right lower corner 2×2 submatrix of $(\mu^{*-1}Y)^{-1}$. But

$$\det(\mu^{*-1}Y)^{-1}_{\{2,3\}} = ((\mu^{*-1}Y)^{-1} \times (\mu^{*-1}Y)^{-1}, e_1)$$
$$= ((\det Y)^{-1}\mu^{*-1}Y, e_1) = (\det Y)^{-1}(\mu^{*-1}Y)_1 .$$

Since

$$Z^{(2)}(Y,s) = 240\,(2\pi)^{-s}\,\Gamma(s)\,\zeta(s)\zeta(s-3)\,\varphi^{(2)}(Y,s)$$
$$= (240)\,2^{-3}\,\pi^{-2}\,(s-1)(s-3)\,\rho(s)\rho(s-3)\,\varphi^{(2)}(Y,s) ,$$

we have

$$\chi(s) = \sum_{\mu \in \mathfrak{I}_0/(\mathcal{P}_1)_0} \Big(\sum_{t \in \mathbb{Z}-0} \frac{2^3\pi^2}{240} (\det Y)^{5/2} (\mu^{*-1}Y)_1^{-5/2}$$
$$\cdot e^{-2\pi|t|(\mu^{*-1}Y)_1} |t|^{-1}$$
$$\cdot \omega_1(4\pi|t|(\mu^{*-1}Y)_1; \frac{s}{2}-4, \frac{s}{2}-4) \tag{6.5}$$
$$\cdot \beta(s-8)$$
$$\cdot Z^{(2)}\Big((\det Y)^{1/2}(\mu^{*-1}Y)_1^{-1/2}(\mu^{*-1}Y)_{\{2,3\}}^{-1}, s-5\Big)\Big).$$

Note that $\det((\det Y)^{1/2}(\mu^{*-1}Y)_1^{-1/2}(\mu^{*-1}Y)_{\{2,3\}}^{-1}) = 1$. Therefore by the functional equation of ρ, β, ω_1 and $Z^{(2)}$, we have the functional equation: $\chi(18-s) = \chi(s)$. Also because of the inequality of ω_1 in the theorem in Section 3, $\chi(s)$ converges and defines a meromorphic function in s which has a pole of order 1 at $s = 5, 13$. Therefore by (6.1), (6.2), (6.3), (6.4) and (6.5),

$$\Psi(s) = (\det Y)^{s/2}\, \rho(s)\rho(s-4)\rho(s-8)\,(s-2)(s-4)\, E_{0,s}(Z)$$
$$= \Psi_0(s,Z) + \Psi_1(s,Z) + \Psi_2(s,Z) + \Psi_0(s,Z),$$

can be continued as a meromorphic function in s to a whole complex plane which has a pole of order 1 at $s = 0, 1, 5, 8, 10, 13, 17, 18$ and satisfies the functional equation:

$$\Psi(18-s) = \Psi(s).$$

This completes the proof of Theorem B.

REMARK. As in Shimura [14], considering residues $E_{k,s}(Z)$ at $s = 2$, we get the following result: (i) If $k = 4$, $E_{4,s}(Z)$ has a simple pole at $s = 2$ and the residue at $s = 2$ is

$$\frac{1}{\pi^2} \sum_{\mu\iota_{(2)} \in \mathfrak{I}_0 \iota_{(2)} N_{0\mathbb{Q}}/N_{0\mathbb{Q}}} \det(\mu^{*-1}Y)_2^{-1} \sum_{\substack{T \in \Lambda_2 \cap \mathfrak{K} \\ \det_{(2)} T = 0}} b(T)\, e^{2\pi i (T, (Z\cdot\mu)_2)},$$

where $b(T) \in \mathbb{Q}$.

(ii) If $k = 8$, $E_{8,s}(Z)$ has a simple pole at $s = 2$ and the residue is $\big((\det Y)^{-1}/\pi^3\big) \times$ (singular modular form of weight 8 with rational

Fourier coefficients). Here we note that for $\mu \in \mathfrak{I}$, $Z \cdot \mu = \mu^{*-1} Z$ and so $\mathrm{Im}(Z \cdot \mu) = \mu^{*-1} Y$.

References.

[1] Baily, W.L., Jr., An exceptional arithmetic group and its Eisenstein series. *Ann. of Math* **91** (1970), 512-549.

[2] Baily, W.L., Jr., *Introductory lectures on automorphic forms.* Princeton University Press, 1973.

[3] Borel, A., Introduction to automorphic forms. *Proc. Symp. Pure Math* **9** (1966), 199-210.

[4] Eie, M. and Krieg, A., The Maass space on the half-plane of Cayley Numbers of degree two. *Math. Z.* **210** (1992), 113-128.

[5] Freudenthal, H., Beziehungen der E_7 and E_8 zur Oktavenebene I. *Proc. Konkl. Ned. Akad. Wet.*, Series A, **57** (1954), 218-230.

[6] Karel, M., Fourier coefficients of certain Eisenstein series. *Ann. of Math.* **99** (1974), 176-202.

[7] Karel, M., Eisenstein series on tube domains. *Abh. Math. Sem. Univ. Hamburg* **62** (1992), 81-116.

[8] Karel, M., Values of certain Whittaker functions on a p-adic reductive group. *Illinois J. Math.* **26** (1982), 552-575.

[9] Kaufhold, G., Dirichletsche Reihe mit Funktionalgleichung in der Theorie der Modulfunktionen 2. Grades. *Math. Ann.* **137** (1959), 454-476.

[10] Krieg, A., *Modular forms on half-spaces of quaternions.* Lecture Notes in Math. **1143** (1980).

[11] Nagaoka, S., Confluent hypergeometric functions on an exceptional domain. *Proc. Japan Acad.*, Series A, **60** (1984), 210-220.

[12] Resnikoff, H. S. On a class of linear differential equations for automorphic forms in several complex variables. *Amer. J. Math.* **95** (1973), 321-332.

[13] Shimura, G., Confluent hypergeometric functions on tube domains. *Math. Ann.* **260** (1982), 269-302.

[14] Shimura, G., On Eisenstein series. *Duke Math. J.* **50** (1983), 417-476.

[15] Shimura, G., On Eisenstein series of half-integral weight. *Duke Math. J.* **52** (1985), 281-314.

[16] Tsao, L. C., The rationality of the Fourier coefficients of certain Eisenstein series in tube domains, I. *Compositio Math.* **32** (1976), 225-291.

Recibido: 31 de agosto de 1.992

Henry H. Kim*
Department of Mathematics
The University of Chicago
Chicago, IL 60637, USA

* This paper represents the author's doctoral thesis at the University of Chicago. The author wishes to express his gratitude to his thesis advisor, Prof. Walter Baily, Jr., for suggesting the topic of this research and for his encouragement and advice.

REVISTA MATEMÁTICA IBEROAMERICANA
VOL. 9, N.° 2, 1993

Complex tangential characterizations of Hardy-Sobolev Spaces of holomorphic functions

Sandrine Grellier

Introduction and results.

Let Ω be a C^∞-domain in $\mathbb{C}^n$. It is well known that a holomorphic function on Ω behaves twice as well in complex tangential directions (see [GS] and [Kr] for instance). It follows from well known results (see [H], [RS]) that some converse is true for any kind of regular functions when Ω satisfies

(P) The tangent space is generated by the Lie brackets of real and imaginary parts of complex tangent vectors.

In this paper, we are interested in the behavior of holomorphic Hardy-Sobolev functions in complex tangential directions. Our aim is to give a characterization of these spaces, defined on a domain which satisfies the property (P), involving only complex tangential derivatives. Our method, which is elementary, is to prove pointwise estimates between gradients and tangential gradients of holomorphic functions and, next, to use them to obtain the characterization of Hardy-Sobolev spaces for $1 \leq p < \infty$. To give precise statements, let us introduce some notations.

Write $\Omega = \{r < 0\}$, where r is a C^∞ function satisfying $dr \neq 0$ on $\partial\Omega = \{r = 0\}$.

Define the holomorphic complex normal vector field

$$N = \sum_{j=1}^{n} \frac{\partial r}{\partial \overline{z_j}} \frac{\partial}{\partial z_j}$$

and the (holomorphic) complex tangential gradient of order k of u, $\{\nabla_T^k u\}$, as follows. It is the vector $\left\{L_{I,J}\, u \,:\, I, J \in \{1, \ldots, n\}^k\right\}$ where

$$L_{i,j} = \frac{\partial r}{\partial z_i} \frac{\partial}{\partial z_j} - \frac{\partial r}{\partial z_j} \frac{\partial}{\partial z_i}, \qquad i, j \in \{1, \ldots, n\}$$

and $L_{I,J} = L_{i_1,j_1} \ldots L_{i_k,j_k}$ when $I = (i_1, \ldots, i_k), J = (j_1, \ldots, j_k)$.

The family $\{L_{i,j} \,:\, i, j \in \{1, \ldots, n\}, i \neq j\}$ gives a total system of complex tangential vector fields on $\partial\Omega$ (respectively on $\partial\Omega_\varepsilon = \{r = -\varepsilon\}$, $0 < \varepsilon < \varepsilon_0$).

For z in Ω near $\partial\Omega$, we set

$$C_2(z) = \max\left\{\left|\partial\overline{\partial} r(L_{i,j}, \overline{L_{k,l}})\right|(z) : i, j, k, l \in \{1, \ldots, n\}, i \neq j, k \neq l\right\}.$$

It is known that C_2 is different from zero on $\partial\Omega$ if and only if Ω satisfies (P).

Denote by $\delta(\cdot)$ the distance to the boundary $\partial\Omega$. We use the following mean-value operator for z in Ω near $\partial\Omega$

$$\mathrm{Mean}^{Q(z)}(u) = \frac{1}{|Q(z)|} \int_{Q(z)} |u(\zeta)|\, dV(\zeta)$$

where $Q(z)$ is a polydisc centered at z whose size is $c\,\delta(z)$ in the complex normal direction and $\sqrt{c\,\delta(z)}$ in the complex tangential ones, c chosen so that, in particular, $Q(z) \subset \Omega$.

Now, let us state our pointwise estimates.

Pointwise estimates. *Let $k \in \mathbb{N}$, $0 < p < \infty$. For each $z_0 \in \partial\Omega$, there exist a neighborhood $V(z_0)$ and a constant C such that, for every holomorphic function g in Ω and every $z \in V(z_0) \cap \Omega$*

$$\delta(z)^{kp/2} \left|\nabla_T^k g(z)\right|^p \leq C\, \mathrm{Mean}^{Q(z)}(|g|^p), \tag{1}$$

$$C_2(z)^{kp}\delta(z)^{kp}\left|\nabla^k g(z)\right|^p \leq C\,\mathrm{Mean}^{Q(z)}\left(\delta^{kp/2}\left|\nabla_T^k g\right|^p + \mathrm{Rest}^k(g)^p\right) \tag{2}$$

where

$$\mathrm{Rest}^k(g) = \delta^{1/2}\Big(\sum_{r=0}^{k-1}\sum_{\substack{1\leq j+r<(k+r)/2\\ j\geq 0}} \mathcal{O}(\delta^{(k-1)/2})\left|\nabla^j\nabla_T^r g\right| + \sum_{r=0}^{k-1}\sum_{\substack{(k+r)/2\leq j+r\leq k\\ j\geq 0}} \mathcal{O}(\delta^{j+r/2})\left|\nabla^j\nabla_T^r g\right|\Big).$$

Inequality (1) is what we call the direct estimates. Such an estimate is implicit in some works but is not explicitely written (see [GS] and [Kr]).

Inequality (2) is the new estimate we prove; it says that, up to a rest, the complex tangential gradient controls all the gradient. It is what we call the converse estimates.

The main terms in these estimates are homogeneous in the following sense: each derivative of order r in the complex tangential directions appears with a factor $\delta^{r/2}$ and each one in the other directions with a factor δ^r. In the remaining terms, they appear with a smaller factor. Compared with the usual mean-value property of holomorphic functions, these pointwise estimates show that $\nabla_T^k g$ behaves as a complete gradient of order $k/2$. Obviously, by the mean-value property and inequality (1), we can majorize $\mathrm{Rest}^k(g)$ by the mean-value of $\delta^{p/2}|g|^p$. However, for technical reasons, we will need this complicated form of $\mathrm{Rest}^k(g)$ (in order to be able to apply Hardy inequalities for example).

Now, we give the precise definition of the Hardy-Sobolev spaces. We will identify a small neighborhood of $\partial\Omega$ in $\overline{\Omega}$, with $\partial\Omega \times [0, s_0[$. More precisely, we choose a map $\Phi : \partial\Omega \times [0, s_0[\to \overline{\Omega}$ such that

- Φ is a diffeomorphism of $\partial\Omega \times [0, s_0[$ onto a neighborhood $\overline{\Omega} \cap U$ of $\partial\Omega$ in $\overline{\Omega}$,
- $\Phi(\zeta, 0) = \zeta$ for every $\zeta \in \partial\Omega$,
- $\delta(\Phi(\zeta, t)) \simeq t$ for every $\zeta \in \partial\Omega$ and every $0 < t < s_0$.

For $0 < p < \infty$ and $k \in \mathbb{N}$, the holomorphic Hardy-Sobolev space $\mathcal{H}_k^p(\Omega)$ is defined to be the space of holomorphic functions g which satisfy

$$\sup_{0<t<s_0}\left|\nabla^k g \circ \Phi(\cdot, t)\right| \in L^p(\partial\Omega),$$

We will see that this definition does not depend on the choice of the function Φ.

To state our main theorem, we need some other definitions.

We know, by [NSW] for instance, that we can define a non-isotropic metric $d(\cdot,\cdot)$ on $\partial\Omega$ which satisfies $d(p,q) \simeq |p-q|^2 + |\langle \mathrm{Im}\, N_p, p-q\rangle|$ (see [S1]).

We define the following quantities for every smooth function u and every aperture $\alpha > 0$.

- The maximal admissible function:

$$\text{for every } \zeta \in \partial\Omega, \quad \mathcal{M}_\alpha u(\zeta) = \sup\{|u(z)| : \quad z \in \mathcal{A}_\alpha(\zeta)\}$$

where $\mathcal{A}_\alpha(\zeta)$ denotes the admissible approach region

$$\mathcal{A}_\alpha(\zeta) = \{\Phi(\eta,t) : \quad \eta \in \partial\Omega,\ 0 < t < s_0, \quad d(\eta,\zeta) < \alpha\, t\}.$$

- The admissible area function:

$$\text{for every } \zeta \in \partial\Omega, \quad S^q_\alpha u(\zeta) = \left(\int_{\mathcal{A}_\alpha(\zeta)} |u|^2\, \delta^q\, \frac{dV}{\delta^{n+1}}\right)^{1/2}.$$

- The Littlewood-Paley function:

$$\text{for every } \zeta \in \partial\Omega, \quad G^q(u)(\zeta) = \left(\int_0^{s_0} |u \circ \Phi(\zeta,t)|^2\, t^q\, \frac{dt}{t}\right)^{1/2}.$$

- The non-isotropic maximal operator on $\partial\Omega$:

$$\text{for every } \zeta \in \partial\Omega, \quad Mf(\zeta) = \sup_{t>0} \frac{1}{|B^d(\zeta,t)|} \int_{B^d(\zeta,t)} |f|\, d\sigma$$

where $B^d(\zeta,t)$ is a non-isotropic ball on $\partial\Omega$, defined with the aid of the metric d, centered at ζ, of radius t.

Since d is a metric and defines a space of homogeneous type, the non-isotropic maximal operator on $\partial\Omega$ is bounded from $L^p(\partial\Omega)$ into itself, for every $1 < p \leq \infty$ and is of weak type $(1,1)$ (see [S1]).

Before stating our results in terms of complex tangential derivatives, we recall some known results about Hardy-Sobolev spaces (where S_α and G stand respectively for S^0_α and G^0).

Auxiliary Theorem. *Let α be a fixed aperture, $k \in \mathbb{N}$. For every $0 < p < \infty$ and every holomorphic function g, the following are equivalent:*

(1) $g \in \mathcal{H}^p_k(\Omega)$
(2) $S_\alpha(\delta \nabla^{k+1} g) \in L^p(\partial\Omega)$
(3) $G(\delta \nabla^{k+1} g) \in L^p(\partial\Omega)$
(4) $\mathcal{M}_\alpha(\nabla^k g) \in L^p(\partial\Omega)$

and the corresponding norms are equivalent.

Now, we can state our main result describing $\mathcal{H}^p_k(\Omega)$ only in terms of complex tangential derivatives.

Main Theorem. *Let α be a fixed aperture, $k \in \mathbb{N}$. For every $1 \leq p < \infty$ and every holomorphic function g, the following are equivalent:*

(1) $g \in \mathcal{H}^p_k(\Omega)$
(2) $S_\alpha\left(\delta \nabla \nabla_T^{2k} g\right) \in L^p(\partial\Omega)$
(3) $\mathcal{M}_\alpha\left(\nabla_T^{2k} g\right) \in L^p(\partial\Omega)$
(4) $\sup_{0<t<s_0} \left|\nabla_T^{2k} g\right| \in L^p(\partial\Omega)$

and the corresponding norms are equivalent.

REMARK. The results of Main Theorem are true for a larger class of p. We will give later the details and precise statements. For instance, it follows from our results the following corollary.

Corollary. *Let α be a fixed aperture. For every $0 < p < \infty$ and every holomorphic function g, we have*

$$g \in \mathcal{H}^p(\Omega) \quad \text{if and only if} \quad \|S_\alpha(\delta \nabla_T^2 g)\|_{L^p(\partial\Omega)} < \infty\,.$$

For the unit ball in $\mathbb{C}^n$, a characterization of Hardy-Sobolev spaces in terms of complex tangential derivatives is given by Ahern and Bruna in [AB]. But, in this particular case, it is easier since the complex tangential derivatives of holomorphic functions are also harmonic. In the case of strictly pseudoconvex domains, Cohn gives a characterization of Hardy-Sobolev spaces $\mathcal{H}^p_k$ with $p > 1$ in terms of maximal function of complex tangential gradients of order $2k$. But his proof uses the representation of the Szegö kernel given by Kerzman and Stein and, so, needs pseudoconvexity (see [Co]).

Our method is to use the pointwise estimates essentially to show that one can define the Hardy-Sobolev spaces in terms of the admissible area function of ordinary gradients as well as in terms of the admissible area function of complex tangential gradients. Then, for the other characterizations, we adapt, when it is possible, the method of [FS]. The technical difficulties are due to the fact that, for a holomorphic function g, $\nabla_T^k g$ is no longer holomorphic nor harmonic, but we can show that, locally and up to a rest, it satisfies some mean-value properties analogous to the ones satisfied by holomorphic functions. When the technics of [FS] do not work, as far as we know, we use a trick which consists in writing $\nabla_T^k g$ as the sum of the solution of a Dirichlet's Problem with data $\Delta \nabla_T^k g$ and a harmonic function -the idea being that the harmonic part is the principal term and the other part a rest. To estimate this rest, we prove an estimate on the Dirichlet's problem in mixed L^p norms with weight.

In a previous paper (see [G1]), we gave analogous pointwise estimates in the more general context of domains of finite type and we applied them to characterize Lipschitz, Besov and Sobolev spaces of holomorphic functions. These estimates allow to generalize some of the results of Main Theorem to domains of finite type. But, as we are not able to deal completely with this case, we restrict ourselves to the case of the (P) property. For more details on finite type domains, see [G2].

1. Pointwise estimates.

1.1 Preliminaries. Change of coordinates and polydiscs.

Let $z_0 \in \partial\Omega$. As $dr(z_0) \neq 0$, we can assume that $\partial r/\partial z_1 \neq 0$ on a neighborhood $V(z_0)$ of z_0. We shall need the following lemma which is well known (see [C] for instance).

Lemma 1.1. *For each $z \in V(z_0) \cap \Omega$, there exists a polynomial biholomorphism Φ_z from $\mathbb{C}^n$ to itself such that*

1) *The coefficients of Φ_z are C^∞ with respect to z and the jacobian of Φ_z is uniformly bounded from both side for $z \in V(z_0)$.*

2) *The defining function $\varrho = r \circ \Phi_z$ of $\Omega_z = \Phi_z^{-1}(\Omega)$ satisfies*

$$\varrho(\zeta) = r(z) + \mathrm{Re}\,\zeta_1 + \sum_{j,k=2}^{n} a_{j,k}(z)\,\zeta_j\overline{\zeta_k} + \mathcal{O}\left(|\zeta_1|\,|\zeta| + |\zeta'|^3\right)$$

where $\zeta' = (\zeta_2, \dots, \zeta_n)$.

3) *There exists a constant* c *on* $V(z_0)$ *such that the polydisc* $R(z)$ *defined by*

$$R(z) = \left\{ \zeta \in \mathbb{C}^n : \quad |\zeta_1| < c\,\delta(z), \quad \sum_{k=2}^{n} |\zeta_k|^2 < c\,\delta(z) \right\}$$

is included in Ω_z. *So*

$$Q(z) = \Phi_z(R(z)) \subset \Omega,$$

and there exist $C_1, C_2 > 0$ *such that*

$$P_{C_1}(z) \subset Q(z) \subset P_{C_2}(z)$$

where

$$P_C(z) = \left\{ \zeta \in \mathbb{C}^n : \quad |(z-\zeta)N_z| \le C\,\delta(z), \quad |z-\zeta|^2 \le C\,\delta(z) \right\}.$$

4) $C_2(z) \simeq \max\{|a_{j,k}(z)| : j,k \in \{2,\dots,n\}\}$ *for every* $z \in V(z_0)$.

REMARKS. This lemma allows to estimate ϱ and its derivatives; it shows that, since ϱ is C^∞,

$$\frac{\partial \varrho}{\partial \zeta_j}(0) = 0 \quad \text{for } j = 2,\dots,n, \quad \frac{\partial \varrho}{\partial \zeta_1}(0) = \frac{1}{2},$$
$$\frac{\partial^2 \varrho}{\partial \zeta_j \partial \zeta_k}(0) = 0, \quad \frac{\partial^2 \varrho}{\partial \zeta_j \partial \overline{\zeta_k}}(0) = a_{j,k}(z) \quad \text{for } j,k = 2,\dots,n.$$

Let us denote by $Q_t(z)$ the set

$$\Phi_z\left(\left\{\zeta \in \mathbb{C}^n : \quad |\zeta_1| < t, \sum_{k=2}^{n} |\zeta_k|^2 < t\right\}\right).$$

It is shown in [NSW] that, for every $\eta \in \partial\Omega$, $Q_t(\eta) \cap \Omega$ is comparable with the "tent"

$$\hat{B}(\eta, t) = \left\{ z \in \overline{\Omega} \cap U : \quad d(\pi(z), \eta) \le t,\ \delta(z) \le t \right\},$$

where π denotes the projection on $\partial\Omega$. This allows to see that there exist $c_1, c_2, c_3 > 0$ such that

$$\Phi^{-1}\left(Q(\Phi(\eta,t))\right) \subset B^d(\eta, c_1 t)\times]c_2 t, c_3 t[\,.$$

(It suffices to remark the two following properties:

- $\eta \in Q(\Phi(\eta, \tilde{C}t))$ for some constant $\tilde{C}$; so, $Q(\Phi(\eta,t)) \subset Q_{c_1 t}(\eta)$ for some constant c_1,
- $\delta(\cdot) \simeq t$ on $Q(\Phi(\eta,t))$.)

PROOF. There exists a biholomorphism

$$\Phi_z(\zeta) = \Big(z_1 + d_0(z)\,\zeta_1 + \sum_{j=2}^{n} d_1^j(z)\,\zeta_j + \sum_{j,k=2}^{n} d_2^{j,k}(z)\,\zeta_j\zeta_k,\ z_2+\zeta_2, \ldots,\ z_n+\zeta_n\Big)$$

such that $\varrho(\zeta) = r \circ \Phi_z(\zeta)$ takes the given form; explicitely

$$d_0(z) = \frac{1}{2}\left(\frac{\partial r}{\partial z_1}(z)\right)^{-1}, \quad d_1^j(z) = \left(\frac{\partial r}{\partial z_1}(z)\right)^{-1}\frac{\partial r}{\partial z_j}(z),$$

$$d_2^{j,k}(z) = -d_0(z)\Big[\frac{\partial^2 r}{\partial z_1^2}(z)\,d_1^j(z)\,d_1^k(z) + \frac{\partial^2 r}{\partial z_1 \partial z_k}(z)\,d_1^j(z) + \frac{\partial^2 r}{\partial z_1 \partial z_j}(z)\,d_1^k(z) + \frac{\partial^2 r}{\partial z_j \partial z_k}(z)\Big].$$

Properties 3) and 4) follow from a direct computation.

Our aim, now, is to show that, after this change of coordinates, $\nabla_T^k g$, for g holomorphic, is, locally and up to a rest, a function which satisfies some mean-value properties.

Let $z \in V(z_0) \cap \Omega$. Let us consider the family

$$L_i' = \frac{\partial \varrho}{\partial \zeta_1}\frac{\partial}{\partial \zeta_i} - \frac{\partial \varrho}{\partial \zeta_i}\frac{\partial}{\partial \zeta_1}, \quad i \in \{2, \ldots, n\},$$

of complex tangential vector fields in Ω_z. Since by assumption $\partial\varrho/\partial\zeta_1 \neq 0$ on a neighborhood of $0 \in \Omega_z$, the family L_i' for $i \in \{2, \ldots, n\}$ gives

a total system of complex tangential vector fields in a neighborhood of $0 \in \Omega_z$. We need the following technical lemma which allows to write locally the field $L'^K = L_2'^{k_2} \dots L_n'^{k_n}$ as a sum of a field with coefficients which are almost harmonic and a rest. As before, we write $\zeta = (\zeta_1, \zeta')$ where $\zeta' = (\zeta_2, \dots, \zeta_n)$.

Lemma 1.2. *Let $K = (k_2, \dots, k_n) \in \mathbb{N}^{n-1}$. On the set*

$$R(z) = \left\{ \zeta \in \mathbb{C}^n : \quad |\zeta_1| \leq c\,\delta(z), \sum_{k=2}^{n} |\zeta_k|^2 \leq c\,\delta(z) \right\} \subset \Omega_z ,$$

we have

$$\begin{aligned} L'^K &= \prod_{j=2}^{n} \left[\frac{1}{2} \frac{\partial}{\partial \zeta_j} - \left(\sum_{l=2}^{n} a_{j,l}(z) \overline{\zeta_l} \right) \frac{\partial}{\partial \zeta_1} \right]^{k_j} + \mathrm{Rest}_{|K|}^K \\ &= F^K + \mathrm{Rest}_{|K|}^K \end{aligned}$$

where Rest_k^K *stands for*

$$\mathrm{Rest}_k^K = \sum_{\substack{1 \leq j+|R| \leq k \\ R \leq K}} b_{j,R} \frac{\partial^{j+|R|}}{\partial \zeta_1^j \partial \zeta^R}$$

with $C^\infty(\Omega_z)$-functions $b_{j,R}$, $1 \leq j+|R| \leq k$, which satisfy the following properties: they are uniformly bounded and if $(2j+|R|-k+1)/2 > 0$, $b_{j,R}(0) = 0$ and, for every $\zeta \in R(z)$, $|b_{j,R}(\zeta)| \leq C\,\delta(z)^{(2j+|R|-k+1)/2}$ for some uniform constant C.

PROOF. We will give the proof in $\mathbb{C}^2$ to simplify.

By convention, we will denote by $\mathcal{O}_r$ any regular function defined on Ω_z, uniformly bounded, satisfying, if $r > 0$, $\mathcal{O}_r(0) = 0$ and, for every $\zeta \in R(z)$, $|\mathcal{O}_r(\zeta)| \leq C\delta(z)^r$ for some constant C. Let

$$L' = \frac{\partial \varrho}{\partial \zeta_1} \frac{\partial}{\partial \zeta_2} - \frac{\partial \varrho}{\partial \zeta_2} \frac{\partial}{\partial \zeta_1}$$

be a complex tangential vector field in Ω_z. We can show by induction on k that there exist some constants $c_{j,r}$, $1 \leq j+r \leq k$, such that

$$L'^k = \sum_{1 \leq j+r \leq k} \sum_{E_{k,j,r}} c_{j,r} \left(\prod_{i=1}^{k} \frac{\partial^{m_i+n_i} \varrho}{\partial \zeta_1^{m_i} \partial \zeta_2^{n_i}} \right) \frac{\partial^{j+r}}{\partial \zeta_1^j \partial \zeta_2^r} ,$$

where $E_{k,j,r}$ denotes the set of couples (m_i, n_i), $i = 1, \dots, k$, which are in alphabetical order and satisfy $\sum_{i=1}^{k} m_i = k - j$, $\sum_{i=1}^{k} n_i = k - r$ with $m_i + n_i \geq 1$.

When $j + r = k$, necessarily, there are j couples which are equal to $(0,1)$ and r couples which are equal to $(1,0)$. So, the corresponding terms get the following form

$$C \left(\frac{\partial \varrho}{\partial \zeta_1}\right)^r \left(\frac{\partial \varrho}{\partial \zeta_2}\right)^{k-r} \frac{\partial^k}{\partial \zeta_1^{k-r} \partial \zeta_2^r} \cdot$$

But, we know by the Taylor expansion of ϱ given in Lemma 1.1 that, for $\zeta \in R(z)$,

$$\frac{\partial \varrho}{\partial \zeta_1}(\zeta) = \frac{1}{2} + \mathcal{O}_{1/2}(\zeta)$$

and

$$\frac{\partial \varrho}{\partial \zeta_2}(\zeta) = a_{2,2}(z)\overline{\zeta_2} + \mathcal{O}_1(\zeta) = \mathcal{O}_{1/2}(\zeta).$$

This allows to see that the terms of order k take the form given in the lemma. Let us look at the terms with $j + r < k$. Since ϱ is a C^∞ function, the coefficients are uniformly bounded on $V(z_0)$. So, it suffices to consider the case when $2j + r \geq k$ and to show that, in this case, the corresponding coefficients are equal to zero at the origin and are bounded by $C\,\delta(z)^{(2j+r-k+1)/2}$ on $R(z)$.

So, let j, r fixed with $j + r < k$ and $2j + r \geq k$. Let us denote by J the number of couples (m_i, n_i) with m_i equal to zero. As $\sum m_i = k - j$, necessarily $J \geq j$. Assume that $m_1 = \cdots = m_J = 0$, necessarily $n_i \geq 1$ for $i \leq J$. Let us denote by K the number of couples $(0, n_i)$, $i \leq J$ with $n_i = 1$. We have $n_1 = \cdots = n_K = 1$ and

$$k - r = \sum_{i=1}^{k} n_i = K + \sum_{j=K+1}^{J} n_i + \sum_{j=J+1}^{k} n_i \geq K + 2(J - K).$$

So, $K \geq 2J - k + r \geq 2j - k + r$. So, if $K \geq 2j - k + r + 1$, the corresponding coefficient which is known to be a $\mathcal{O}_{K/2}$ (as there are at least K factors $\partial\varrho/\partial\zeta_2$), is bounded by $C\delta^{(2j-k+r+1)/2}$ and is equal to zero at the origin since, by assumption on the indices, $K \geq 1$. Otherwise, if $K < 2j - k + r + 1$ then, necessarily $J = j$, $K = 2j - k + r$ and there exists at least one couple $(0, n_i)$ with $n_i = 2$ for $K + 1 \leq i \leq j$ (since

otherwise all the n_i, for $K+1 \leq i \leq j$, should be strictly bigger than 2 and we would have

$$\begin{aligned} k-r &= \sum_{i=1}^{k} n_i \\ &> K + 2(J-K) \\ &= 2j - k + r + 2(k-j-r) = k-r\,, \end{aligned}$$

which is impossible).

So, we use the fact that $\partial^2\varrho/\partial\zeta_2^2\,(\zeta) = \mathcal{O}_{1/2}(\zeta)$ for every $\zeta \in R(z)$. This gives that the corresponding coefficient is a $\mathcal{O}_{(K+1)/2}$ and so, is equal to zero at the origin and is bounded by $C\delta^{(2j-k+r+1)/2}$ since there are $K = 2j-k+r$ factors $\partial\varrho/\partial\zeta_2$ and at least one $\partial^2\varrho/\partial\zeta_2^2$.

In the new system of coordinates, near the origin, $\partial/\partial\zeta_i \simeq L_i'$, this allows to show the following corollary.

Corollary 1.3. *For every $l \in \mathbb{N}$, every $K \in \mathbb{N}^{n-1}$ and every function $u \in C^\infty(\overline{\Omega_z})$, we have, on $R(z)$*

$$\left|\frac{\partial^{l+|K|}u}{\partial\zeta_1^l\partial\zeta'^K}\right| \leq C \sum_{\substack{1\leq j+|R|\leq l+|K| \\ R\leq K}} \left|L'^R\frac{\partial^j u}{\partial\zeta_1^j}\right|.$$

PROOF. Lemma 1.2 allows to write on $R(z)$

$$\begin{aligned} \left(\frac{\partial\varrho}{\partial\zeta_1}\right)^{|K|}\frac{\partial^{|K|}}{\partial\zeta'^K} &= L'^K + \sum_{|J|=0}^{|K|-1}\mathcal{O}(\delta(z)^{(|K|-|J|)/2})\frac{\partial^{|K|}}{\partial\zeta_1^{|K|-|J|}\partial\zeta'^J} \\ &\quad + \mathrm{Rest}_{|K|-1}^K f\,, \end{aligned}$$

and we can assume that $\partial\varrho/\partial\zeta_1 \neq 0$ on $R(z)$. So, this allows us to estimate each derivative $|\partial^{l+|K|}u/\partial\zeta_1^l\partial\zeta'^K|$ in terms of $|L'^K\partial^l u/\partial\zeta_1^l|$ and of derivatives either of order strictly less than $l+|K|$, or of order with respect to ζ' strictly less than $|K|$.

Applying successively this estimate, we obtain the lemma.

Now, we are going to see that, for f holomorphic in Ω_z, $F^K f$ satisfies some mean-value property. Let us give the following definition (*cf.* [AB] and [G1] for instance).

Definition. *Let Ω be an open set in $\mathbb{C}^n$. Let $K = (k_1, \dots, k_n)$ be a multi-index of integers. A function $F \in C^\infty(\Omega)$ is called* $(\mathrm{AB})_K$ *if*

$$\frac{\partial^{k_j} F}{\partial \overline{\zeta_j}^{k_j}} = 0\,, \qquad \text{for } j = 1, \dots, n\,.$$

To simplify, we will assume that K is fixed in the following and we will write (AB) instead of $(\mathrm{AB})_K$. For every $\zeta \in \mathbb{C}$, $r > 0$, we denote by $D(\zeta, r)$ the disc $\{z \in \mathbb{C} : |z - \zeta| \leq r\}$. Then, we have the following lemma (see [AB]).

Lemma 1.4. *For every $L, M \in \mathbb{N}^n$, $0 < p < \infty$, there exists a constant C such that, for every* (AB) *function F in Ω, every $\zeta = (\zeta_1, \dots, \zeta_n) \in \Omega$ and every $r = (r_1, \dots, r_n) \in (]0, +\infty[)^n$ with $D(\zeta_1, r_1) \times \cdots \times D(\zeta_n, r_n) \subset \Omega$, we have*

$$\left| \frac{\partial^{|L|+|M|} F}{\partial \overline{\zeta}^L \partial \zeta^M}(\zeta) \right|^p \leq \frac{C}{\prod_{j=1}^n r_j^{p(L_j + M_j)+2}} \int_{D(\zeta_1, r_1) \times \cdots \times D(\zeta_n, r_n)} |F|^p dV.$$

So, for f holomorphic, $F^K f$ is an (AB) function. This allows to prove the following result for example which gives a mean-value property with rest for $\nabla_T^k \nabla^l g$.

Corollary 1.5. *For every $l, k \in \mathbb{N}$ and every holomorphic function g in Ω, we have, for every $z \in V(z_0) \cap \Omega$,*

$$\delta(z)^{k/2+l} \left| \nabla_T^k \nabla^l g \right|(z) \leq C \,\mathrm{Mean}^{Q(z)} \left(\delta^{k/2+l} \left| \nabla_T^k \nabla^l g \right| + \delta^l \,\mathrm{Rest}^k(\nabla^l g) \right).$$

(Rest^k *has the same meaning as in the Pointwise Estimates*).

PROOF. It suffices to consider the case $l = 0$. The general case follows replacing g by $\nabla^l g$.

We set $f = g \circ \Phi_z$. Since, by assumption, the family

$$L_i' = \frac{\partial \varrho}{\partial \zeta_1} \frac{\partial}{\partial \zeta_i} - \frac{\partial \varrho}{\partial \zeta_i} \frac{\partial}{\partial \zeta_1}\,, \qquad i \in \{2, \dots, n\}$$

gives a total system of complex tangential vector fields in a neighborhood of $0 \in \Omega_z$, each iterated complex tangential vector field of order

k at 0 is obtained as a linear combination with smooth coefficients of L'^K at 0, where $K \in \mathbb{N}^{n-1}$, $1 \leq |K| \leq k$.

So, to estimate $\nabla_T^k g(z)$, we have to estimate $L'^K f(0)$ for every $K \in \mathbb{N}^{n-1}$ with $1 \leq |K| \leq k$.

To simplify, we will only estimate $L'^k_i f(0)$ and we will write $L'^k_i f = F^k_i + \mathrm{Rest}^k_i$.

By Lemma 1.2, we have

$$\begin{aligned}\left|L'^k_i f(0)\right| &\leq C\left(\left|\frac{\partial^k f}{\partial \zeta_i^k}(0)\right| + \sum_{1\leq 2j+r\leq k-1}\left|\frac{\partial^{j+r} f}{\partial \zeta_1^j \partial \zeta_i^r}(0)\right|\right)\\ &= C\left(\left|F^k_i f(0)\right| + \sum_{1\leq 2j+r\leq k-1}\left|\frac{\partial^{j+r} f}{\partial \zeta_1^j \partial \zeta_i^r}(0)\right|\right)\end{aligned}$$

Now $F^k_i f$ is an (AB) function since f is holomorphic and so, by Lemma 1.4,

$$\begin{aligned}\left|L'^k_i f(0)\right| &\leq C\,\mathrm{Mean}^{R(z)}\left(\left|F^k_i f\right| + \sum_{1\leq 2j+r\leq k-1}\left|\frac{\partial^{j+r} f}{\partial \zeta_1^j \partial \zeta_i^r}\right|\right)\\ &\leq C\,\mathrm{Mean}^{R(z)}\left(\left|L'^k_i f\right| + \mathrm{Rest}^k_i f\right).\end{aligned}$$

Now, we will show that, after a change of coordinates, $\delta(z)^{k/2}\sum_i \mathrm{Rest}^k_i f$ is bounded by $\mathrm{Rest}^k(g)$. By Corollary 1.3, we have on $R(z)$

$$\begin{aligned}\mathrm{Rest}^k_i f \leq C\Big(&\sum_{r=0}^{k-1}\sum_{1\leq j+r<(k+r)/2}\left|L'^r_i \frac{\partial^j f}{\partial \zeta_1^j}\right|\\ &+\sum_{r=0}^{k-1}\sum_{(k+r)/2\leq j+r\leq k}\delta(z)^{(2j+r-k+1)/2}\left|L'^r_i \frac{\partial^j f}{\partial \zeta_1^j}\right|\Big).\end{aligned}$$

This inequality allows to conclude.

1.2. Direct estimates.

We are going to prove the following theorem.

Theorem A. *Let Ω be a C^∞ domain in $\mathbb{C}^n$. Let $k \in \mathbb{N}$, $l \in \mathbb{N}^\star$, $0 < p < \infty$. For every $z_0 \in \partial\Omega$, there exist a neighborhood $V(z_0)$ of z_0 and a constant C such that, for every holomorphic functions g in Ω and every $z \in V(z_0) \cap \Omega$, we have*

$$(1) \quad \delta(z)^{kp/2} \left|\nabla_T^k g(z)\right|^p \leq C \,\mathrm{Mean}^{Q(z)}(|g|^p).$$

$$(2) \quad \delta(z)^{kp/2+lp} \left|\nabla^l \nabla_T^k g(z)\right|^p \leq C \,\mathrm{Mean}^{Q(z)} \left(\delta^{lp} \sum_{j=1}^{l} \left|\nabla^j g\right|^p\right).$$

PROOF. Let $z \in V(z_0)$. We set $f(\zeta) = g(\Phi_z(\zeta))$, then f is holomorphic in $\Omega_z = \Phi_z^{-1}(\Omega)$.

In order to show the first part of Theorem A, we will apply Cauchy Formula to f. Since f is holomorphic, by subharmonicity property of f^p, we have, for every $j \in \mathbb{N}$, every $R \in \mathbb{N}^{n-1}$ and every $p > 0$

$$\left|\frac{\partial^{j+|R|} f}{\partial\zeta_1^j \partial\zeta'^R}(0)\right|^p \leq \frac{1}{(c\,\delta(z))^{n+1+p(j+|R|/2)}} \int_{|\zeta_1|,\,|\zeta'|^2 \leq c\delta(z)} |f(\zeta)|^p \, dV(\zeta).$$

The domain of integration is $R(z)$. In order to conclude, we recall that each iterated complex tangential vector field of order k at 0 is obtained as a linear combination with smooth coefficients of L'^K at 0, where $K \in \mathbb{N}^{n-1}$, $1 \leq |K| \leq k$. Furthermore, by Lemma 1.2, $L'^K f(0)$ is almost equal to $(1/2)^{|K|}\, \partial^{|K|} f / \partial\zeta'^K(0)$. More precisely, for each $K \in \mathbb{N}^{n-1}$, we can write, with the help of Lemma 1.2, that

$$\begin{aligned}
\left|L'^K f(0)\right| &\leq C \left(\left|\frac{\partial^{|K|} f}{\partial\zeta'^K}(0)\right| + \sum_{1 \leq 2j+|R| \leq |K|-1} \left|\frac{\partial^{j+|R|} f}{\partial\zeta_1^j \partial\zeta'^R}(0)\right|\right) \\
&\leq C \left(\delta(z)^{-|K|p/2} + \sum_{1 \leq 2j+|R| \leq |K|-1} \delta(z)^{-p(j+|R|/2)}\right) \\
&\quad + \left(\frac{1}{|R(z)|} \int_{R(z)} |f|^p dV\right) \\
&\leq C\,\delta(z)^{-|K|p/2} \left(\frac{1}{|R(z)|} \int_{R(z)} |f|^p dV\right).
\end{aligned}$$

This allows to conclude for the first part of Theorem A.

To show the second part, it suffices to apply the preceding result to the derivatives of g and then, to use the first part of the following elementary lemma.

Lemma 1.6. *For every $l, k \in \mathbb{N}$ and every $u \in C^\infty(\Omega)$, we have*

$$|\nabla^l \nabla_T^k u| \leq |\nabla_T^k \nabla^l u| + \mathcal{O}\Big(\sum_{\substack{1 \leq j \leq l, \\ 0 \leq r \leq k-1}} |\nabla_T^r \nabla^j u| \Big)$$

$$|\nabla_T^k \nabla^l u| \leq |\nabla^l \nabla_T^k u| + \mathcal{O}\Big(\sum_{\substack{1 \leq j \leq l, \\ 0 \leq r \leq k-1}} |\nabla_T^r \nabla^j u| \Big)$$

where the $\mathcal{O}$ are uniform on $V(z_0)$.

1.3. Converse estimates.

We are going to show the following theorem.

Theorem B. *Let Ω be a C^∞ domain in $\mathbb{C}^n$. Let $k \in \mathbb{N}$, $0 < p < \infty$. For every $z_0 \in \partial\Omega$, there exist a neighborhood $V(z_0)$ of z_0 and a constant C such that, for every holomorphic function g in Ω and every $z \in V(z_0) \cap \Omega$, we have*

$$C_2(z)^{kp} \delta(z)^{kp} \left|\nabla^k g(z)\right|^p \leq C \operatorname{Mean}^{Q(z)} \left(\delta^{kp/2} \left|\nabla_T^k g\right|^p + \operatorname{Rest}^k(g)^p \right),$$

where

$$\operatorname{Rest}^k(g) = \delta^{1/2} \Big(\sum_{r=0}^{k-1} \sum_{\substack{1 \leq j+r < (k+r)/2 \\ j \geq 0}} \mathcal{O}(\delta^{(k-1)/2}) \left|\nabla^j \nabla_T^r g\right| + \sum_{r=0}^{k-1} \sum_{\substack{(k+r)/2 \leq j+r \leq k \\ j \geq 0}} \mathcal{O}(\delta^{j+r/2}) \left|\nabla^j \nabla_T^r g\right| \Big).$$

PROOF. As before, for every $z \in V(z_0)$ fixed, we set $f(\zeta) = g(\Phi_z(\zeta))$. We begin with the following lemma.

Lemma 1.7. *Let Ω be a C^∞ domain in $\mathbb{C}^n$, let $k \in \mathbb{N}$ and $0 < p < \infty$. For each $z_0 \in \partial\Omega$, there exist a neighborhood $V(z_0)$ and a constant C*

such that, for every $z \in V(z_0) \cap \Omega$, there exists a transverse vector field M_z, with $M_z r(z) = 1$, such that, for every holomorphic function g in Ω, we have

$$C_2(z)^{kp}\delta(z)^{kp}\left|M_z^k g(z)\right|^p \leq C\,\mathrm{Mean}^{Q(z)}\left(\delta^{kp/2}|\nabla_T^k g|^p + \mathrm{Rest}^k(g)^p\right)$$

where $\mathrm{Rest}^k(g)$ *take the form given in Theorem* B.

PROOF. We write $L'^k_i f(\zeta) = F_i^k f(\zeta) + \mathrm{Rest}_i^k f(\zeta)$ for $\zeta \in R(z)$, where

$$F_i^k f(\zeta) = \left(\frac{1}{2}\frac{\partial}{\partial\zeta_i} - \sum_{l=2}^n a_{i,l}(z)\overline{\zeta_l}\frac{\partial}{\partial\zeta_1}\right)^k f(\zeta)\,.$$

Since f is holomorphic, $F_i^k f$ is an (AB) function. Furthermore

$$\frac{\partial^k F_i^k f}{\partial\overline{\zeta_l}^k}(0) = (-a_{i,l}(z))^k\,\frac{\partial^k f}{\partial\zeta_1^k}(0)\,, \qquad l = 2,\dots,n\,.$$

So, Lemma 1.4 allows us to deduce that, for $l = 2,\dots,n$,

$$\begin{aligned}\delta(z)^{kp/2}\,|a_{i,l}(z)|^{kp}\left|\frac{\partial^k f}{\partial\zeta_1^k}(0)\right|^p &\leq C\,\mathrm{Mean}^{R(z)}(|F_i^k f|^p)\\ &\leq C\,\mathrm{Mean}^{R(z)}\left(|L'^k_i f|^p + (\mathrm{Rest}_i^k f)^p\right).\end{aligned}$$

Then, summing on i and l, we obtain

$$\begin{aligned}&C_2(z)^{kp}\delta(z)^{kp/2}\left|\frac{\partial^k f}{\partial\zeta_1^k}(0)\right|^p\\ &\qquad\leq C\,\mathrm{Mean}^{R(z)}\left(\sum_{i=2}^n\left(|L'^k_i f|^p + (\mathrm{Rest}_i^k f)^p\right)\right).\end{aligned}$$

And we estimate $\mathrm{Rest}_i^k f$ as in the proof of Corollary 1.5. In the ordinary system of coordinates, this gives

$$\begin{aligned}&C_2(z)^{kp}\delta(z)^{kp}\left|\left(\frac{\partial r}{\partial z_1}(z)\right)^{-k}\frac{\partial^k g}{\partial z_1^k}(z)\right|^p\\ &\qquad\leq C\,\mathrm{Mean}^{Q(z)}\left(\delta(z)^{kp/2}|\nabla_T^k g|^p + \mathrm{Rest}^k(g)^p\right).\end{aligned}$$

Lemma 1.7 follows with

$$M_z = \left(\frac{\partial r}{\partial z_1}(z)\right)^{-1} \frac{\partial}{\partial z_1}.$$

In order to conclude for Theorem B, we remark that there exists a constant C such that, for every $z \in V(z_0)$, we have

$$|\nabla^k g(z)| \leq C\Big\{|M_z^k g(z)| + \sum_{\substack{j+r=k \\ r\geq 1}} |\nabla_T^r \nabla^j g(z)| + \sum_{1\leq j\leq k-1} |\nabla^j g(z)|\Big\}.$$

But, by Corollary 1.5, we can estimate each $|\nabla_T^r \nabla^j g(z)|$ by its mean-value on $Q(z)$, disregarding some remaining terms. This allows to see that the terms

$$\sum_{\substack{j+r=k \\ r\geq 1}} \delta(z)^k |\nabla_T^r \nabla^j g(z)| + \sum_{1\leq j\leq k-1} \delta(z)^k |\nabla^j g(z)|$$

can be majorized by $C\,\mathrm{Mean}^{Q(z)}(\mathrm{Rest}^k(g))$.

REMARK. The estimate of Theorem B is intrinsic: explicitely, if $\hat{\nabla}_T^k$ is a tangential gradient defined with the help of another defining function $\hat{r}$, then the right member of the estimate defined with the help of these $\hat{\nabla}_T$ is equivalent to the one defined with the help of ∇_T.

We will assume now that Ω is bounded in $\mathbb{C}^n$; so, the estimates of Theorem A and B are uniformly true on $\Omega \cap U$, where U is a neighborhood of $\partial\Omega$ sufficiently small such that the projection on $\partial\Omega$ is well defined. Furthermore, if Ω satisfies (P), we assume U sufficiently small so that C_2 is uniformly bounded from below on $\overline{\Omega} \cap U$. In this case, we obtain the following corollary.

Corollary. *Let Ω be a bounded C^∞ domain satisfying* (P). *For every $0 < p < \infty$, $k \in \mathbb{N}$, there exists a constant C such that, for every holomorphic function g in Ω and every $z \in \Omega \cap U$, we have*

$$\delta(z)^{kp} |\nabla^k g(z)|^p \leq C\,\mathrm{Mean}^{Q(z)}\left(\delta^{kp/2} |\nabla_T^k g|^p + \mathrm{Rest}^k(g)^p\right).$$

2. Hardy-Sobolev Spaces.

In the following K will denote a compact set contained in the complement of $\Omega \cap U$.

2.1. Proof of the Auxiliary Theorem.

We give the main ideas and references for the proof of this theorem, which follows from standard methods but which is nowhere explicitely written (as far as we know).

Assume that $k = 0$ to simplify. The equivalence between (1) and (3) is well known from Fefferman-Stein work and is valid for harmonic functions (see [FS]). We have to prove that (1) implies (4).

For every $0 < p < \infty$, every $\Phi(\eta, t) \in \mathcal{A}_\alpha(\zeta)$, we have, by subharmonicity of $|g|^{p/2}$

$$\begin{aligned} |g \circ \Phi(\eta,t)|^{p/2} &\leq C\,\mathrm{Mean}^{Q(\Phi(\eta,t))}(|g|^{p/2}) \\ &\leq \frac{C}{|Q(\Phi(\eta,t))|}\int_{\Phi^{-1}(Q(\Phi(\eta,t)))} |g\circ\Phi(\eta',s)|^{p/2}d\sigma(\eta')ds \\ &\leq \frac{C}{|B^d(\zeta,ct)|}\int_{B^d(\zeta,ct)}\sup_{0<s<s_0}|g\circ\Phi(\eta',s)|^{p/2}d\sigma(\eta')\,, \end{aligned}$$

since if $\Phi(\eta,t) \in \mathcal{A}_\alpha(\zeta)$, the projection of $\Phi^{-1}\left(Q(\Phi(\eta,t))\right)$ on $\partial\Omega$ is contained in $B^d(\zeta,ct)$, for some constant c.

So,

$$\mathcal{M}_\alpha(|g|)^{p/2}(\zeta) \leq C\,M\left(\sup_{0<s<s_0}|g\circ\Phi(\cdot,s)|^{p/2}\right)(\zeta)\,,$$

where M is the non-isotropic maximal operator. We conclude by the L^2-continuity of the operator M.

The fact that (2) implies (3) follows from a similar argument. For $p > 2$, (3) implies (2) is true for any kind of regular functions (see Lemma 2.5 further on) and follows from an argument of duality and from the L^q-continuity of the non-isotropic maximal operator for $q > 1$. It remains to prove that (4) implies (2) when $p \leq 2$. The proof of Fefferman-Stein can be adapted in this context. We postpone this proof as we shall adapt the method of Fefferman-Stein in a more general context for Theorem D. We can also see [B1] and [B2].

REMARK. As an immediate consequence of the equivalence between

$$\left\| \sup_{0<t<s_0} |\nabla^k g \circ \Phi(\cdot,t)| \right\|_{L^p(\partial\Omega)} \quad \text{and} \quad \|S_\alpha(\delta \nabla^{k+1} g)\|_{L^p(\partial\Omega)},$$

we obtain that the spaces $\mathcal{H}^p_k(\Omega)$ are independent of the choice of the map Φ, and that different choices of Φ yield equivalent norms.

2.2. Admissible area functions.

First, we give an auxiliary result which can be proved by the same method as in the case of classical area functions (we do not give the proof because the method will be largely used in the following, we can also see [CMS] for instance).

Lemma 2.1 *Let Ω be a C^2-domain in $\mathbb{C}^n$. For every apertures $\alpha, \beta > 0$, every $0 < p < \infty$, every $q \in \mathbb{R}$,*

$$\|S^q_\alpha(u)\|_{L^p(\partial\Omega)} \text{ and } \left\|S^q_\beta(u)\right\|_{L^p(\partial\Omega)}$$

a. equivalent for every regular function u defined on Ω.

Now, we need a particular Hardy Inequality.

Lemma 2.2. (A Hardy Inequality on a region over a graph). *Let $\mathcal{R}$ be a region over a graph in $\Omega \cap U$, $\mathcal{R}$ given by*

$$\Phi^{-1}(\mathcal{R}) = \{(\eta, t) \in \partial\Omega \times]0, s_0[: \quad t \geq \phi(\eta)\}$$

for some function ϕ.

Let $q > 0$. There exists a constant C such that, for every mesurable function u on Ω, we have

$$\iint_{\mathcal{R}} \delta^q |u|^2 \frac{dV}{\delta} \leq C\Big(\iint_{\mathcal{R}} \delta^{q+2k} |\nabla^k u|^2 \frac{dV}{\delta} + \sum_{j=0}^{k-1} \|\nabla^j u\|^2_{L^2(K)} \Big).$$

PROOF OF LEMMA 2.2. First, we recall the usual Hardy-inequality:

Let $p \geq 1$. For every $q > 0$, there exists a constant C such that, for any positive, mesurable function v defined on $\mathbb{R}^+$, we have

$$\int_0^\infty t^q V(t)^p \frac{dt}{t} \leq C \int_0^\infty t^q \left(tv(t)\right)^p \frac{dt}{t},$$

where $V(t) = \int_t^\infty v(s)ds$ for $t \geq 0$.

In order to obtain the lemma, for each $\eta \in \partial\Omega$, we apply this inequality with $p = 2$ successively to the function

$$V_\eta(t_1) = \int_{t_1}^\infty \cdots \int_{t_k}^\infty |v_\eta(t)|\, dt dt_k \ldots dt_2$$

$$\text{where } v_\eta(t) = \begin{cases} \dfrac{d^k u}{dt^k}\left(\Phi(\eta, t)\right) & \text{if } \phi(\eta) \leq t \leq s_0, \\ 0 \quad \text{otherwise.} \end{cases}$$

Integrating over $\partial\Omega$, this gives the result.

We are going to prove, now, that the admissible area functions of different orders are equivalent. This equivalence follows from standard arguments and from an appropriate Hardy Inequality.

Theorem 2.3. *Let Ω be a bounded C^∞ domain in $\mathbb{C}^n$. For every holomorphic function g, every $k \in \mathbb{N}$, every $0 < p < \infty$, every $q > 0$ and every aperture $\alpha > 0$*

$$\|S_\alpha^q(g)\|_{L^p(\partial\Omega)} \text{ and } \left\|S_\alpha^q(\delta^k \nabla^k g)\right\|_{L^p(\partial\Omega)}$$

are equivalent, modulo an error of $\|g\|_{L^2(K)}$.

In [S2], Stein gives this result for $p = 2$ and harmonic functions in $\mathbb{R}^n$. For general p, the corresponding result for harmonic functions in $\mathbb{R}^n$ follows from equivalent definitions of $\mathcal{H}^p$ (see [CT] for instance).

PROOF. The proof will be given in two steps. The first step is devoted to show that

$$\left\|S_\alpha^q(\delta^k \nabla^k g)\right\|_{L^p(\partial\Omega)} \leq C\left(\|S_\alpha^q(g)\|_{L^p(\partial\Omega)} + \|g\|_{L^2(K)}\right)$$

and the second to the converse inequality.

First inequality. In this part, we are going to show that, for every $\zeta \in \partial\Omega$, every $\alpha > 0$, there exists $\beta > \alpha$ such that

$$S^q_\alpha(\delta^k \nabla^k g)(\zeta) \leq C \left(S^q_\beta(g)(\zeta) + \|g\|_{L^2(K)} \right).$$

Lemma 2.1 will allow to conclude that, for every $0 < p < \infty$,

$$\left\| S^q_\alpha(\delta^k \nabla^k g) \right\|_{L^p(\partial\Omega)} \leq C \left(\|S^q_\alpha(g)\|_{L^p(\partial\Omega)} + \|g\|_{L^2(K)} \right).$$

This inequality follows easily from the fact that, since g is holomorphic in Ω, for every $z \in \Omega \cap U$,

$$\delta(z)^{2k} |\nabla^k g|^2(z) \leq C \, \mathrm{Mean}^{Q(z)}(|g|^2).$$

So, it suffices to choose β sufficiently large such that, for every $z \in \mathcal{A}_\alpha(\zeta)$, $Q(z) \subset \mathcal{A}_\beta(\zeta)$.

Converse inequality. In order to show the converse inequality, we are going to distinguish three cases: $p = 2$, $0 < p < 2$ and $p > 2$.

1. *Case* $p = 2$. This case is the simplest one since

$$\begin{aligned} \|S^q_\alpha(g)\|^2_{L^2(\partial\Omega)} &\simeq \iint_{\mathcal{R}_\alpha} \delta^q(z) \, |g(z)|^2 \, \sigma\left(\{\zeta \in \partial\Omega : \, z \in \mathcal{A}_\alpha(\zeta)\}\right) \frac{dV(z)}{\delta^{n+1}(z)} \\ &\simeq \iint_{\mathcal{R}_\alpha} |g|^2 \delta^q \frac{dV}{\delta} \end{aligned}$$

where $\mathcal{R}_\alpha$ is the set $\cup_{\zeta \in \partial\Omega} \mathcal{A}_\alpha(\zeta) = \Omega \cap U$. So

$$\|S^q_\alpha(g)\|^2_{L^2(\partial\Omega)} \simeq \int_{\Omega \cap U} |g|^2 \delta^q \frac{dV}{\delta}$$

and it suffices to apply Hardy-inequality to conclude.

2. *Case* $0 < p < 2$. We use again ideas of [FS] and [CT]. The proof will be given in two parts.

2.1. *First part:* For every $\lambda > 0$ and every $\beta > \alpha > 0$, let $E \, (= E^\lambda)$ be the set of points of $\partial\Omega$ where $S^q_\beta(\delta^k \nabla^k g) \leq \lambda$. Now, let $E_0 \, (= E^\lambda_0)$ be the points of E of relative density $1/2$; more precisely E_0 is the set

$$\left\{ \zeta \in \partial\Omega : \text{ for every ball } B^d \text{ containing } \zeta, \, \sigma(E \cap B^d) \geq \frac{1}{2} \sigma(B^d) \right\}.$$

If χ is the characteristic function of $D\,(=D^\lambda) = E^c$ (complementary of E), then

$$D_0\,(=D_0^\lambda) = E_0^c = \left\{\zeta \in \partial\Omega :\ M(\chi) > \frac{1}{2}\right\},$$

where M is the non-isotropic maximal operator. Thus, there exists a constant c such that $\sigma(D_0) \leq c\,\sigma(D)$.

We are going to prove the following lemma.

Lemma 2.4. *Under the assumptions of Theorem* 2.3, *there exists a constant C such that*

$$\int_{E_0} S_\alpha^q(g)^2 d\sigma \leq C\left(\int_E S_\beta^q(\delta^k\nabla^k g)^2 d\sigma + \int_K |g|^2 dV\right).$$

PROOF. We have

$$\begin{aligned}(*) &= \int_{E_0} S_\alpha^q(g)^2 d\sigma \\ &= \iint_{\mathcal{R}_\alpha} \delta^q(z)\,|g(z)|^2\,\sigma\left(\{\zeta \in E_0 :\ z \in \mathcal{A}_\alpha(\zeta)\}\right)\frac{dV(z)}{\delta^{n+1}(z)} \\ &\leq C\iint_{\mathcal{R}_\alpha} \delta^q\,|g|^2\,\frac{dV}{\delta}\,,\end{aligned}$$

where we denote by $\mathcal{R}_\alpha$ the set $\cup_{\zeta\in E_0}\mathcal{A}_\alpha(\zeta)$. Observe that

$$\Phi^{-1}\left(\mathcal{R}_\alpha\right) = \left\{(\eta,t) \in \partial\Omega\times]0,s_0[:\quad t \geq \frac{1}{\alpha}d(\eta,E_0)\right\}$$

(where $d(\eta,E_0) = \inf_{\zeta\in E_0} d(\eta,\zeta)$) and apply Lemma 2.2 in order to obtain

$$\iint_{\mathcal{R}_\alpha} \delta^q|g|^2\frac{dV}{\delta} \leq C\left(\iint_{\mathcal{R}_\alpha} \delta^{q+2k}|\nabla^k g|^2\frac{dV}{\delta} + \int_K |g|^2 dV\right),$$

since, by assumption $q > 0$. So, we have

$$\int_{E_0} S_\alpha^q(g)^2 d\sigma \leq C\left(\iint_{\mathcal{R}_\alpha} \delta^{q+2k}|\nabla^k g|^2\frac{dV}{\delta} + \int_K |g|^2 dV\right).$$

Now, it is sufficient to observe that $z = \Phi(\eta, t) \in \mathcal{R}_\alpha$ if and only if $d(\eta, \zeta) \leq \alpha t$ for some $\zeta \in E_0$. But then $z = \Phi(\eta, t) \in \mathcal{A}_\beta(w)$ whenever $d(\zeta, w) \leq (\beta - \alpha)t$. Thus

$$\begin{aligned}\sigma\left(\{w \in E:\ z \in \mathcal{A}_\beta(w)\}\right) &\geq \sigma(E \cap B^d(\zeta, (\beta-\alpha)t)) \\ &\geq \frac{1}{2}\,\sigma(B^d(\zeta, (\beta-\alpha)t))\end{aligned}$$

in view of the definition of E_0. So, the later quantity exceeds $Ct^n \simeq C\delta(z)^n$. So,

$$\begin{aligned}(*) &\leq C\Big(\iint_{\mathcal{R}_\alpha} \delta^{q+2k}|\nabla^k g(z)|^2 \sigma\left(\{w \in E:\ z \in \mathcal{A}_\beta(w)\}\right) \frac{dV}{\delta^{n+1}} \\ &\qquad + \int_K |g|^2 dV\Big) \\ &\leq C\left(\int_E S_\beta^q(\delta^k \nabla^k g)^2 d\sigma + \int_K |g|^2 dV\right).\end{aligned}$$

2.2. *Second part:* We conclude from Lemma 2.4 as in [FS] that, when $0 < p < 2$,

$$\|S_\alpha^q(g)\|_{L^p(\partial\Omega)} \leq C\left(\left\|S_\alpha^q(\delta^k \nabla^k g)\right\|_{L^p(\partial\Omega)} + \|g\|_{L^2(K)}\right).$$

Let us give the proof once for all for completeness.

Observe that

$$\begin{aligned}\sigma\left(\{S_\alpha^q(g) \geq \lambda\}\right) &\leq \sigma(D_0^\lambda) + \sigma\left(\left\{\zeta \in E_0^\lambda:\quad S_\alpha^q(g)(\zeta) \geq \lambda\right\}\right) \\ &\leq \sigma(D_0^\lambda) + \frac{1}{\lambda^2}\int_{E_0^\lambda} S_\alpha^q(g)^2 d\sigma\,.\end{aligned}$$

Then, we write

$$\begin{aligned}\|S_\alpha^q(g)\|_{L^p(\partial\Omega)}^p &= p\int_0^\infty \lambda^{p-1}\sigma\left(\{S_\alpha^q(g) \geq \lambda\}\right) d\lambda \\ &\leq p\int_0^\infty \lambda^{p-1}\sigma(D_0^\lambda)\, d\lambda \\ &\quad + p\int_M^\infty \lambda^{p-3}\int_{E_0^\lambda} S_\alpha^q(g)^2\, d\sigma d\lambda \\ &\quad + \sigma(\partial\Omega)M^p\end{aligned}$$

$$\leq C\Big(p\int_0^\infty \lambda^{p-1}\sigma(D^\lambda)\,d\lambda + p\int_M^\infty \lambda^{p-3}\int_{E^\lambda} S_\beta^q(\delta^k\nabla^k g)^2\,d\sigma d\lambda + \|g\|_{L^2(K)}^2 M^{p-2} + \sigma(\partial\Omega)M^p\Big)$$
$$\leq C\left(\left\|S_\beta^q(\delta^k\nabla^k g)\right\|_{L^p(\partial\Omega)}^p + \|g\|_{L^2(K)}^p\right)$$

(by choosing $M = \|g\|_{L^2(K)}$). Lemma 2.1 allows to conclude.

3. *Case* $2 < p < \infty$. First, it follows easily from the usual Hardy-inequality in $L^2(0, s_0)$ that, for every $q > 0$ and every $\zeta \in \partial\Omega$,

$$G^q(g)(\zeta) \leq C\left(G^q(\delta^k\nabla^k g)(\zeta) + \|g\|_{L^2(K)}\right).$$

So, for every $0 < p < \infty$,

$$\|G^q(g)\|_{L^p(\partial\Omega)} \leq C\left(\left\|G^q(\delta^k\nabla^k g)\right\|_{L^p(\partial\Omega)} + \|g\|_{L^2(K)}\right).$$

Now, since for every function $u \in C^\infty(\Omega)$ and every $\zeta \in \partial\Omega$,

$$G^q(\mathrm{Mean}^Q(u))(\zeta) \leq C\left(S_\alpha^q(u)(\zeta) + \|u\|_{L^2(K)}\right)$$

for some $\alpha > 0$, we have

$$\|G^q(g)\|_{L^p(\partial\Omega)} \leq C\left(\|S_\alpha^q(g)\|_{L^p(\partial\Omega)} + \|g\|_{L^2(K)}\right)$$

for every holomorphic function g and every $0 < p < \infty$. So, it remains to show the following lemma.

Lemma 2.5. *For every $\alpha > 0$, every $2 < p < \infty$, every $u \in C^\infty(\Omega)$ and every $q \in \mathbb{R}$, there exists a constant C such that*

$$\|S_\alpha^q(u)\|_{L^p(\partial\Omega)} \leq C\,\|G^q(u)\|_{L^p(\partial\Omega)}.$$

PROOF. As in [S2], we use the fact that

$$\|S_\alpha^q(u)\|_{L^p(\partial\Omega)}^2 = \left\|S_\alpha^q(u)^2\right\|_{L^{p/2}(\partial\Omega)}$$

$$= \sup \int_{\partial\Omega} S^q_\alpha(u)(\zeta)^2 v(\zeta) d\sigma(\zeta)$$

where the supremum is taken over all the functions $v \in L^{p'}(\partial\Omega)$, with $2/p + 1/p' = 1$ and $\|v\|_{L^{p'}(\partial\Omega)} \leq 1$.

$$\begin{aligned}
&\int_{\partial\Omega} S^q_\alpha(u)^2(\zeta) v(\zeta) d\sigma(\zeta) \\
&\quad = C \int_{\partial\Omega} \int_{\Phi^{-1}(\mathcal{A}_\alpha(\zeta))} t^q |u|^2 \circ \Phi(\eta, t) \frac{dt}{t^{n+1}} \, d\sigma(\eta) v(\zeta) \, d\sigma(\zeta) \\
&\quad = C \int_{\partial\Omega} \int_0^{s_0} t^q |u|^2 \circ \Phi(\eta, t) \left(\frac{1}{t^n} \int_{B^d(\eta, \alpha t)} v(\zeta) d\sigma(\zeta) \right) \frac{dt}{t} \, d\sigma(\eta) \\
&\quad \leq C \int_{\partial\Omega} M v(\eta) G^q(u)(\eta)^2 d\sigma(\eta) \,,
\end{aligned}$$

where M is the non-isotropic maximal operator,

$$\begin{aligned}
&\leq C \, \|Mv\|_{L^{p'}(\partial\Omega)} \, \|G^q(u)\|^2_{L^p(\partial\Omega)} \\
&\leq C \|v\|_{L^{p'}(\partial\Omega)} \, \|G^q(u)\|^2_{L^p(\partial\Omega)} \\
&\leq C \, \|G^q(u)\|^2_{L^p(\partial\Omega)}
\end{aligned}$$

by the $L^{p'}$-continuity of the non-isotropic maximal operator.

2.3. Proof of the first part of Main Theorem.

We are going to prove the following result.

Theorem C. *Let Ω be a bounded C^∞-domain in $\mathbb{C}^n$ satisfying* (P), *$\alpha > 0$ be a fixed aperture and $0 < p < \infty$. For every holomorphic function g in Ω, every $q \in \mathbb{R}$, $k, l, r \in \mathbb{N}$ with $q + k + 2l > 0$ and $q + 2r > 0$,*

$$\left\| S^q_\alpha(\delta^{k/2+l} \nabla^k_T \nabla^l g) \right\|_{L^p(\partial\Omega)}, \quad \left\| S^q_\alpha(\delta^{k/2+l} \nabla^l \nabla^k_T g) \right\|_{L^p(\partial\Omega)},$$

$$\text{and} \qquad \|S^q_\alpha(\delta^r \nabla^r g)\|_{L^p(\partial\Omega)}$$

are equivalent, modulo an error of $\|g\|_{L^2(K)}$.

These results are also true for any permutation of ∇ and ∇_T in a product $\nabla^l \nabla_T^k$.

As an immediate application, we obtain the following corollary.

Corollary. *Let Ω be a bounded C^∞-domain in $\mathbb{C}^n$ satisfying* (P). *For every $0 < p < \infty$, every $k \in \mathbb{N}$ and every holomorphic function g, we have*

$$g \in \mathcal{H}_k^p(\Omega) \quad \textit{if and only if} \quad \left\| S_\alpha^{-2k}(\delta^{j+l/2} \nabla^j \nabla_T^l g) \right\|_{L^p(\partial\Omega)} < \infty$$

for every $l, j \in \mathbb{N}$ with $j + l/2 > k$.

In particular for $l = 2k$ and $j = 1$, this corollary gives the equivalence between (1) and (2) of Main Theorem and for $k = 0$, $l = 2$ and $j = 0$, this gives the corollary stated in the introduction.

PROOF OF THEOREM C. We are going to show that

$$\left\| S_\alpha^q(\delta^{k/2+l} \nabla_T^k \nabla^l g) \right\|_{L^p(\partial\Omega)} \quad \text{and} \quad \| S_\alpha^q(\delta^r \nabla^r g) \|_{L^p(\partial\Omega)}$$

are equivalent, under the assumption that $q + 2r > 0$ and $q + 2l + k > 0$.

In order to obtain the last equivalence of Theorem C, we just have to use Lemma 2.2 which allows us to write $\nabla^l \nabla_T^k g$ as the sum of $\nabla_T^k \nabla^l g$ and of terms involving smaller derivatives like $\nabla_T^r \nabla^j g$, with $0 \leq r \leq k - 1$ and $1 \leq j \leq l$ (the terms involving smaller derivatives are smaller than $\| S_\alpha^q(\delta^r \nabla^r g) \|_{L^p(\partial\Omega)}$).

So, let us show this equivalence.

First inequality. We are going to show that

$$\left\| S_\alpha^q(\delta^{k/2+l} \nabla_T^k \nabla^l g) \right\|_{L^p(\partial\Omega)} \leq C \left(\| S_\alpha^q(\delta^r \nabla^r g) \|_{L^p(\partial\Omega)} + \| g \|_{L^2(K)} \right),$$

under the assumption that $q + k + 2l > 0$ and $q + 2r > 0$. As before, we will distinguish the cases $p = 2$, $0 < p < 2$ and $p > 2$.

The case $p = 2$ is the simplest one as before. Since it follows easily from the results of [G1], we will not repeat the proof here.

1. *Case* $0 < p < 2$. In view of the proof of Theorem 2.3, it is sufficient to prove the following inequality, the second step being the same as in the proof of Theorem 2.3.

There exists $\gamma > \alpha$ such that

$$\int_{E_0} S_\alpha^q(\delta^{k/2+l}\nabla_T^k\nabla^l g)^2 d\sigma \leq C\left(\int_E S_\gamma^q(\delta^r\nabla^r g)^2 d\sigma + \|g\|^2_{L^2(K)}\right)$$

with E, E_0 corresponding to $S_\gamma^q(\delta^r\nabla^r g)$. This inequality will follow from the following lemma.

Lemma 2.6. *Let $k, l, r \in \mathbb{N}$, $q \in \mathbb{R}$ with $q + 2l + k > 0$ and $q + 2r > 0$. Then, for every $\alpha > 0$, there exist $\beta > \alpha$ and a constant C such that, for every $E_0 \subset \partial\Omega$ and every holomorphic function g in Ω, we have*

$$\iint_{\mathcal{R}_\alpha} \delta^{q+2l+k} \left|\nabla_T^k\nabla^l g\right|^2 \frac{dV}{\delta} \leq C\left(\iint_{\mathcal{R}_\beta} \delta^{q+2r} \left|\nabla^r g\right|^2 \frac{dV}{\delta} + \|g\|^2_{L^2(K)}\right),$$

$$\iint_{\mathcal{R}_\alpha} \delta^{q+2l+k} \left|\nabla^l\nabla_T^k g\right|^2 \frac{dV}{\delta} \leq C\left(\iint_{\mathcal{R}_\beta} \delta^{q+2r} \left|\nabla^r g\right|^2 \frac{dV}{\delta} + \|g\|^2_{L^2(K)}\right),$$

where $\mathcal{R}_\alpha = \cup_{\zeta\in E_0}\mathcal{A}_\alpha(\zeta)$.

PROOF. We only give the proof for $l = 0$. The general result for the first inequality will follow applying the result with $l = 0$ to the components of $\nabla^l g$, changing q into $q + 2l$ and using the same method as in the proof of Theorem 2.3. For the second inequality, we use Lemma 1.6 as before.

Let us apply the Hardy Inequality of Lemma 2.2, this gives

$$\iint_{\mathcal{R}_\alpha} \delta^{q+k} \left|\nabla_T^k g\right|^2 \frac{dV}{\delta} \leq C\left(\iint_{\mathcal{R}_\alpha} \delta^{q+2r+k} \left|\nabla^r\nabla_T^k g\right|^2 \frac{dV}{\delta} + \|g\|^2_{L^2(K)}\right),$$

under the assumption that $q + k > 0$.

Now, we apply successively the results of Theorem A and the Hardy Inequality of Lemma 2.2 to the terms involving derivatives less than r in order to obtain

$$\iint_{\mathcal{R}_\alpha} \delta^{q+k} \left|\nabla_T^k g\right|^2 \frac{dV}{\delta} \leq C \left(\sum_{j=1}^{r} \iint_{\mathcal{R}_\beta} \delta^{q+2r} \left|\nabla^j g\right|^2 \frac{dV}{\delta} + \|g\|^2_{L^2(K)} \right)$$
$$\leq C \left(\iint_{\mathcal{R}_\beta} \delta^{q+2r} \left|\nabla^r g\right|^2 \frac{dV}{\delta} + \|g\|^2_{L^2(K)} \right)$$

since, by assumption, $q + 2r > 0$.

2. *Case* $2 < p < \infty$. As before, we will only consider the case $l = 0$. We have, by Lemma 2.5,

$$\left\| S_\alpha^q(\delta^{k/2} \nabla_T^k g) \right\|_{L^p(\partial\Omega)} \leq C \left\| G^q(\delta^{k/2} \nabla_T^k g) \right\|_{L^p(\partial\Omega)}.$$

But, by Hardy inequality in $L^2(0, s_0)$, we have, for every $\zeta \in \partial\Omega$,

$$G^q(\delta^{k/2} \nabla_T^k g)(\zeta) \leq C \left(G^q(\delta^{k/2+r} \nabla^r \nabla_T^k g)(\zeta) + C\|g\|_{L^2(K)} \right)$$
$$\leq C \left(S_\alpha^q(\delta^r \sum_{j=1}^{r} |\nabla^j g|)(\zeta) + \|g\|_{L^2(K)} \right)$$

by Theorem A, since for every function $u \in C^\infty(\Omega)$ and every $\zeta \in \partial\Omega$, there exists $\alpha > 0$ such that

$$G^q(\mathrm{Mean}^Q(u))(\zeta) \leq C \left(S_\alpha^q(u)(\zeta) + \|u\|_{L^2(K)} \right).$$

So, by Theorem 2.3, we obtain the result.

$$\left\| S_\alpha^q(\delta^{k/2} \nabla_T^k g) \right\|_{L^p(\partial\Omega)} \leq C \left(\|S_\alpha^q(\delta^r \nabla^r g)\|_{L^p(\partial\Omega)} + \|g\|_{L^2(K)} \right).$$

Converse inequality. We want to show that,

$$\|S_\alpha^q(\delta^r \nabla^r g)\|_{L^p(\partial\Omega)} \leq C \left(\left\| S_\alpha^q(\delta^{k/2+l} \nabla_T^k \nabla^l g) \right\|_{L^p(\partial\Omega)} + \|g\|_{L^2(K)} \right),$$

under the assumptions that $q + k + 2l > 0$ and $q + 2r > 0$. We only consider the case $l = 0$. For general l, we use the same method as before.

By Theorem 2.3, it suffices to show that

$$\left\|S_\alpha^q(\delta^k\nabla^k g)\right\|_{L^p(\partial\Omega)} \leq C\left(\left\|S_\alpha^q(\delta^{k/2}\nabla_T^k g)\right\|_{L^p(\partial\Omega)} + \|g\|_{L^2(K)}\right).$$

We begin with the following lemma.

Lemma 2.7. *Under the assumptions of Theorem* C, *we have*

$$\left\|S_\alpha^q(\delta^k\nabla^k g)\right\|_{L^p(\partial\Omega)} \leq C\Big(\left\|S_\alpha^q(\delta^{k/2}\nabla_T^k g)\right\|_{L^p(\partial\Omega)} + \left\|S_\alpha^{q+1}(\delta^k\nabla^k g)\right\|_{L^p(\partial\Omega)} \|g\|_{L^2(K)}\Big).$$

PROOF. 1. *Case* $0 < p < 2$. Lemma 2.7 will follow from the following estimate.

There exists $\gamma > \alpha$ and a constant C such that,

$$\int_{E_0} S_\alpha^q(\delta^k\nabla^k g)^2 d\sigma \leq C\Big(\int_E S_\gamma^q(\delta^{k/2}\nabla_T^k g)^2\, d\sigma + \int_E S_\gamma^{q+1}(\delta^k\nabla^k g)^2 d\sigma + \|g\|_{L^2(K)}^2\Big)$$

where

$$E = \left\{S_\gamma^{q+1}(\delta^k\nabla^k g) \leq \lambda \text{ and } S_\gamma^q(\delta^{k/2}\nabla_T^k g) \leq \lambda\right\}$$

and E_0 is the set of points of E of relative density $1/2$. As in the proof of Theorem 2.3, we will denote by D_0 and by D the complements of E_0 and E respectively; then $\sigma(D_0) \leq c\,\sigma(D)$ by the non-isotropic maximal Theorem.

The preceding inequality will follow from the following estimate.

There exists $\beta' > \beta > \alpha$ such that

$$\begin{aligned}(*) &= \iint_{\mathcal{R}_\alpha} \delta^{q+2k}|\nabla^k g|^2 \frac{dV}{\delta} \\ &\leq C\Big(\iint_{\mathcal{R}_\beta} \delta^{q+k}|\nabla_T^k g|^2 \frac{dV}{\delta} + \iint_{\mathcal{R}_{\beta'}} \delta^{q+2k+1}|\nabla^k g|^2 \frac{dV}{\delta} + \|g\|_{L^2(K)}^2\Big).\end{aligned}$$

Let us prove this inequality. By Theorem B, we have

$$(*) \leq C\Big(\iint_{\mathcal{R}_\beta} \delta^{q+k} |\nabla_T^k g|^2 \frac{dV}{\delta} + \iint_{\mathcal{R}_\beta} \delta^q |\mathrm{Rest}^k(g)|^2 \frac{dV}{\delta} + \|g\|^2_{L^2(K)} \Big).$$

Let us estimate the remaining term. We are going to show that it is bounded by

$$C\left(\iint_{\mathcal{R}_{\beta'}} \delta^{q+2k+1} |\nabla^k g|^2 \frac{dV}{\delta} + \|g\|^2_{L^2(K)} \right).$$

We have

$$\begin{aligned}(**) &= \iint_{\mathcal{R}_\beta} \delta^q |\mathrm{Rest}^k(g)|^2 \frac{dV}{\delta} \\ &\leq C\Big(\iint_{\mathcal{R}_\beta} \sum_{1 \leq j+r<(k+r)/2} \delta^{k+q} \left|\nabla^j \nabla_T^r g\right|^2 \frac{dV}{\delta} \\ &\qquad + \iint_{\mathcal{R}_\beta} \sum_{(k+r)/2 \leq j+r \leq k} \delta^{2j+r+1+q} \left|\nabla^j \nabla_T^r g\right|^2 \frac{dV}{\delta} \Big).\end{aligned}$$

By Hardy inequality of Lemma 2.2, we have

$$\begin{aligned}(**) \leq C \Bigg(&\iint_{\mathcal{R}_\beta} \sum_{1 \leq j+r<(k+r)/2} \delta^{k+q+2(k-j)} \left|\nabla^k \nabla_T^r g\right|^2 \frac{dV}{\delta} \\ &+ \iint_{\mathcal{R}_\beta} \sum_{(k+r)/2 \leq j+r \leq k} \delta^{2k+r+1+q} \left|\nabla^k \nabla_T^r g\right|^2 \frac{dV}{\delta} \Bigg).\end{aligned}$$

Then, by Theorem A, we obtain

$$\begin{aligned}(**) \leq C\Big(&\iint_{\mathcal{R}_\beta} \sum_{1 \leq j+r<(k+r)/2} \sum_{l=1}^{k} \delta^{k+q+2(k-j)-r} \left|\nabla^l g\right|^2 \frac{dV}{\delta} \\ &+ \iint_{\mathcal{R}_\beta} \sum_{l=1}^{k} \delta^{2k+1+q} \left|\nabla^l g\right|^2 \frac{dV}{\delta} \Big)\end{aligned}$$

$$\leq C \iint_{\mathcal{R}_{\beta'}} \delta^{q+2k+1} |\nabla^k g|^2 \frac{dV}{\delta}.$$

So, we obtain the good estimate.

Using this inequality, we obtain Lemma 2.7 by the same method as in the proof of Theorem 2.3 when $0 < p < 2$ since

$$E \subset \left\{ S_\gamma^q(\delta^{k/2} \nabla_T^k g) \leq \lambda \right\}, \quad E \subset \left\{ S_\gamma^{q+1}(\delta^k \nabla^k g) \leq \lambda \right\}$$

and

$$\sigma(D) \leq \sigma\left(\left\{ S_\gamma^{q+1}(\delta^k \nabla^k g) \geq \lambda \right\} \right) + \sigma\left(\left\{ S_\gamma^q(\delta^{k/2} \nabla_T^k g) \geq \lambda \right\} \right).$$

2. *Case* $2 < p < \infty$. We have

$$\left\| S_\alpha^q(\delta^k \nabla^k g) \right\|_{L^p(\partial\Omega)} \leq C \left\| G^q(\delta^k \nabla^k g) \right\|_{L^p(\partial\Omega)}.$$

But, by Theorem B, we have, for every $\zeta \in \partial\Omega$,

$$G^q(\delta^k \nabla^k g)(\zeta) \leq C \left(S_\alpha^q(\delta^{k/2} \nabla_T^k g)(\zeta) + S_\alpha^q(\mathrm{Rest}^k(g))(\zeta) + \|g\|_{L^2(K)} \right).$$

We estimate the remaining terms $\left\| S_\alpha^q(\mathrm{Rest}^k(g)) \right\|_{L^p(\partial\Omega)}$ by Theorem 2.3 and the first inequality we have just proved, in order to obtain

$$\begin{aligned} \left\| G^q(\delta^k \nabla^k g) \right\|_{L^p(\partial\Omega)} \leq C \Big(& \left\| S_\alpha^q(\delta^{k/2} \nabla_T^k g) \right\|_{L^p(\partial\Omega)} \\ & + \left\| S_\alpha^{q+1}(\delta^k \nabla^k g) \right\|_{L^p(\partial\Omega)} + \|g\|_{L^2(K)} \Big). \end{aligned}$$

Lemma 2.7 follows.

END OF THE PROOF OF THEOREM C. It remains to show that Theorem C follows from Lemma 2.7. We distinguish two cases.

1. *Case* $q + k \geq 1$. Observe that, since $\delta(z) \leq s_0$ on $\mathcal{A}_\alpha(\zeta)$ for every $\zeta \in \partial\Omega$,

$$\left\| S_\alpha^{q+1}(\delta^k \nabla^k g) \right\|_{L^p(\partial\Omega)} \leq C s_0^{1/2} \left\| S_\alpha^q(\delta^k \nabla^k g) \right\|_{L^p(\partial\Omega)}.$$

So, we apply Lemma 2.7 in $\Omega_\varepsilon = \{\Phi(\zeta, t) \in \Omega,\ t > \varepsilon\}$ to g (which belongs to $C^\infty(\overline{\Omega_\varepsilon})$). Reducing s_0 if necessary, we get

$$\left\| S_{\alpha,\varepsilon}^q(\delta^k \nabla^k g) \right\|_{L^p(\partial\Omega_\varepsilon)} \leq C \left(\left\| S_{\alpha,\varepsilon}^q(\delta^{k/2} \nabla_T^k g) \right\|_{L^p(\partial\Omega_\varepsilon)} + \|g\|_{L^2(K)} \right)$$

where $S^q_{\alpha,\varepsilon}$ denotes the admissible area function corresponding to Ω_ε. We want to let $\varepsilon \to 0$.

Using Fatou's Lemma, it is sufficient to show that, for ε small enough,

$$\left\| S^q_{\alpha,\varepsilon}(\delta^{k/2}\nabla^k_T g) \right\|_{L^p(\partial\Omega_\varepsilon)} \leq C \left\| S^q_\alpha(\delta^{k/2}\nabla^k_T g) \right\|_{L^p(\partial\Omega)}.$$

We have

$$\left\| S^q_{\alpha,\varepsilon}(\delta^{k/2}\nabla^k_T g) \right\|^p_{L^p(\partial\Omega_\varepsilon)} = \int_{\partial\Omega_\varepsilon} \left(\int_{\mathcal{A}_\alpha(\zeta_\varepsilon)} \delta_\varepsilon^{k+q} \left|\nabla^k_T g\right|^2 \frac{dV}{\delta_\varepsilon} \right)^{p/2} d\sigma_\varepsilon.$$

Now, for $\zeta_\varepsilon = \Phi(\zeta,\varepsilon) \in \partial\Omega_\varepsilon$, $\mathcal{A}_\alpha(\zeta_\epsilon) \subset \mathcal{A}_\beta(\zeta)$ for some $\beta > \alpha$ and obviously, $\delta_\varepsilon \leq \delta$. This allows to conclude.

2. *Case* $q + k < 1$. We are going to use the method called "bootstrapping". Without loss of generality, we can assume that

$$\left\| S^q_\alpha(\delta^{k/2}\nabla^k_T g) \right\|_{L^p(\partial\Omega)} < \infty.$$

Then, in particular, for any ϱ sufficiently large such that $k + \varrho \geq 1$,

$$\left\| S^\varrho_\alpha(\delta^{k/2}\nabla^k_T g) \right\|_{L^p(\partial\Omega)} < \infty.$$

and by the preceding result

$$\left\| S^\varrho_\alpha(\delta^k \nabla^k g) \right\|_{L^p(\partial\Omega)} < \infty.$$

Now, choose $\varrho = q + m$ with $m \in \mathbb{N}$. Then, we apply Lemma 2.7 in order to obtain

$$\left\| S^{\varrho-1}_\alpha(\delta^k \nabla^k g) \right\|_{L^p(\partial\Omega)} < \infty.$$

Now, we repeat the same argument as long as the index of the weight in S_α is different from q.

As a corollary of Theorem C, we obtain the following.

Corollary 2.8. *Let Ω be a bounded C^∞-domain in $\mathbb{C}^n$, satisfying* (P). *For every $0 < p < \infty$, for every $\alpha > 0$, every $k, l \in \mathbb{N}$ and $q \in \mathbb{R}$ with*

$q+k+2l > 0$, there exists a constant C such that, for every holomorphic function g in Ω, we have

$$\left\| G^q(\delta^{l+k/2}\nabla^l\nabla_T^k g)\right\|_{L^p(\partial\Omega)} \leq C\left(\left\| S_\alpha^q(\delta^{l+k/2}\nabla^l\nabla_T^k g)\right\|_{L^p(\partial\Omega)} + \|g\|_{L^2(K)}\right).$$

PROOF. As in Corollary 1.5, we write

$$\delta^{2l+k}(z)|\nabla^l\nabla_T^k g|^2(z) \leq C\,\mathrm{Mean}^{Q(z)}\Big(\delta^{2l+k}|\nabla^l\nabla_T^k g|^2 + \delta^l\left[\mathrm{Rest}^k(\nabla^l g)\right]^2\Big).$$

So,

$$\left\| G^q(\delta^{l+k/2}\nabla^l\nabla_T^k g)\right\|_{L^p(\partial\Omega)} \leq C\Big(\left\| S_\alpha^q(\delta^{l+k/2}\nabla^l\nabla_T^k g)\right\|_{L^p(\partial\Omega)} + \left\| S_\alpha^q(\delta^l\mathrm{Rest}^k(\nabla^l g))\right\|_{L^p(\partial\Omega)}\Big).$$

So, it suffices to estimate the remaining term by Theorem C. We obtain

$$\left\| S_\alpha^q(\delta^l\mathrm{Rest}^k(\nabla^l g))\right\|_{L^p(\partial\Omega)} \leq C\, s_0^{1/2}\left(\left\| S_\alpha^q(\delta^{l+k/2}\nabla^l\nabla_T^k g)\right\|_{L^p(\partial\Omega)} + \|g\|_{L^2(K)}\right).$$

2.4. Admissible area and maximal functions.

In this paragraph, we continue the proof of Main Theorem.

Theorem D. *Let Ω be a bounded C^∞-domain in $\mathbb{C}^n$ satisfying* (P), *α be an aperture > 0. For every $0 < p < \infty$, there exists a constant C such that, for every holomorphic function g, we have*

$$\left\| S_\alpha(\delta\nabla\nabla_T^k g)\right\|_{L^p(\partial\Omega)} \leq C\left(\left\| \mathcal{M}_\alpha(\nabla_T^k g)\right\|_{L^p(\partial\Omega)} + \|g\|_{L^2(K)}\right).$$

As an immediate application of this result, we obtain that (3) implies (2) in Main Theorem.

PROOF OF THEOREM D. The proof of Theorem D will follow the same lines as the corresponding one of Fefferman-Stein (see [FS]). The differences are due to the fact that $\nabla_T^k g$ is no longer holomorphic or even harmonic in Ω although g is holomorphic.

First, we assume that $g \in C^\infty(\overline{\Omega})$ and we are going to show the following a priori inequality:

$$\left\|S_\alpha(\delta \nabla \nabla_T^k g)\right\|_{L^p(\partial\Omega)} \leq C\left(\left\|\mathcal{M}_\alpha(\nabla_T^k g)\right\|_{L^p(\partial\Omega)} + \|g\|_{L^2(K)}\right).$$

Let us assume this inequality proved. Then, it remains to show that this inequality is still valid for general g. We apply this inequality in Ω_ε to g holomorphic in Ω. One can verify that the constant involved is independent of $\varepsilon > 0$. We want to let $\varepsilon \to 0$ in the inequality. Let us observe that, for $\zeta_\varepsilon = \Phi(\zeta, \varepsilon) \in \partial\Omega_\varepsilon$, $\mathcal{A}_\alpha(\zeta_\epsilon) \subset \mathcal{A}_\beta(\zeta)$, for some $\beta > \alpha$. This allows to show that

$$\left\|\mathcal{M}_\alpha(\nabla_T^k g)\right\|_{L^p(\partial\Omega_\epsilon)} \leq C\left\|\mathcal{M}_\beta(\nabla_T^k g)\right\|_{L^p(\partial\Omega)}.$$

Now, we just have to apply Fatou's Lemma and the following usual result (see [FS] for instance).

Let $0 < p < \infty$ and α, $\beta > 0$. There exists a constant C such that, for every function u on Ω

$$\|\mathcal{M}_\alpha u\|_{L^p(\partial\Omega)} \leq C\|\mathcal{M}_\beta u\|_{L^p(\partial\Omega)}.$$

So, let us show the a priori inequality. As in the preceding, we are going to distinguish three cases: $p = 2$, $0 < p < 2$, $p > 2$. As the case when $p = 2$ is the simplest one and follows the same line as the case when $0 < p < 2$, we will only consider the cases $0 < p < 2$ and $p > 2$.

1. *Case* $0 < p < 2$. In the following, it will be convenient to have a defining function for Ω which is harmonic near $\partial\Omega$. We choose a point $x_0 \in K$ and denote by δ the Green's function for Ω with singularity x_0. Thus, δ is harmonic in $\Omega \setminus \{x_0\}$ and $\delta(z)$ is comparable with the distance to the boundary, for $z \in \Omega \cap U$.

Let λ and ε be any real positive numbers and E be the set

$$\left\{\zeta \in \partial\Omega : \quad \mathcal{M}_\alpha(\nabla_T^k g)(\zeta) \le \lambda\,, \quad S_\gamma^{-k}(\delta^{k+1}\nabla^{k+1} g)(\zeta) \le C(\varepsilon, s_0)\,\lambda\right\}\,,$$

for some $\gamma > \alpha$ and $C(\varepsilon, s_0) = (\varepsilon^2 + s_0)^{-1/2}$.

Let E_0 be those points of E of relative density $1/2$, D_0, D their complements. We are going to prove the following lemma which gives a "good λ" inequality of a new type.

Lemma 2.9. *Under the assumptions of Theorem* D, *there exists a constant C such that, for every $\varepsilon > 0$*

$$\begin{aligned}\int_{E_0} S_\alpha(\delta\nabla\nabla_T^k g)^2(z)d\sigma(z) \le C\Bigg(&\left(\frac{1}{\varepsilon^2}+1\right)\lambda^2\sigma(D_0)\\ &+\int_0^\lambda t\sigma\left(\left\{\mathcal{M}_\alpha(\nabla_T^k g)\ge t\right\}\right)dt\\ &+(\varepsilon^2+s_0)\int_E S_\gamma^{-k}(\delta^{k+1}\nabla^{k+1}g)^2(z)\,d\sigma(z)\\ &+\|g\|^2_{L^2(K)}\Bigg)\,.\end{aligned}$$

PROOF. We denote by $\mathcal{R}_\alpha$ the set $\cup_{z\in E_0}\mathcal{A}_\alpha(z)$ and

$$I_{E_0} = \int_{E_0} S_\alpha(\delta\nabla\nabla_T^k g)^2(\zeta)d\sigma(\zeta)\,.$$

Then

$$\begin{aligned}I_{E_0} &= \iint_{\mathcal{R}_\alpha} \delta^2|\nabla\nabla_T^k g|^2(z)\,\sigma\left(\{\zeta\in E_0 :\ z\in\mathcal{A}_\alpha(\zeta)\}\right)\frac{dV(z)}{\delta^{n+1}}\\ &\le C\iint_{\mathcal{R}_\alpha}\delta\,|\nabla\nabla_T^k g|^2(z)dV(z).\end{aligned}$$

We write that

$$2\,|\nabla\nabla_T^k g|^2 \le 2\,|\Delta(\nabla_T^k g)\cdot\nabla_T^k g| + \Delta|\nabla_T^k g|^2$$

and, following the method of Fefferman and Stein, we will estimate

$$\iint_{\mathcal{R}_\alpha}\delta\,\Delta|\nabla_T^k g|^2 dV$$

by applying Green's Theorem (we recall that δ is a Green's function for Ω). Let us denote by $d\hat{\sigma}$ the surface measure on $\partial\mathcal{R}_\alpha$. So, we obtain

$$\begin{aligned} I_{E_0} \leq & C\left(\iint_{\mathcal{R}_\alpha} \delta\,|\nabla_T^k g\cdot\Delta(\nabla_T^k g)|dV\right. \\ & \left. + \left(\int_{\partial\mathcal{R}_\alpha}\delta\,\frac{\partial|\nabla_T^k g|^2}{\partial\nu}d\hat{\sigma} - \int_{\partial\mathcal{R}_\alpha}\frac{\partial\delta}{\partial\nu}|\nabla_T^k g|^2 d\hat{\sigma}\right)\right) \\ = & (1)+(2)+(3)\,, \end{aligned}$$

where $\partial/\partial\nu$ denotes the outer normal derivative on $\partial\mathcal{R}_\alpha$.

1.1. *Estimate of the first term* (1). As g is holomorphic in Ω, we have $\Delta\nabla_T^k g = [\Delta, \nabla_T^k]g$ and so,

$$|\Delta\nabla_T^k g| \leq C \sum_{\substack{0\leq j\leq 2\\ 0\leq r\leq k-1}} |\nabla^j\nabla_T^r g|\,.$$

Now

$$\begin{aligned} (1) &\leq \iint_{\mathcal{R}_\alpha}\delta|\nabla_T^k g| \sum_{\substack{0\leq j\leq 2\\ 0\leq r\leq k-1}} |\nabla^j\nabla_T^r g|\,dV \\ &\leq \Big(\iint_{\mathcal{R}_\alpha}\delta^{-1+\mu}|\nabla_T^k g|^2 dV\Big)^{1/2} \\ &\quad\cdot\Big(\iint_{\mathcal{R}_\alpha}\delta^{3-\mu}\sum_{\substack{0\leq j\leq 2\\ 0\leq r\leq k-1}}|\nabla^j\nabla_T^r g|^2 dV\Big)^{1/2} \end{aligned}$$

for every $\mu > 0$. Let us apply Lemma 2.6 with $0 < \mu < 1$, in order to obtain

$$(1) \leq C\Big(s_0\iint_{\mathcal{R}_\beta}\delta^{k+2}|\nabla^{k+1}g|^2\frac{dV}{\delta} + \int_K |g|^2 dV\Big)$$

for some $\beta > \alpha$.

1.2. *Estimate of the second term* (2).

$$(2) \leq C\int_{\partial\mathcal{R}_\alpha}\delta\,|\nabla\nabla_T^k g|\,|\nabla_T^k g|\,d\hat{\sigma}\,,$$

We split $\partial\mathcal{R}_\alpha$ into three pieces $\partial\mathcal{R}_\alpha = F \cup F^{E_0} \cup F^{D_0}$ where

$$\Phi^{-1}(F) \subset \partial\Omega \times \{s_0\}, \quad \Phi^{-1}\left(F^{E_0}\right) \subset E_0$$

and

$$\Phi^{-1}\left(F^{D_0}\right) \subset D_0 \times (0, s_0).$$

So, we write

$$(2) \leq C\left(\int_F + \int_{F^{E_0}} + \int_{F^{D_0}}\right).$$

First, we have trivially

$$\int_F \delta\,|\nabla\nabla_T^k g|\,|\nabla_T^k g|\,d\hat{\sigma} \leq C\,\|g\|^2_{L^2(K)}.$$

Then, as $d\hat{\sigma} \leq C\,d\sigma$ along $F^{E_0} \cup F^{D_0}$, we have

$$\int_{F^{E_0}} \delta\,|\nabla\nabla_T^k g|\,|\nabla_T^k g|\,d\hat{\sigma} = 0 \qquad textsince\ F^{E_0} \subset \partial\Omega.$$

For every $\varepsilon > 0$, the last part is majorized by

$$\leq C\left(\frac{1}{\varepsilon^2}\int_{F^{D_0}} |\nabla_T^k g|^2 d\hat{\sigma} + \varepsilon^2 \int_{F^{D_0}} \delta^2 |\nabla\nabla_T^k g|^2 d\hat{\sigma}\right).$$

As $\mathcal{M}_\alpha(\nabla_T^k g) \leq \lambda$ on E, we deduce that

$$\frac{1}{\varepsilon^2}\int_{F^{D_0}} |\nabla_T^k g|^2 d\hat{\sigma} \leq \frac{1}{\varepsilon^2}\lambda^2 \int_{F^{D_0}} d\hat{\sigma} \leq \frac{C}{\varepsilon^2}\lambda^2\sigma(D_0).$$

We are going to prove now that, under the assumptions of Theorem D

$$\varepsilon^2 \int_{F^{D_0}} \delta^2 |\nabla\nabla_T^k g|^2 d\hat{\sigma} \leq C\left(\varepsilon^2 \iint_{\mathcal{R}_\beta} \delta^{k+2} |\nabla^{k+1} g|^2 \frac{dV}{\delta} + \int_K |g|^2 dV\right).$$

This will follow from the fact that, by Corollary 1.5,

$$|\nabla\nabla_T^k g|^2(\zeta) \leq C\,\mathrm{Mean}^{Q(\zeta)}\left(|\nabla\nabla_T^k g|^2 + \left[\delta^{-k/2}\mathrm{Rest}^k(\nabla g)\right]^2\right),$$

and the fact that

$$\int_{\partial\mathcal{R}_\alpha} \delta^{l+1}\mathrm{Mean}^Q(|f|^2)\,d\hat{\sigma} \leq C \iint_{\mathcal{R}_\beta} \delta^l |f|^2 dV,$$

for some β sufficiently large. Then, we apply Lemma 2.6 in order to obtain the result. So

$$(2) \leq \frac{C}{\varepsilon^2}\lambda^2\sigma(D_0) + C\left(\varepsilon^2 \iint_{\mathcal{R}_\beta} \delta^{k+2}|\nabla^{k+1}g|^2\frac{dV}{\delta} + \int_K |g|^2 dV\right).$$

1.3. *Estimate of the third term.* The third term is majorized by

$$\begin{aligned}(3) &\leq C\int_{\partial\mathcal{R}_\alpha} |\nabla_T^k g|^2 d\hat{\sigma} \leq C\left(\int_F + \int_{F^{E_0}} + \int_{F^{D_0}}\right)\\ &\leq C\left(\int_K |g|^2 dV + \int_0^\lambda t\,\sigma\left(\{\mathcal{M}_\alpha(\nabla_T^k g) \geq t\}\right) dt + \lambda^2\,\sigma(D_0)\right),\end{aligned}$$

since $\mathcal{M}_\alpha(\nabla_T^k g) \leq \lambda$ on E.

To conclude for Lemma 2.9, it suffices to remark that, as in the proof of Theorem 2.3,

$$\iint_{\mathcal{R}_\beta} \delta^{k+2}\left|\nabla^{k+1}g\right|^2 \frac{dV}{\delta} \leq \int_E S_\gamma^{-k}(\delta^{k+1}\nabla^{k+1}g)^2 d\sigma,$$

for some $\gamma > \beta$.

1.4. Let us prove now the a priori inequality in case $0 < p < 2$.

$$\begin{aligned}&\left\|S_\alpha(\delta\nabla\nabla_T^k g)\right\|_{L^p(\partial\Omega)}^p\\ &= p\int_0^\infty \lambda^{p-1}\sigma\left(\{S_\alpha(\delta\nabla\nabla_T^k g) \geq \lambda\}\right) d\lambda\\ &\leq p\int_0^\infty \lambda^{p-1}\sigma(D_0)d\lambda + M^p\sigma(\partial\Omega)\\ &\quad + p\int_M^\infty \lambda^{p-3}\int_{E_0} S_\alpha(\delta\nabla\nabla_T^k g)^2(\zeta)d\sigma(\zeta)\,d\lambda\\ &\leq C\Big(\Big(\frac{1}{\varepsilon^2}+1\Big)p\int_0^\infty \lambda^{p-1}\sigma(D)\,d\lambda + M^p\sigma(\partial\Omega)\\ &\quad + (\varepsilon^2 + s_0)p\int_M^\infty \lambda^{p-3}\int_E S_\gamma^{-k}(\delta^{k+1}\nabla^{k+1}g)^2(\zeta)\,d\sigma(\zeta)d\lambda\\ &\quad + p\int_M^\infty \lambda^{p-3}\int_0^\lambda t\,\sigma\left(\{\mathcal{M}_\alpha(\nabla_T^k g) \geq t\}\right) dt\,d\lambda\end{aligned}$$

$$
\begin{aligned}
&\quad + \|g\|^2_{L^2(K)} M^{p-2}\Big) \\
&\leq C\Big(\Big(\frac{1}{\varepsilon^2}+1\Big)\Big(\big\|\mathcal{M}_\alpha(\nabla_T^k g)\big\|^p_{L^p(\partial\Omega)} \\
&\quad + \frac{1}{C(\varepsilon,s_0)^p}\big\|S_\gamma^{-k}(\delta^{k+1}\nabla^{k+1}g)\big\|^p_{L^p(\partial\Omega)}\Big) \\
&\quad + M^p\sigma(\partial\Omega) \\
&\quad + \|g\|^2_{L^2(K)} M^{p-2} \\
&\quad + \frac{(\varepsilon^2+s_0)}{C(\varepsilon,s_0)^{p-2}}\big\|S_\gamma^{-k}(\delta^{k+1}\nabla^{k+1}g)\big\|_{L^p(\partial\Omega)}\Big),
\end{aligned}
$$

since

$$
\sigma(D) \leq \sigma\left(\{\mathcal{M}_\alpha(\nabla_T^k g) \geq \lambda\}\right) + \sigma\left(\{S_\gamma^{-k}(\delta^{k+1}\nabla^{k+1}g) \geq C(\varepsilon,s_0)\lambda\}\right).
$$

By Theorem C and under the assumptions of Theorem D, we have

$$
\big\|S_\gamma^{-k}(\delta^{k+1}\nabla^{k+1}g)\big\|_{L^p(\partial\Omega)} \leq C\left(\big\|S_\gamma(\delta\nabla\nabla_T^k g)\big\|_{L^p(\partial\Omega)} + \|g\|_{L^2(K)}\right).
$$

So, it is sufficient to apply Lemma 2.1 and to choose ε and s_0 sufficiently small and $M = \|g\|_{L^2(K)}$ in order to conclude that, for $g \in C^\infty(\overline{\Omega})$ holomorphic in Ω , we have

$$
\big\|S_\alpha(\delta\nabla\nabla_T^k g)\big\|_{L^p(\partial\Omega)} \leq C\left(\big\|\mathcal{M}_\alpha(\nabla_T^k g)\big\|_{L^p(\partial\Omega)} + \|g\|_{L^2(K)}\right).
$$

2. *Case* $2 < p < \infty$. In this part, we will use an auxiliary result on Dirichlet's problem. This will be proved in the appendix since we did not find any reference. We need the following definitions.

- We denote by $L_\alpha^{(p,2)}(\Omega)$ the space of all functions u on Ω such that

$$
u \circ \Phi \in L^p\left(\partial\Omega; L^2(]0,s_0[, t^\alpha dt)\right)
$$

with the induced Banach norm on $L_\alpha^{(p,2)}(\Omega)$

$$
\|u\|^p_{L_\alpha^{(p,2)}(\Omega)} = \int_{\partial\Omega}\left(\int_0^{s_0} |u\circ\Phi(\zeta,t)|^2\, t^\alpha dt\right)^{p/2} d\sigma(\zeta).
$$

- We denote by $W_\alpha^{l;(p,2)}(\Omega)$, $l \in \mathbb{N}$, the space of all functions u such

that

$$D^j u \in L^{(p,2)}_\alpha(\Omega) \qquad \text{for } |j| \leq l\,,$$

where $D^j u$ denotes the distribution derivative. We define a Banach norm on $W^{l;(p,2)}_\alpha(\Omega)$ by

$$\|u\|^p_{W^{l;(p,2)}_\alpha} = \Big(\sum_{|j|\leq l} \|D^j u\|^p_{L^{(p,2)}_\alpha} \Big)^{1/p}.$$

- For $l \in \mathbb{N}$, we denote by $\overset{\circ}{W}{}^{l;(p,2)}_\alpha(\Omega)$ the closure of $\mathcal{D}(\Omega)$ in $W^{l;(p,2)}_\alpha(\Omega)$.
- For $l \in \mathbb{N}$, we denote by $W^{-l;(p,2)}_\alpha(\Omega)$ the dual space of $\overset{\circ}{W}{}^{l;(p',2)}_{-\alpha}(\Omega)$, $1/p + 1/p' = 1$.

Now, we can state our result.

Theorem on Dirichlet's Problem. *Let $1 < p < \infty$, Ω be a bounded C^∞-domain and A be a differential operator of order 2, strongly elliptic, with smooth coefficients. For every smooth function v defined on Ω, let u be the solution of the problem*

$$\begin{cases} Au = v & \text{in } \Omega\,, \\ u = 0 & \text{on } \partial\Omega\,. \end{cases}$$

Then, for every $-1 < \theta < 1$, there exists a constant C independent of v such that

$$\int_{\partial\Omega} \left(\int_0^{s_0} |\nabla u \circ \Phi(z,t)|^2 t^\theta dt \right)^{p/2} d\sigma(z) \leq C \left(\|v\|_{W^{-1;(p,2)}_\theta(\Omega)} + \sup_K |u|^p \right).$$

To prove our estimate, we have to majorize, for $g \in C^\infty(\overline{\Omega})$,

$$\left\| S_\alpha \left(\delta \nabla \nabla^k_T g \right) \right\|_{L^p(\partial\Omega)} \quad \text{by} \quad \left\| \sup_{0<t<s_0} |\nabla^k_T g| \right\|_{L^p(\partial\Omega)},$$

then the a priori estimate will follow.

First, we are going to show that

$$\left\|G\left(\delta\nabla\nabla_T^k g\right)\right\|_{L^p(\partial\Omega)} \leq C\left(\left\|\sup_{0<t<s_0}\left|\nabla_T^k g\right|\right\|_{L^p(\partial\Omega)} + s_0^{1/2}\left\|S_\alpha\left(\delta\nabla\nabla_T^k g\right)\right\|_{L^p(\partial\Omega)}\right).$$

The desired estimate will follow since

$$\left\|S_\alpha\left(\delta\nabla\nabla_T^k g\right)\right\|_{L^p(\partial\Omega)} \leq C\left\|G\left(\delta\nabla\nabla_T^k g\right)\right\|_{L^p(\partial\Omega)}.$$

We write $\nabla_T^k g = \left(\nabla_T^k g\right)_0 + \left(\nabla_T^k g\right)_h$ where $\left(\nabla_T^k g\right)_0$ is the solution of the problem

$$\begin{cases} \Delta w = \Delta(\nabla_T^k g) & \text{in } \Omega, \\ w = 0 & \text{on } \partial\Omega. \end{cases}$$

We know that such a solution exists in $C^\infty(\overline{\Omega})$ (since $g \in C^\infty(\overline{\Omega})$ by assumption).

So,

$$\left\|G\left(\delta\nabla\nabla_T^k g\right)\right\|_{L^p(\partial\Omega)} \leq \left\|G\left(\delta\nabla(\nabla_T^k g)_0\right)\right\|_{L^p(\partial\Omega)} + \left\|G\left(\delta\nabla(\nabla_T^k g)_h\right)\right\|_{L^p(\partial\Omega)}.$$

2.1. *Estimate of* $\left\|G(\delta\nabla(\nabla_T^k g)_h)\right\|_{L^p(\partial\Omega)}$. It is well known that the Littlewood-Paley function of a harmonic function is majorized, in L^p-norm, by the L^p-norm of its trace on the boundary. So, we have

$$\left\|G(\delta\nabla(\nabla_T^k g)_h)\right\|_{L^p(\partial\Omega)} \leq \left\|\left|(\nabla_T^k g)_h\right|\right\|_{L^p(\partial\Omega)} \leq \left\|\sup_{0<t<s_0}\left|\nabla_T^k g\right|\right\|_{L^p(\partial\Omega)}$$

since, by assumption, $\nabla_T^k g = (\nabla_T^k g)_h$ on $\partial\Omega$ and $g \in C^\infty(\overline{\Omega})$.

2.2. *Estimate of* $\left\|G(\delta\nabla(\nabla_T^k g)_0)\right\|_{L^p(\partial\Omega)}$. As g is holomorphic in Ω,

$$\Delta(\nabla_T^k g) = [\Delta, \nabla_T^k] g, \quad \left|\Delta(\nabla_T^k g)\right| \leq C\left|\nabla v\right|$$

with

$$|v| \leq C \sum_{\substack{0\leq j\leq 1 \\ 0\leq r\leq k-1}} \left|\nabla^j \nabla_T^r g\right|$$

and we can apply the Theorem on the Dirichlet's problem. For every $0 < \theta < 1$

$$\begin{aligned}
\left\| G\left(\delta\nabla(\nabla_T^k g)_0\right)\right\|_{L^p(\partial\Omega)} &\leq s_0^{(1-\theta)/2}\left\| G^{\theta-1}\left(\delta\nabla(\nabla_T^k g)_0\right)\right\|_{L^p(\partial\Omega)} \\
&= s_0^{(1-\theta)/2}\Big(\int_{\partial\Omega}\Big(\int_0^{s_0}\left|\nabla(\nabla_T^k g)_0\right|^2 t^\theta dt\Big)^{p/2} d\sigma\Big)^{1/p} \\
&\leq C\, s_0^{(1-\theta)/2}\Big(\int_{\partial\Omega}\Big(\int_0^{s_0}\sum_{\substack{0\leq j\leq 1\\ 0\leq r\leq k-1}}|\nabla^j\nabla_T^r g|^2 t^\theta dt\Big)^{p/2} d\sigma\Big)^{1/p} \\
&\quad + C\,\|g\|_{L^2(K)} \\
&\leq C\, s_0^{(1-\theta)/2}\Big(\int_{\partial\Omega}\Big(\int_0^{s_0}\sum_{0\leq r\leq k-1}|\nabla^k\nabla_T^r g|^2 t^{\theta+2k-2} dt\Big)^{p/2} d\sigma\Big)^{1/p} \\
&\quad + C\,\|g\|_{L^2(K)},
\end{aligned}$$

(by Hardy inequality in $L^2(]0, s_0[)$),

$$\begin{aligned}
&\leq C\, s_0^{(1-\theta)/2}\Big(\int_{\partial\Omega}\Big(\int_0^{s_0}\mathrm{Mean}^Q\Big(\sum_{j=1}^k|\nabla^j g|^2\Big) t^{\theta+k-1} dt\Big)^{p/2} d\sigma\Big)^{1/p} \\
&\quad + C\,\|g\|_{L^2(K)}\,,
\end{aligned}$$

(by Theorem A),

$$\begin{aligned}
&= C\, s_0^{(1-\theta)/2}\Big(\Big\|S_\alpha^{\theta+k}\Big(\sum_{j=1}^k|\nabla^j g|\Big)\Big\|_{L^p(\partial\Omega)} + \|g\|_{L^2(K)}\Big) \\
&\leq C\, s_0^{1/2}\left(\left\|S_\alpha(\delta\nabla\nabla_T^k g)\right\|_{L^p(\partial\Omega)} + \|g\|_{L^2(K)}\right),
\end{aligned}$$

(by Theorem 2.3).

2.5. Admissible and radial maximal functions.

We are going to show the following theorem.

Theorem E. *Let Ω be a bounded C^∞-domain in $\mathbb{C}^n$, satisfying* (P). *For every $0 < p < \infty$, every $k \in \mathbb{N}$, every $\alpha > 0$, there exists a constant C such that, for every holomorphic function g, the following holds*

$$\left\|\mathcal{M}_\alpha(\nabla_T^k g)\right\|_{L^p(\partial\Omega)} \leq C\left(\left\|\sup_{0<t<s_0}\left|\nabla_T^k g\right|\right\|_{L^p(\partial\Omega)} + \|g\|_{L^2(K)}\right).$$

As an application, we obtain that (4) implies (3) in Main Theorem.

Lemma 2.10. *Let $l > 0$, α be a fixed aperture. There exists a constant C such that, for every function $u \in C^\infty(\Omega)$, we have, for every $\mu > 0$, every $\zeta \in \partial\Omega$*

$$\mathcal{M}_\alpha\left(\delta^\mu |u|\right)(\zeta) \leq C\mathcal{M}_\alpha\left(\delta\left|\frac{\partial u}{\partial \nu}\right|\right)(\zeta) + C\sup_K |u|,$$
$$\mathcal{M}_\alpha\left(\delta^l |u|\right)(\zeta) \leq C\mathcal{M}_\alpha\left(\delta^{l+1}\left|\frac{\partial u}{\partial \nu}\right|\right)(\zeta) + C\sup_K |u|,$$

where $\partial/\partial\nu$ denotes the normal derivative on $\partial\Omega$.

PROOF. Let $\zeta \in \partial\Omega$. We assume that

$$\mathcal{M}_\alpha\left(\delta^{l+1}\left|\frac{\partial u}{\partial \nu}\right|\right)(\zeta) \leq C.$$

For every $\Phi(\eta, t) \in \mathcal{A}_\alpha(\zeta)$, we have

$$\begin{aligned}|u \circ \Phi(\eta, t)| &= \left|\int_0^{s_0} \frac{d}{ds} u \circ \Phi(\eta, s+t) ds - u \circ \Phi(\eta, s_0 + t)\right| \\ &\leq \mathcal{M}_\alpha\left(\delta^{l+1}\left|\frac{\partial u}{\partial \nu}\right|\right)(\zeta) \cdot \left|\int_0^{s_0} \frac{ds}{(t+s)^{l+1}}\right| + C\sup_K |u|\end{aligned}$$

(since $\Phi(\eta, s+t) \in \mathcal{A}_\alpha(\zeta) \cup K$)

$$\leq C\left(t^{-l}\mathcal{M}_\alpha\left(\delta^{l+1}\left|\frac{\partial u}{\partial \nu}\right|\right)(\zeta) + \sup_K |u|\right).$$

PROOF OF THEOREM E. As in the proof of Theorem D, we are going to prove an a priori estimate. Explicitely, for $g \in C^\infty(\overline{\Omega})$, we are going to show that

$$\left\|\mathcal{M}_\alpha(|\nabla_T^k g|)\right\|_{L^p(\partial\Omega)} \leq C\left(\left\|\sup_{0<t<s_0} |\nabla_T^k g|\right\|_{L^p(\partial\Omega)} + \|g\|_{L^p(K)}\right).$$

For general g, we apply the preceding inequality in Ω_ε and we let $\varepsilon \to 0$ as before.

So, let us assume that $g \in C^\infty(\overline{\Omega})$. By Lemma 1.2, after a change of coordinates, we can write, around any point of $\mathcal{A}_\beta(\zeta)$, each component of $\nabla_T^k g$ as a sum of an (AB) function and of a rest $\delta^{-k/2}\mathrm{Rest}^k(g)$. This allows us to show that, for every $\Phi(\eta,t) \in \mathcal{A}_\alpha(\zeta)$

$$
\begin{aligned}
&\left|\nabla_T^k g \circ \Phi(\eta,t)\right|^{p/2} \\
&\leq \frac{C}{|Q(\Phi(\eta,t))|}\int_{Q(\Phi(\eta,t))} \left|\nabla_T^k g\right|^{p/2}\, dV \\
&\quad + \frac{C}{|Q(\Phi(\eta,t))|}\int_{Q(\Phi(\eta,t))} \left|\delta^{-k/2}\mathrm{Rest}^k(g)\right|^{p/2}\, dV \\
&\quad + \left|\delta^{-k/2}\mathrm{Rest}^k(g) \circ \Phi(\eta,t)\right|^{p/2} \\
&\leq \frac{C}{|B^d(\eta,t)|}\int_{B^d(\eta,t)} \sup_{0<s<s_0} \left|\nabla_T^k g \circ \Phi(\eta',s)\right|^{p/2} d\sigma(\eta') \\
&\quad + \frac{C}{|B^d(\eta,t)|}\int_{B^d(\eta,t)} \sup_{0<s<s_0} \left|\delta^{-k/2}\mathrm{Rest}^k(g) \circ \Phi(\eta',s)\right|^{p/2} d\sigma(\eta') \\
&\quad + \left|\delta^{-k/2}\mathrm{Rest}^k(g) \circ \Phi(\eta,t)\right|^{p/2}.
\end{aligned}
$$

So, we obtain, by the L^2-continuity of the non-isotropic maximal operator

$$
\begin{aligned}
&\left\|\mathcal{M}_\alpha(\nabla_T^k g)\right\|_{L^p(\partial\Omega)}^p = \left\|\mathcal{M}_\alpha(\nabla_T^k g)^{p/2}\right\|_{L^2(\partial\Omega)}^p \\
&\leq C\Big(\left\| \sup_{0<s<s_0} |\nabla_T^k g| \right\|_{L^p(\partial\Omega)}^p + \left\| \sup_{0<s<s_0} |\delta^{-k/2}\mathrm{Rest}^k(g)| \right\|_{L^p(\partial\Omega)}^p \\
&\qquad + \left\|\mathcal{M}_\alpha(\delta^{-k/2}\mathrm{Rest}^k(g))\right\|_{L^p(\partial\Omega)}^p \Big) \\
&\leq C\left(\left\| \sup_{0<s<s_0} |\nabla_T^k g| \right\|_{L^p(\partial\Omega)}^p + \left\|\mathcal{M}_\alpha(\delta^{-k/2}\mathrm{Rest}^k(g))\right\|_{L^p(\partial\Omega)}^p \right).
\end{aligned}
$$

So, it suffices to estimate $\left\|\mathcal{M}_\alpha(\delta^{-k/2}\mathrm{Rest}^k(g))\right\|_{L^p(\partial\Omega)}^p$ in terms of $\left\|\mathcal{M}_\alpha(\nabla_T^k g)\right\|_{L^p(\partial\Omega)}$ with a small constant.

$$
\begin{aligned}
\mathcal{M}_\alpha\left(\delta^{-k/2}\mathrm{Rest}^k(g)\right) \leq C\Big(&\mathcal{M}_\alpha\Big(\sum_{1\leq j+r<(k+r)/2} |\nabla^j \nabla_T^r u| \Big) \\
&+ \mathcal{M}_\alpha\Big(\sum_{(k+r)/2\leq j+r\leq k} \delta^{j+(r+1-k)/2} |\nabla^j \nabla_T^r u| \Big)\Big).
\end{aligned}
$$

Let β be chosen so that if $z \in \mathcal{A}_\alpha(\zeta)$, $Q(z) \subset \mathcal{A}_\beta(\zeta)$. We apply successively Lemma 2.10 and Theorem A in order to obtain, for every $\mu > 0$,

$$\begin{aligned}\mathcal{M}_\alpha\left(\delta^{-k/2}\mathrm{Rest}^k(g)\right) &\leq C\left(\mathcal{M}_\beta\Big(\sum_{1\leq j+r<(k+r)/2} \delta^{k-\mu-j-r/2}|\nabla^k g|\Big)\right.\\ &\qquad \left. + \mathcal{M}_\beta\Big(\sum_{(k+r)/2\leq j+r\leq k} \delta^{(k+1)/2}|\nabla^k g|\Big)\right)\\ &\leq C\,\mathcal{M}_\beta\left(\delta^{(k+1)/2-\mu}|\nabla^k g|\right).\end{aligned}$$

So, we have to estimate

$$\|\mathcal{M}_\beta\left(\delta^{(k+1)/2-\mu}|\nabla^k g|\right)\|_{L^p(\partial\Omega)}$$

in terms of

$$\left\|\mathcal{M}_\alpha(\nabla_T^k g)\right\|_{L^p(\partial\Omega)}$$

with a small constant.

By the converse estimates of Theorem B, we have, by choosing $\mu < 1/2$,

$$\begin{aligned}(*) &= \mathcal{M}_\beta\left(\delta^{(k+1)/2-\mu}|\nabla^k g|\right)\\ &\leq C\left(\mathcal{M}_\gamma\left(\delta^{1/2-\mu}|\nabla_T^k g|\right) + \mathcal{M}_\gamma\left(\delta^{-\mu+(1-k)/2}|\mathrm{Rest}^k(g)|\right) + \sup_K |g|\right)\\ &\leq C\left(s_0^{1/2-\mu}\mathcal{M}_\gamma\left(|\nabla_T^k g|\right) + s_0^{1/2}\mathcal{M}_\gamma\left(\delta^{(k+1)/2-\mu}|\nabla^k g|\right) + \sup_K |g|\right).\end{aligned}$$

Then, we choose s_0 sufficiently small in order to obtain

$$\begin{aligned}&\left\|\mathcal{M}_\beta\left(\delta^{(k+1)/2-\mu}|\nabla^k g|\right)\right\|_{L^p(\partial\Omega)}\\ &\qquad \leq C\left(s_0^{1/2-\mu}\left\|\mathcal{M}_\alpha(|\nabla_T^k g|)\right\|_{L^p(\partial\Omega)} + \|g\|_{L^p(K)}\right).\end{aligned}$$

Inserting this inequality in the estimate of

$$\left\|\mathcal{M}_\alpha(\delta^{-k/2}\mathrm{Rest}^k(g))\right\|_{L^p(\partial\Omega)},$$

we are able to conclude, reducing again s_0 if necessary, that, for $g \in C^\infty(\overline{\Omega})$

$$\left\|\mathcal{M}_\alpha(|\nabla_T^k g|)\right\|_{L^p(\partial\Omega)} \leq C\left(\left\|\sup_{0<t<s_0}|\nabla_T^k g|\right\|_{L^p(\partial\Omega)} + \|g\|_{L^p(K)}\right).$$

2.6. End of the proof of Main Theorem.

In this paragraph, we are going to show the following theorem.

Theorem F. *Let Ω be a bounded C^∞-domain in $\mathbb{C}^n$, satisfying* (P). *For every $1-1/(2n+1) < p < \infty$, every $k \in \mathbb{N}$, there exists a constant C such that, for every holomorphic function g in Ω, we have*

$$\left\| \sup_{0<t<s_0} |\nabla_T^k g| \right\|_{L^p(\partial\Omega)} \leq C \left\| S_\alpha(\delta \nabla \nabla_T^k g) \right\|_{L^p(\partial\Omega)} + \|g\|_{L^2(K)} \,.$$

As an immediate corollary of this theorem, we obtain that (2) implies (4) in Main Theorem.

PROOF OF THEOREM F. In this part, we will use theorem on Dirichlet's problem stated in paragraph 2.4.

We want to majorize

$$\left\| \sup_{0<t<s_0} |\nabla_T^k g| \right\|_{L^p(\partial\Omega)}$$

As before, we are going to prove Theorem F only for $g \in C^\infty(\overline{\Omega})$. We write $\nabla_T^k g = \left(\nabla_T^k g\right)_0 + \left(\nabla_T^k g\right)_h$ where $\left(\nabla_T^k g\right)_0$ is the solution of the problem

$$\begin{cases} \Delta w = \Delta(\nabla_T^k g) & \text{in } \Omega\,, \\ w = 0 & \text{on } \partial\Omega\,. \end{cases}$$

So,

$$\begin{aligned} \left\| \sup_{0<t<s_0} |\nabla_T^k g| \right\|_{L^p(\partial\Omega)} &\leq \left\| \sup_{0<t<s_0} |(\nabla_T^k g)_0| \right\|_{L^p(\partial\Omega)} \\ &\quad + \left\| \sup_{0<t<s_0} |(\nabla_T^k g)_h| \right\|_{L^p(\partial\Omega)} \\ &= (1) + (2)\,. \end{aligned}$$

First, for every $0 < \theta < 1$, we have

$$\left|(\nabla_T^k g)_0 \circ \Phi(\eta, t)\right| \leq \int_0^{s_0} \left|\nabla(\nabla_T^k g)_0 \circ \Phi(\eta, s)\right| ds + \sup_K |g|$$

$$\leq C\, s_0^{(1-\theta)/2} \left(\int_0^{s_0} \left| \nabla(\nabla_T^k g)_0 \circ \Phi(\eta, s) \right|^2 s^{\theta+1} \frac{ds}{s} \right)^{1/2} + \sup_K |g|$$

$$\leq C\, s_0^{(1-\theta)/2}\, G^{\theta-1}(\delta \nabla(\nabla_T^k g)_0)(\eta) + \sup_K |g| .$$

So,

$$(1) = \left\| \sup_{0<t<s_0} |(\nabla_T^k g)_0| \right\|_{L^p(\partial\Omega)} \leq C \left(s_0^{(1-\theta)/2} \left\| G^{\theta-1}(\delta \nabla(\nabla_T^k g)_0) \right\|_{L^p(\partial\Omega)} + \|g\|_{L^p(K)} \right).$$

It is well known that, for harmonic functions, the L^p-norm of the radial maximal functions are majorized by the L^p-norm of the Littlewood-Paley functions. So,

$$(2) = \left\| \sup_{0<t<s_0} |(\nabla_T^k g)_h| \right\|_{L^p(\partial\Omega)} \leq C \left\| G(\delta \nabla(\nabla_T^k g)_h) \right\|_{L^p(\partial\Omega)}$$

and we obtain

$$(2) \leq C \left(\left\| G(\delta \nabla \nabla_T^k g) \right\|_{L^p(\partial\Omega)} + s_0^{(1-\theta)/2} \left\| G^{\theta-1}(\delta \nabla(\nabla_T^k g)_0) \right\|_{L^p(\partial\Omega)} \right)$$

$$\leq C \left(\left\| S_\alpha(\delta \nabla \nabla_T^k g) \right\|_{L^p(\partial\Omega)} + s_0^{(1-\theta)/2} \left\| G^{\theta-1}(\delta \nabla(\nabla_T^k g)_0) \right\|_{L^p(\partial\Omega)} \right),$$

by Corollary 2.8.

So, it remains to estimate $\left\| G^{\theta-1}(\delta \nabla(\nabla_T^k g)_0) \right\|_{L^p(\partial\Omega)}$. Since in the proof of Theorem D, we have shown that, when $2 < p < \infty$,

$$s_0^{(1-\theta)/2} \left\| G^{\theta-1}(\delta \nabla(\nabla_T^k g)_0) \right\|_{L^p(\partial\Omega)} \leq C \left(s_0^{1/2} \left\| S_\alpha(\delta \nabla \nabla_T^k g) \right\|_{L^p(\partial\Omega)} + \|g\|_{L^2(K)} \right),$$

it is sufficient to consider the case $0 < p \leq 2$. We are going to show that

$$\left\| G^{\theta-1}(\delta \nabla(\nabla_T^k g)_0) \right\|_{L^p(\partial\Omega)}$$

is bounded by

$$\left\| \sup_{0<t<s_0} |\nabla_T^k g| \right\|_{L^p(\partial\Omega)}$$

when $1 - 1/(2n+1) < p \leq 2$.

First, let us give the following lemma.

Lemma 2.11. *Let u be any continuous function defined on Ω. Then, for every $\alpha > 0$, every $q \geq p > 0$ and every $\theta > 0$ satisfying $\theta \geq n/p - n/q$, we have*

$$\left\|\mathcal{M}_\alpha(\delta^\theta u)\right\|_{L^q(\partial\Omega)} \leq C \left\|\mathcal{M}_\alpha(u)\right\|_{L^p(\partial\Omega)}.$$

PROOF. The proof is based on the atomic decomposition of spaces of homogeneous type (see [AN]). We will denote by T^p the space of all continuous functions u on Ω such that $\mathcal{M}_\alpha(u) \in L^p(\partial\Omega)$ with norm $\|u\|_{T^p} = \|\mathcal{M}_\alpha(u)\|_{L^p(\partial\Omega)}$. For every $E \subset \partial\Omega$, we define the tent over E to be the subset $\hat{E}$ of $\partial\Omega \times (0, s_0)$ by $\partial\Omega \times (0, s_0) \setminus \hat{E} = \cup \{\mathcal{A}(\zeta),\, \zeta \in \partial\Omega \setminus E\}$. A non-negative function a on $\partial\Omega \times (0, s_0)$ is an atom over $\hat{B^d}$ if a vanishes outside $\hat{B^d}$ and if $a \leq \sigma(B^d)^{-1}$. The atomic decomposition Theorem can be formulated as follows.

Theorem. ([AN]). *There is a constant C such that, for every $u \in T^1$, there are a sequence of balls $\{B_j^d = B^d(x_j, \delta_j)\}$, a sequence of atoms a_j over B_j^d and a sequence $\{\lambda_j\}$ of positive numbers such that $|u \circ \Phi| \leq \sum \lambda_j a_j$ on $\partial\Omega \times (0, s_0)$ and $\sum \lambda_j \leq \|u\|_{T^1}$.*

Let $u \in T^p$ with $\|u\|_{T^p} \leq 1$, then $|u|^p \in T^1$, so by the preceding theorem, there are a sequence $\{a_j\}$ of atoms and a sequence $\{\lambda_j\}$ of positive numbers such that $|u \circ \Phi|^p \leq \sum \lambda_j a_j$ on $\partial\Omega \times (0, s_0)$ and $\sum \lambda_j \leq \|u\|_{T^p}^p \leq 1$. Thus, we have

$$\left[\mathcal{M}_\alpha(\delta^\theta |u|)\right]^p \leq \mathcal{M}_\alpha(\delta^{\theta p} |u|^p) \leq \left(\sum \lambda_j \mathcal{M}_\alpha(\delta^{\theta p} |a_j|)\right).$$

So

$$\begin{aligned}\left\|\mathcal{M}_\alpha(\delta^\theta |u|)\right\|_{L^q(\partial\Omega)} &= \left\|\mathcal{M}_\alpha(\delta^\theta |u|)^p\right\|_{L^{q/p}(\partial\Omega)} \\ &\leq \sum \lambda_j \left\|\mathcal{M}_\alpha(\delta^{\theta p} |a_j|)\right\|_{L^{q/p}(\partial\Omega)}\end{aligned}$$

and it suffices to see that

$$\left\|\mathcal{M}_\alpha(\delta^{\theta p} |a_j|)\right\|_{L^{q/p}(\partial\Omega)} \leq C.$$

But

$$\left\|\mathcal{M}_\alpha(\delta^{\theta p}\,|a_j|)\right\|_{L^{q/p}(\partial\Omega)} \le \delta_j^{\theta p}\,\sigma(B_j^d)^{-1}\,\sigma(B_j^d)^{p/q}$$

since $\delta(z) \le c\,\delta_j$ on $\hat{B_j^d}$. This allows to conclude, since $\sigma(B_j^d) \simeq \delta_j^n$.

1. *Estimate of* $\left\|G^{\theta-1}(\delta\nabla(\nabla_T^k g)_0)\right\|_{L^p(\partial\Omega)}$ *when* $1-1/(2n+1) < p \le 2$. As g is holomorphic in Ω, $\Delta(\nabla_T^k g) = [\Delta, \nabla_T^k]g$, $\left|\Delta(\nabla_T^k g)\right| \le C\,|\nabla v|$ with

$$|v| \le C \sum_{\substack{0\le j\le 1\\ 0\le r\le k-1}} |\nabla^j \nabla_T^r g|$$

and we can apply the result of the theorem on Dirichlet's problem.

For every $0 < \theta < 1$

$$(*) = \left\|G^{\theta-1}(\delta\nabla(\nabla_T^k g)_0)\right\|_{L^p(\partial\Omega)} \le \left\|G^{\theta-1}(\delta\nabla(\nabla_T^k g)_0)\right\|_{L^q(\partial\Omega)}$$

where $q > 1$ will be chosen later. Then, we can apply our result on Dirichlet's problem to obtain

$$\begin{aligned}(*) \le\ & C\Big(\int_{\partial\Omega}\Big(\int_0^{s_0} \sum_{\substack{0\le j\le 1\\ 0\le r\le k-1}} |\nabla^j\nabla_T^r g|^2\, t^\theta\, dt\Big)^{q/2} d\sigma\Big)^{1/q}\\ & + C\,\|g\|_{L^2(K)}\\ \le\ & C\Big(\int_{\partial\Omega}\Big(\int_0^{s_0} \sum_{0\le r\le k-1} |\nabla^k\nabla_T^r g|^2\, t^{\theta+2k-2}\, dt\Big)^{q/2} d\sigma\Big)^{1/q}\\ & + C\,\|g\|_{L^2(K)}\ ,\end{aligned}$$

(by Hardy inequality in $L^2(]0, s_0[)$),

$$\begin{aligned}\le\ & C\Big(\int_{\partial\Omega}\Big(\int_0^{s_0} \mathrm{Mean}^Q\Big(\sum_{j=1}^k |\nabla^j g|^2\Big) t^{\theta+k-1} dt\Big)^{q/2} d\sigma\Big)^{1/q}\\ & + C\,\|g\|_{L^2(K)}\ ,\end{aligned}$$

(by Theorem A),

$$\le C\left(\left\|S_\alpha^{\theta-k}(\delta^k\nabla^k g)\right\|_{L^q(\partial\Omega)} + \|g\|_{L^2(K)}\right)\ ,$$

(by Theorem 2.3).

So, we obtain by Theorem C

$$\left\|S_\alpha^{\theta-k}(\delta^k\nabla^k g)\right\|_{L^q(\partial\Omega)} \leq C\left(\left\|S_\alpha^\theta(\nabla_T^k g)\right\|_{L^q(\partial\Omega)} + \|g\|_{L^2(K)}\right).$$

So we have, for every $0 < \varepsilon < 1$,

$$(*) \leq C\left(\int_{\partial\Omega}\left(\mathcal{M}_\alpha(\delta^\beta|\nabla_T^k g|)\right)^q\left(\int_{A_\alpha(\zeta)}\delta^{2\varepsilon}\frac{dV}{\delta^{n+1}}\right)^{q/2}d\sigma\right)^{1/q} + C\|g\|_{L^2(K)}$$

(with $\beta = \theta/2 - \varepsilon$),

$$\leq C\left(s_0^\varepsilon\left\|\mathcal{M}_\alpha(\delta^\beta|\nabla_T^k g|)\right\|_{L^q(\partial\Omega)} + \|g\|_{L^2(K)}\right)$$
$$\leq C\left(s_0^\varepsilon\left\|\mathcal{M}_\alpha(\nabla_T^k g)\right\|_{L^p(\partial\Omega)} + \|g\|_{L^2(K)}\right),$$

by choosing q such that Lemma 2.11 holds with $\theta/2 - \varepsilon \geq n/p - n/q$ which is possible if $1 - 1/(2n+1) < p < \infty$ (since we can choose $\theta - 1$ and ε arbitrarily close to 0). But, by Theorem E, we have

$$\left\|\mathcal{M}_\alpha\left(\nabla_T^k g\right)\right\|_{L^p(\partial\Omega)} \leq C\left(\left\|\sup_{0<t<s_0}|\nabla_T^k g|\right\|_{L^p(\partial\Omega)} + \|g\|_{L^2(K)}\right).$$

So, we obtain an a priori estimate for every $1 - 1/(2n+1) < p < \infty$ and every holomorphic function $g \in C^\infty(\overline{\Omega})$

$$\left\|\sup_{0<t<s_0}|\nabla_T^k g|\right\|_{L^p(\partial\Omega)} \leq C\left(\left\|S_\alpha(\delta\nabla\nabla_T^k g)\right\|_{L^p(\partial\Omega)} + \|g\|_{L^2(K)}\right).$$

3. Appendix: Dirichlet problem in mixed Sobolev Spaces with weights.

Let us recall the result we are going to prove.

Theorem on Dirichlet's Problem. *Let $1 < p < \infty$, Ω be a bounded C^∞-domain and A be a differential operator of order 2, strongly elliptic,*

with smooth coefficients. For every smooth function v defined on Ω, let u be the solution of the problem

$$\begin{cases} Au = v & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega. \end{cases}$$

Then, for every $-1 < \theta < 1$, there exists a constant C independent of v such that

$$\int_{\partial\Omega} \left(\int_0^{s_0} |\nabla u \circ \Phi(z,t)|^2 t^\theta dt \right)^{p/2} d\sigma(z) \leq C \left(\|v\|_{W_\theta^{-1;(p,2)}(\Omega)} + \sup_K |u|^p \right).$$

The proof follows the same line as the one given by Grisvard in his book for the usual problem (see [Gr]). By routine arguments (partition of unity, change of coordinates, freezing of coefficients) it suffices to solve the problem for the Laplacian and for smooth functions with compact support in

$$F = \{(x', x_n): \quad x' \in \mathbb{R}^{n-1},\ 0 \leq x_n \leq s_0\} \subset \mathbb{R}^n_+, \quad s_0 \text{ fixed.}$$

Explicitely, we just have to show the following lemma.

Lemma 3.1. *Let $1 < p < \infty$. Then, for every $-1 < \theta < 1$, there exists a constant C such that, for any smooth function u with compact support in F, we have*

$$\int_{\mathbb{R}^{n-1}} \left(\int_0^{s_0} |\nabla u|^2 x_n^\theta dx_n \right)^{p/2} dx' \leq C \left(\|\Delta u\|_{W_\theta^{-1;(p,2)}(F)} + \sup_K |u|^p \right).$$

PROOF. Let u be any smooth function with compact support in F. We denote by v the function Δu. Assume that v belongs to $W_\theta^{-1,(p,2)}(\mathbb{R}^n_+)$; then there exists v_1, $v_2^{(J)} \in L_\theta^{(p,2)}(\mathbb{R}^n_+)$, with compact support in F, such that

$$v = v_1 + \sum_{|J|=1} D^J v_2^{(J)}$$

$$\text{with } \|v\|_{W_\theta^{-1,(p,2)}(\mathbb{R}^n_+)} \simeq \|v_1\|_{L_\theta^{(p,2)}(\mathbb{R}^n_+)} + \sum_{|J|=1} \|v_2^{(J)}\|_{L_\theta^{(p,2)}(\mathbb{R}^n_+)}.$$

Then, u can be written as the sum of two functions u_1, u_2 satisfying

$$\begin{cases} \Delta u_1 = v_1 & \text{in } \mathbb{R}^n_+ , \\ u_1 = 0 & \text{on } \{x_n = 0\} , \end{cases} \qquad \begin{cases} \Delta u_2 = \sum_{|J|=1} D^J v_2^{(J)} & \text{in } \mathbb{R}^n_+ , \\ u_2 = 0 & \text{on } \{x_n = 0\} . \end{cases}$$

We will only estimate the term corresponding to u_2 since the other term is better.

An argument of symetry allows us to write

$$u_2(x', x_n) = \int_{\mathbb{R}^n_+} [E(x' - y', x_n - y_n) - E(x' - y', x_n + y_n)] \cdot \sum_{|J|=1} D^J v_2^{(J)}(y', y_n) dy' dy_n ,$$

where E is the normalized fundamental solution of Laplace's Equation. We can assume that u_2 is smooth with compact support in F.

By Green's Theorem, we have

$$u_2(x', x_n) = \int_{\mathbb{R}^n_+} \sum_{|J|=1} D^J [E(x' - y', x_n - y_n) - E(x' - y', x_n + y_n)] \cdot v_2^{(J)}(y', y_n) dy' dy_n ,$$

so, for $|K| = 1$

$$\begin{aligned} D^K u_2(x', x_n) &= \int_{\mathbb{R}^n_+} \sum_{|J|=1} D^{K+J} \big[E(x' - y', x_n - y_n) \\ &\qquad - E(x' - y', x_n + y_n)\big] v_2^{(J)}(y', y_n) dy' dy_n \\ &= \sum_{|\mu|=2} (D^\mu E * V_2^{(\mu)} + D^\mu E * V'^{(\mu)}_2) , \end{aligned}$$

where $V_2^{(\mu)}$ and $V'^{(\mu)}_2$ are zero outside $\mathbb{R}^n_+$. It is usual to see that $K = \sum_{|\mu|=2} D^\mu E$ is a Calderón-Zygmund kernel. We are going to show that the corresponding operator is bounded from $L^p(dx', L^2(x_n^\theta dx_n))$ into itself for every $-1 < \theta < 1$. Assume it is done, then we obtain

$$\int_{\mathbb{R}^{n-1}} \Big(\int_0^{s_0} |\nabla u_2|^2 x_n^\theta dx_n \Big)^{p/2} dx'$$

$$\leq C \int_{\mathbb{R}^{n-1}} \Big(\int_0^{s_0} \Big| \sum_{|J|=1} v_2^{(J)} \Big|^2 x_n^\theta dx_n \Big)^{p/2} dx'$$

$$\leq C \, \|v\|_{W_\theta^{-1,(p,2)}}$$

and this gives the result.

So, it remains to show that a Calderón-Zygmund kernel K defines a bounded operator from $L^p(dx', L^2(x_n^\theta dx_n))$ into itself, for every $-1 < \theta < 1$ and every $1 < p \leq 2$: the result for general p follows from duality. Let us give the proof for completeness.

First, it is easy to see that the weight x_n^θ belongs to the class of Muckenhoupt (A_2), for every $-1 < \theta < 1$. This implies the continuity of the operator from $L^2(x_n^\theta dx_n dx')$ into itself. It remains to show (see [S2]) that

$$\|\nabla_{x'} K\| \leq \frac{C}{|x'|^n}$$

where $K(x')$ is the operator defined by

$$K(x')h(t) = \int K(x', t-s)h(s)ds = \int \frac{\omega(x', t-s)}{\big((t-s)^2 + |x'|^2\big)^{n/2}} h(s)ds$$

for $h \in L^2(x_n^\theta dx_n)$, where ω is homogeneous of order 0 and $\|\nabla_{x'} K\|$ denotes the norm, from $L^2(x_n^\theta dx_n)$ into itself, of the corresponding operator which satisfies, for every $h \in L^2(x_n^\theta dx_n)$ and every $t \in \mathbb{R}$,

$$|\nabla_{x'} K(x')h(t)| \leq C \int \frac{\omega(x', t-s)}{\big((t-s)^2 + |x'|^2\big)^{(n+1)/2}} \, h(s)ds \, .$$

So, by homogeneity, it suffices to show that $\|k\| \leq C$, where k is the convolution operator with kernel $1/(t^2+1)^{(n+1)/2}$, acting on $L^2(x_n^\theta dx_n)$.

It is well known that, for every $t \in \mathbb{R}$, every $h \in L^2(x_n^\theta dx_n)$, $k(h)(t) \leq C \, Mh(t)$; where M denotes the maximal Hardy-Littlewood operator. This finishes the proof since M is bounded in $L^2(x_n^\theta dx_n)$ as x_n^θ belongs to (A_2).

Acknowledgements. Part of this work comes from my thesis (see [G2]), and I would like to express all my thanks to my advisor Aline Bonami. I would like to thank also Joaquim Bruna for his encouragements and helpful suggestions about this work.

References.

[AB] Ahern, P. and Bruna, J., Maximal and area integral characterizations of Hardy-Sobolev spaces in the unit ball of $\mathbb{C}^n$. *Revista Mat. Iberoamericana* **4** (1988), 123-153.

[AN] Ahern, P. and Nagel, A., Strong L^p estimates for maximal functions with respect to singular measures; with applications to exceptional sets. *Duke Math. J.* **53** (1986), 359-393.

[B1] Beatrous, F., Behavior of holomorphic functions near weakly pseudoconvex boundary points. *Indiana Univ. J. Math.* **40** (1991), 915-966.

[B2] Beatrous, F., Boundary estimates for derivatives of harmonic functions. *Studia Math.* **98** (1991), 53-71.

[C] Catlin, D., Estimates of invariant metrics on pseudoconvex domains of dimension two. *Math. Z.* **200** (1989), 429-466.

[CT] Calderón, A. P. and Torchinsky, A., Parabolic maximal functions associated with a distribution. *Advances in Math.* **16** (1975), 1-64.

[Co] Cohn, W., Tangential characterizations of Hardy-Sobolev spaces. *Indiana. Univ. J. Math.* **40** (1991), 1221-1249.

[CMS] Coifman, R., Meyer, Y. and Stein, E. M., Some new function spaces and their application to harmonic analysis. *J. Funct. Anal.* **62** (1985), 304-335.

[FS] Fefferman, C. and Stein, E. M., $\mathcal{H}^p$-spaces of several variables. *Acta Math.* **129** (1972), 137-193.

[G1] Grellier, S., Behavior of holomorphic functions in complex tangential directions in a domain of finite type in $\mathbb{C}^n$. *Publications Matematiques* **36** (1992), 1-41.

[G2] Grellier, S., Espaces de fonctions holomorphes dans les domaines de type fini. Thèse de l'Université d'Orléans (1991).

[GS] Greiner, P. C. and Stein, E. M., *Estimates for the $\overline{\partial}$-Neumann problem.* Princeton University Press (1977).

[Gr] Grisvard, P., *Elliptic problems in non smooth domains.* Pittman (1985).

[H] Hörmander, L., Subelliptic operators. Seminar on singularities of solutions. *Ann. of Math. Studies.* (1978) 127-208.

[Kr] Krantz, S. G., *Function Theory of Several Complex Variables.* John Wiley & sons (1982).

[NSW] Nagel, A., Stein, E. M. and Wainger, S., Boundary behavior of functions holomorphic in domains of finite type. *Proc. Nat. Acad. Sci. USA* **78** (1981), 6596-6599.

[RS] Rotschild, L. P. and Stein, E. M., Hypoelliptic differential operator and nilpotent groups. *Acta. Math.* **137** (1977), 248-315.

[S1] Stein, E.M., *Boundary behavior of holomorphic functions of several complex variables*, Princeton University Press (1972).

[S2] Stein, E. M., *Singular integrals and differentiability properties of functions*. Princeton University Press (1970).

Recibido: 1 de julio de 1.992

Sandrine Grellier
Département de Mathématiques
Université de Paris-Sud
91405 Orsay Cedex, FRANCE

Revista Matemática Iberoamericana
Vol. 9, N.° 2, 1993

Wiener-Hopf integral operators with PC symbols on spaces with Muckenhoupt weight

Albrecht Böttcher and Ilya M. Spitkovsky

Abstract. We describe the spectrum and the essential spectrum and give an index formula for Wiener-Hopf integral operators with piecewise continuous symbols on the space $L^p(\mathbb{R}_+, \omega)$ with a Muckenhoupt weight ω. Our main result says that the essential spectrum is a set resulting from the essential range of the symbol by joining the two endpoints of each jump by a certain sickle-shaped domain, whose shape is completely determined by the value of p and the behavior of the weight ω at the origin and at infinity.

1. Introduction.

Given $p \in (1, \infty)$, let A_p denote the set of all nonnegative functions w on $\mathbb{R}$ such that the singular integral operator S,

$$(Sf)(x) = \frac{1}{\pi i} \int_{-\infty}^{\infty} \frac{f(t)}{t - x}\, dt, \qquad x \in \mathbb{R},$$

is bounded on the space $L^p(\mathbb{R}, w)$ with the norm

$$||f||_{p,w} = \Big(\int_{-\infty}^{\infty} |w(x)f(x)|^p \Big)^{1/p}.$$

If $w \in A_p$, then the compression S_+ of S to the positive half-line $\mathbb{R}_+ = [0, \infty)$,

$$(S_+ f)(x) = \frac{1}{\pi i} \int_0^{\infty} \frac{f(t)}{t-x}\, dt, \qquad x \in \mathbb{R}_+,$$

is a bounded operator on $L^p(\mathbb{R}_+, w)$ $(= L^p(\mathbb{R}_+, w|\mathbb{R}_+))$. The operator S_+ is the archetypal example of a Wiener-Hopf integral operator with a piecewise continuous symbol: by definition, the symbol of S_+ is the function

$$\sigma(\xi) = -\operatorname{sgn} \xi = \begin{cases} 1 & \text{for } \xi \in (-\infty, 0), \\ -1 & \text{for } \xi \in (0, \infty). \end{cases}$$

A fairly general class of Wiener-Hopf integral operators is constituted by operators W of the form

$$(Wf)(x) = \sum_{j=1}^{m} \frac{c_j}{\pi i} \int_0^{\infty} \frac{e^{i\alpha_j (t-x)} f(t)}{t-x}\, dt + \int_0^{\infty} k(x-t) f(t)\, dt, \qquad x > 0,$$

where $c_j \in \mathbb{C}$ and $\alpha_j \in \mathbb{R}$ are given numbers and $k \in L^1(\mathbb{R})$ is a given function. The symbol of W is defined as the function

$$a(\xi) = -\sum_{j=1}^{m} c_j \operatorname{sgn}(\xi - \alpha_j) + \hat{k}(\xi), \qquad \xi \in \mathbb{R},$$

where $\hat{k}$ stands for the Fourier transform of k,

$$\hat{k}(\xi) = (Fk)(\xi) = \int_{-\infty}^{\infty} k(x) e^{i\xi x} dx, \qquad \xi \in \mathbb{R}.$$

Notice that a is a piecewise continuous function with jumps at $\alpha_1, \ldots, \alpha_m$ and at infinity.

What we are interested in here is the spectrum and essential spectrum of a Wiener-Hopf integral operator with a piecewise continuous symbol on $L^p(\mathbb{R}_+, w)$. As usual, the spectrum of W is the set of all $\lambda \in \mathbb{C}$ for which $W - \lambda I$ is not invertible. An operator W on $L^p(\mathbb{R}_+, w)$ is said to be Fredholm if it is invertible modulo the compact operators

or, equivalently, if its range is closed and the kernel and cokernel dimensions $\alpha(W)$ and $\beta(W)$ are finite; in that case the index of W is defined as $\alpha(W) - \beta(W)$. Finally, the essential spectrum of W is the set of all $\lambda \in \mathbb{C}$ for which $W - \lambda I$ is not Fredholm.

It has been well known for a long time that the spectrum and the essential spectrum of a Wiener-Hopf operator with a discontinuous symbol on $L^p(\mathbb{R}_+)$ or $L^p(\mathbb{R}_+, w)$ depend very sensitively on the value of p and the behavior of the weight w.

The pioneering work in this direction is undoubtedly Harold Widom's 1960 paper [16]. He observed that the spectrum of S_+ on $L^p(\mathbb{R}_+)$ is a certain circular arc depending on the value of p, namely the circular arc between -1 and 1 containing the point $-i\cot(\pi/p)$, which enabled him to identify the spectrum, the essential spectrum and the index of Toeplitz operators with piecewise continuous symbols on the (unweighted) Hardy spaces $H^p(\mathbb{R})$; the Toeplitz operators studied by Widom are just the operators we shall define in Section 2.9 below.

The hey-day of the development was the late sixties and early seventies. During that period Gohberg, Krupnik [8], and Duduchava [4],[5], to mention only the principal figures, considered pure Wiener-Hopf operators with piecewise continuous symbols on $L^p(\mathbb{R}_+)$ (without weight) and they proved that the essential spectrum of W is the continuous closed curve resulting from the range of the symbol a by joining the two endpoints of each jump by a certain circular arc. All these arcs are similar to one another and their shape is determined by p. The results of Gohberg, Krupnik, and Duduchava were extended by Schneider [14] to weights w of the form

$$w(x) = |x+i|^{\mu} \prod_{l=1}^{n} |x-\beta_l|^{\mu_l}, \qquad x \in \mathbb{R}.$$

He showed that the essential spectrum of W is again obtained from the range of a by filling in certain circular arcs. The interesting point of Schneider's criterion is that the circular arcs between $a(\alpha_j - 0)$ and $a(\alpha_j + 0)$ are all similar to one another and that their shape is determined solely by p and the behavior of the weight at infinity (*i.e.* the value of $\mu + \mu_1 + \cdots + \mu_n$), while the shape of the arc joining $a(+\infty)$ to $a(-\infty)$ depends only on p and the behavior of the weight at the point $x = 0$.

The main result of the present paper describes the essential spectrum of W in case w is any weight belonging to A_p. We show that this spectrum is obtained from the range of a by filling in a certain

sickle-shaped domain (which will be called a "horn") for each jump. A circular arc is regarded as a degenerate horn. It turns out that the horns joining $a(\alpha_j - 0)$ and $a(\alpha_j + 0)$ are again similar to one another and that their shape is given by merely the value of p and the behavior of the weight at infinity, whereas the shape of the horn between $a(+\infty)$ and $a(-\infty)$ depends on p and the behavior of the weight at $x = 0$ alone.

2. Toeplitz Operators.

2.1. Muckenhoupt weights on the circle.

Let $\mathbb{T}$ denote the complex unit circle and let ρ be a nonegative function on $\mathbb{T}$ which does not vanish identically. For $1 < p < \infty$, consider the space $L^p(\mathbb{T}, \rho)$ with the norm

$$||f||_{p,\rho} = \Big(\int_{\mathbb{T}} |f|^p \rho^p dm \Big)^{1/p},$$

where dm is Lebesgue measure on $\mathbb{T}$. If $\rho \equiv 1$, we abbreviate $L^p(\mathbb{T}, \rho)$ to $L^p(\mathbb{T})$. The weight ρ is said to be a Muckenhoupt weight, and we write $\rho \in A_p(\mathbb{T})$ in this case, if $\rho \in L^p(\mathbb{T})$, $\rho^{-1} \in L^q(\mathbb{T})$ $(1/p + 1/q = 1)$, and

$$\sup_I \Big(\frac{1}{|I|} \int_I \rho^p dm \Big)^{1/p} \Big(\frac{1}{|I|} \int_I \rho^{-q} dm \Big)^{1/q} < \infty,$$

where the supremum is over all subarcs I of $\mathbb{T}$ and $|I|$ denotes the arc length of I. Weights of this type first appeared in connection with the boundedness of the Hardy maximal function operator in Muckenhoupt's paper [11].

The singular integral operator S_0 ,

$$(S_0 f)(t) = \frac{1}{\pi i} \int_{\mathbb{T}} \frac{f(\tau)}{\tau - t} \, d\tau, \qquad t \in \mathbb{T},$$

is bounded on $L^p(\mathbb{T})$ by a theorem of Marcel Riesz. The problem of describing all the weights ρ such that S_0 maps $L^p(\mathbb{T}) \cap L^p(\mathbb{T}, \rho)$ into itself and extends from $L^p(\mathbb{T}) \cap L^p(\mathbb{T}, \rho)$ to a bounded operator on $L^p(\mathbb{T}, \rho)$ was solved by Hunt, Muckenhoupt and Wheeden [9]: S_0 extends to a bounded operator on $L^p(\mathbb{T}, \rho)$ if and only if $\rho \in A_p(\mathbb{T})$. We remark that

nice discussions of the Hunt-Muckenhoupt-Wheeden Theorem are also in [6], [7], [10] and [13]. Notice that a so-called power weight, given by

$$\rho(t) = \prod_{l=1}^{n} |t - \tau_l|^{\mu_l}, \qquad t \in \mathbb{T},$$

where $\tau_l \in \mathbb{T}$ and $\mu_l \in \mathbb{R}$, belongs to $A_p(\mathbb{T})$ if and only if $-1/p < \mu_l < 1/q$ for all l.

Troughout what follows let $\rho \in A_p(\mathbb{T})$. Then the two projections $P_0 = (I + S_0)/2$ and $Q_0 = (I - S_0)/2$ are bounded on $L^p(\mathbb{T}, \rho)$. The Hardy space $H^p(\mathbb{T}, \rho)$ is defined as the image of P_0 in $L^p(\mathbb{T}, \rho)$, *i.e.* $H^p(\mathbb{T}, \rho) = P_0 L^p(\mathbb{T}, \rho)$.

2.2. Toeplitz operators on the unit circle.

The Toeplitz operator $T_0(a)$ generated by a function $a \in L^\infty(\mathbb{T})$ is the operator on $H^p(\mathbb{T}, \rho)$ that sends f to $P_0(af)$. Since $\rho \in A_p(\mathbb{T})$, the operator $T_0(a)$ is bounded. The function a is usually referred to as the symbol of $T_0(a)$.

A well known theorem by L.A. Coburn says that $T_0(a)$ is invertible if and only if $T_0(a)$ is Fredholm with index zero (see *e.g.* [2, p. 216]). Hence, in order to study invertibility of Toeplitz operators (or, equivalently, in order to describe their spectrum), it suffices to establish a Fredholm criterion (or to describe the essential spectrum) and to have an index formula.

Given a Banach algebra $\mathfrak{A}$ with identity element, we denote by $G\mathfrak{A}$ the invertible elements in $\mathfrak{A}$. The Hartman-Wintner Theorem (again see [2, p. 216]) tells us that if $T_0(a)$ is Fredholm, then $a \in GL^\infty(\mathbb{T})$. If a is continuous, $a \in C(\mathbb{T})$, then the invertibility of a in $L^\infty(\mathbb{T})$ (and thus in $C(\mathbb{T})$) is also sufficient for $T_0(a)$ to be Fredholm, and the index of $T_0(a)$ is then minus the winding number of $a(\mathbb{T})$ about the origin. In general, however, symbols $a \in GL^\infty(\mathbb{T})$ do not induce Fredholm Toeplitz operators.

Let $PC(\mathbb{T})$ denote the C^*-algebra of all piecewise continuous functions on $\mathbb{T}$. A function $a \in PC(\mathbb{T})$ has at most countably many jumps and the limits

$$a(t \pm 0) = \lim_{\varepsilon \to 0 \pm 0} a(te^{i\varepsilon})$$

exist for each $t \in \mathbb{T}$. Under the sole assumption that $\rho \in A_p(\mathbb{T})$, a Fredholm criterion and an index formula for Toeplitz operators on

$H^p(\mathbb{T},\rho)$ with symbols in $PC(\mathbb{T})$ were only recently obtained by one of the authors [15]. Before citing this result we need a (crucial) lemma and the definition of what we call horns.

Lemma 2.3. ([15]). *Let $\rho \in A_p(\mathbb{T})$ and $\tau \in \mathbb{T}$. Then the set*

$$I_\tau(p,\rho) = \{\mu \in \mathbb{R} : |t-\tau|^\mu \rho(t) \in A_p(\mathbb{T})\}$$

is an open interval of a length not greater than 1 containing the origin.

REMARK 2.4. If ρ is a power weight as in Section 2.1, then clearly

$$I_{\tau_l}(p,\rho) = (-1/p-\mu_l\,, 1/q-\mu_l)\,,$$
$$I_\tau(p,\rho) = (-1/p\,, 1/q) \quad \text{for } \tau \notin \{\tau_1,\dots,\tau_n\}\,,$$

i.e. all $I_\tau(p,\rho)$ have length 1. To produce a weight $\rho \in A_p(\mathbb{T})$ such that $I_\tau(p,\rho)$ is any prescribed interval $(-\alpha,\beta) \ni 0$ of a length $\alpha+\beta<1$, let first PQC denote the C^*-algebra of all piecewise quasicontinuous functions on $\mathbb{T}$ (see [2] or [3]). In [3], we showed that there exist $a \in PQC$ with a logarithm $\log a \in PQC$ such that $T_0(a)$ is invertible on $H^p(\mathbb{T},|t-\tau|^\mu)$ if and only if $1/p+\mu \in (0,\alpha+\beta)$. This implies (see [15]) that if we put

$$\rho(t) = |\exp(P_0(\log a))|\,|t-\tau|^{\alpha-1/p}\,,$$

then $\rho \in A_p(\mathbb{T})$ and $I_\tau(p,\rho) = (-\alpha,\beta)$.

Definition 2.5. *For $\rho \in A_p(\mathbb{T})$ and $\tau \in \mathbb{T}$, let $I_\tau(p,\rho)$ be the interval determined by Lemma 2.3 and define the numbers $\nu_\tau^\pm(p,\rho)$ by*

$$(-\nu_\tau^-(p,\rho), 1-\nu_\tau^+(p,\rho)) = I_\tau(p,\rho)\,.$$

Because $I_\tau(p,\rho)$ contains the origin and is of a length not greater than 1, we have

$$0 < \nu_\tau^-(p,\rho) \le \nu_\tau^+(p,\rho) < 1\,.$$

2.6. Horns.

In what follows the argument of a nonzero complex number is always specified to belong to $[0,2\pi)$. Given two real numbers γ,δ such

that $0 < \gamma \leq \delta < 1$ and two distinct complex numbers z, w, we define the γ, δ horn joining z and w to be the set

$$\mathcal{H}(z, w; \gamma, \delta) = \{\zeta \in \mathbb{C}\backslash\{z, w\} : \arg \frac{\zeta - z}{\zeta - w} \in [2\pi\gamma, 2\pi\delta]\} \cup \{z, w\}.$$

Notice that for each $\phi \in (0, 1)$ the set

$$\{\zeta \in \mathbb{C}\backslash\{z, w\} : \arg \frac{\zeta - z}{\zeta - w} = 2\pi\phi\}$$

is a circular arc. If $\phi = 1/2$, this arc degenerates to the open line segment (z, w). For $\phi \in (0, 1/2)$ (respectively $\phi \in (1/2, 1)$) this arc is located on the right (respectively, left) of the straight line passing first z and then w, and it consists just of the points at which the segment $[z, w]$ is seen at the angle $2\pi\phi$ (respectively, $2\pi(1 - \phi)$). To cover the case $z = w$, we also define $\mathcal{H}(z, z; \gamma, \delta) = \{z\}$.

Note that $0 \notin \mathcal{H}(z, w; \gamma, \delta)$ if and only if $z \neq 0$, $w \neq 0$, and $\arg(z/w)$ does not belong to $[2\pi\gamma, 2\pi\delta]$.

For $a \in PC(\mathbb{T})$, the set

$$a_{p,\rho} = \bigcup_{\tau \in \mathbb{T}} \mathcal{H}(a(\tau - 0), a(\tau + 0); \nu_\tau^-(p, \rho), \nu_\tau^+(p, \rho))$$

results from the (essential) range of a by filling in a well-defined horn into each jump. The set $a_{p,\rho}$ is clearly connected. Any closed continuous curve obtained from the (essential) range of a by joining $a(t - 0)$ to $a(t + 0)$ by a circular arc contained in the horn between $a(t - 0)$ and $a(t + 0)$ inherits an orientation in a natural fashion. If $0 \notin a_{p,\rho}$, we denote by $\operatorname{wind} a_{p,\rho}$ the winding number of that curve about the origin.

Theorem 2.7. ([15]). *If $\rho \in A_p(\mathbb{T})$ and $a \in PC(\mathbb{T})$, then the essential spectrum of $T_0(a)$ on $H^p(\mathbb{T}, \rho)$ is the set $a_{p,\rho}$. In case $0 \notin a_{p,\rho}$, the index of $T_0(a)$ on $H^p(\mathbb{T}, \rho)$ equals $-\operatorname{wind} a_{p,\rho}$.*

2.8. Muckenhoupt weights on the real line.

Our next concern is to carry over Theorem 2.7 to Toeplitz operators on Hardy spaces of the real line.

For $p \in (1, \infty)$ and a nonnegative function w on $\mathbb{R}$ which is not identically zero, we consider the space $L^p(\mathbb{R}, w)$, whose norm is given by

$$||f||_{p,w} = \Big(\int_{\mathbb{R}} |w(x)f(x)|^p dx \Big)^{1/p}.$$

Again we abbreviate $L^p(\mathbb{R}, 1)$ to $L^p(\mathbb{R})$.

We write $w \in A_p$ and call w a Muckenhoupt weight if $w \in L^p(\mathbb{R})$, $w^{-1} \in L^q(\mathbb{R})$ $(1/p + 1/q = 1)$, and

$$\sup_I \Big(\frac{1}{|I|} \int_I w(x)^p dx \Big)^{1/p} \Big(\frac{1}{|I|} \int_I w(x)^{-q} dx \Big)^{1/q} < \infty,$$

where I ranges over all finite intervals $I \subset \mathbb{R}$ and $|I|$ stands for the length of the interval I.

The singular integral operator S (see the Introduction) is bounded on $L^p(\mathbb{R})$, and it was also Hunt, Muckenhoupt and Wheeden [9] who showed that S maps $L^p(\mathbb{R}) \cap L^p(\mathbb{R}, w)$ into itself and extends to a bounded operator on $L^p(\mathbb{R}, w)$ if and only if $w \in A_p$.

Henceforth let always $w \in A_p$. The projections $P = (I + S)/2$ and $Q = (I - S)/2$ are bounded on $L^p(\mathbb{R}, w)$, and the image of P in $L^p(\mathbb{R}, w)$ is denoted by $H^p(\mathbb{R}, w)$ and called the p-th Hardy space of $\mathbb{R}$ with the weight function w.

2.9. Toeplitz operators on the real line.

Given $a \in L^\infty(\mathbb{R})$, define the Toeplitz operator $T(a)$ on $H^p(\mathbb{R}, w)$ by $T(a)f = P(af)$. Since $w \in A_p$, this is a bounded operator. Again the function a is called the symbol of $T(a)$.

The Coburn and Hartman-Wintner theorems extend to Toeplitz operators on $H^p(\mathbb{R}, w)$: the operator $T(a)$ is invertible if and only if it is Fredholm of index zero, and the Fredholmness of $T(a)$ implies the invertibility of a in $L^\infty(\mathbb{R})$. If $a \in C(\dot{\mathbb{R}})$, which means that $a \in C(\mathbb{R}) \cap L^\infty(\mathbb{R})$ and that the limits $a(\pm\infty)$ exist and are equal to each other, then for $T(a)$ to be Fredholm on $H^p(\mathbb{R}, w)$ it is necessary and sufficient that $a \in GL^\infty(\mathbb{R})$; in that case the index of $T(a)$ is minus the winding number of the range of a about the origin.

Let PC be the C^*-subalgebra of $L^\infty(\mathbb{R})$ consisting of all functions a in $L^\infty(\mathbb{R})$ which have limits $a(\xi \pm 0)$ for each $\xi \in \mathbb{R}$ and for which the limits $a(\pm\infty)$ exist. Note that functions in PC have at most countably

many jumps. Also notice that PC contains $C(\bar{\mathbb{R}})$, the C^*-algebra of all functions $a \in C(\mathbb{R}) \cap L^\infty(\mathbb{R})$ with finite (but not necessarily equal) limits $a(\pm\infty)$.

Theorem 2.10. *Let $w \in A_p$ and $a \in PC$.*

(1) *Each of the sets*

$$I_\xi(p,w) = \{\mu \in \mathbb{R} : \left|\frac{x-\xi}{x-i}\right|^\mu w(x) \in A_p\}, \qquad \xi \in \mathbb{R},$$
$$I_\infty(p,w) = \{\mu \in \mathbb{R} : |x-i|^{-\mu} w(x) \in A_p\}$$

is an open interval of a length not greater than 1 which contains the origin

$$I_\xi(p,w) = (-\nu_\xi^-(p,w), 1-\nu_\xi^+(p,w)) \qquad (\xi \in \dot{\mathbb{R}} \stackrel{\text{def}}{=} \mathbb{R} \cup \{\infty\})$$

with $0 < \nu_\xi^-(p,w) \le \nu_\xi^+(p,w) < 1$.

(2) *The essential spectrum of $T(a)$ on $H^p(\mathbb{R},w)$ equals*

$$a_{p,w} = \Big(\bigcup_{\xi\in\mathbb{R}} \mathcal{H}\big(a(\xi-0), a(\xi+0); \nu_\xi^-(p,w), \nu_\xi^+(p,w)\big)\Big)$$
$$\bigcup \mathcal{H}\big(a(+\infty), a(-\infty); \nu_\infty^-(p,w), \nu_\infty^+(p,w)\big).$$

If $0 \notin a_{p,w}$, then the index of $T(a)$ is $-\operatorname{wind} a_{p,w}$.

PROOF. We reduce the case of the real line to the situation on the circle in a standard way (see *e.g.* [8, p. 307]).

Define the weight ρ on $\mathbb{T}$ by

$$\rho(t) = w\Big(i\frac{t+1}{t-1}\Big)|t-1|^{1-2/p}, \qquad t \in \mathbb{T}.$$

Then the operator $B : L^p(\mathbb{R},w) \to L^p(\mathbb{T},\rho)$ given by

$$(B\phi)(t) = \frac{1}{t-1}\,\phi\Big(i\frac{t+1}{t-1}\Big), \qquad t \in \mathbb{T},\ \phi \in L^p(\mathbb{R},w),$$

is an isomorphism, the inverse operator being

$$(B^{-1}\psi)(x) = \frac{2i}{x-i}\,\psi\Big(\frac{x+i}{x-i}\Big), \qquad x \in \mathbb{R},\ \psi \in L^p(\mathbb{T},\rho).$$

Moreover, $B^{-1}S_0B = -S$. The latter equality in conjunction with the Hunt-Muckenhoupt-Wheeden theorems implies that $\rho \in A_p(\mathbb{T})$ if and only if $w \in A_p$ and also that, for $\xi = i(\tau+1)/(\tau-1) \in \mathbb{R}$,

$$\left|\frac{x-\xi}{x-i}\right|^{\mu} w(x) \in A_p$$

if and only if

$$\left|\frac{i\frac{t+1}{t-1} - i\frac{\tau+1}{\tau-1}}{i\frac{t+1}{t-1} - i}\right|^{\mu} w\Big(i\frac{t+1}{t-1}\Big)|t-1|^{1-2/p} = |t-\tau|^{\mu}\rho(t) \in A_p(\mathbb{T}),$$

and, analogously, that

$$|x-i|^{-\mu} w(x) \in A_p$$

exactly if

$$\left|i\frac{t+1}{t-1} - i\right|^{-\mu} w\Big(i\frac{t+1}{t-1}\Big)|t-1|^{1-2/p} = 2^{-\mu}|t-1|^{\mu}\rho(t) \in A_p(\mathbb{T}).$$

So part (1) of the present theorem is an immediate consequence of Lemma 2.3.

Let us now show that $0 \notin a_{p,w}$ if and only if $T(a)$ is Fredholm. Since the essential range of a is a subset of $a_{p,w}$ and the Fredholmness of $T(a)$ necessitates the invertibility of a in $L^{\infty}(\mathbb{R})$, we may without loss of generality a priori assume that $a \in GL^p(\mathbb{R})$.

It is easily seen that $T(a) = Pa|\mathrm{Im}\ P$ is Fredholm of index κ on $H^p(\mathbb{R}, w) = PL^p(\mathbb{R}, w)$ if and only if the operator $Q + PaP$ is Fredholm of index κ on $L^p(\mathbb{R}, w)$. Because $BSB^{-1} = -S_0$, it follows that

$$\begin{aligned} B(Q+PaP)B^{-1} &= P_0 + Q_0BaB^{-1}Q_0 \\ &= P_0 + Q_0bQ_0 \\ &= (P_0 + bQ_0)(I - P_0bQ_0) \\ &= b(b^{-1}P_0 + Q_0)(I - P_0bQ_0) \\ &= b(Q_0 + P_0b^{-1}P_0)(I + Q_0b^{-1}P_0)(I - P_0bQ_0), \end{aligned}$$

where $b(t) = a\big(i(t+1)/(t-1)\big)$ for $t \in \mathbb{T}$. The operators $I+Q_0b^{-1}P_0$ and $I-P_0bQ_0$ are invertible, the inverses being $I-Q_0b^{-1}P_0$ and $I+P_0bQ_0$, respectively. Since $a \in GL^{\infty}(\mathbb{R})$, the operator of multiplication by b is

invertible as well. Hence, $T(a)$ is Fredholm of index κ on $H^p(\mathbb{R}, w)$ if and only if

$$T_0(b^{-1}) = P_0 b^{-1}|\mathrm{Im}\ P_0$$

is Fredholm of index κ on $H^p(\mathbb{T}, \rho) = P_0 L^p(\mathbb{T}, \rho)$.

From Theorem 2.7 we infer that $T_0(b^{-1})$ is Fredholm if and only if

$$\begin{aligned} 0 \notin (b^{-1})_{p,\rho} &= \bigcup_{\tau\in\mathbb{T}} \mathcal{H}\big(b^{-1}(\tau-0), b^{-1}(\tau+0); \nu_\tau^-(p,\rho), \nu_\tau^+(p,\rho)\big) \\ &= \bigcup_{\tau\in\mathbb{T}\setminus\{1\}} \mathcal{H}\big(b^{-1}(\tau-0), b^{-1}(\tau+0); \nu_\tau^-(p,\rho), \nu_\tau^+(p,\rho)\big) \\ &\qquad \bigcup\ \mathcal{H}\big(b^{-1}(1-0), b^{-1}(1+0); \nu_0^-(p,\rho), \nu_0^+(p,\rho)\big) \\ &= \bigcup_{\xi\in\mathbb{R}} \mathcal{H}\big(a^{-1}(\xi+0), a^{-1}(\xi-0); \nu_\xi^-(p,\rho), \nu_\xi^+(p,\rho)\big) \\ &\qquad \bigcup\ \mathcal{H}\big(a^{-1}(-\infty), a^{-1}(+\infty); \nu_\infty^-(p,\rho), \nu_\infty^+(p,\rho)\big)\,. \end{aligned}$$

Consequently, $0 \notin (b^{-1})_{p,\rho}$ exactly if

$$\arg\frac{a^{-1}(\xi+0)}{a^{-1}(\xi-0)} = \arg\frac{a(\xi-0)}{a(\xi+0)} \notin [2\pi\nu_\xi^-(p,w), 2\pi\nu_\xi^+(p,w)]$$

for all $\xi \in \mathbb{R}$ and

$$\arg\frac{a^{-1}(-\infty)}{a^{-1}(+\infty)} = \arg\frac{a(+\infty)}{a(-\infty)} \notin [2\pi\nu_\infty^-(p,w), 2\pi\nu_\infty^+(p,w)]\,,$$

which is equivalent to the condition that $0 \notin a_{p,w}$.

3. Convolution operators on the real line.

3.1. Fourier multipliers.

Again let $1 < p < \infty$ and $w \in A_p$. Denote by $F : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ the Fourier transform and by F^{-1} its inverse. If a is any function defined on $\mathbb{R}$, the multiplication operator $f \mapsto af$ is traditionally denoted by aI, and in case aI is applied after another operator, B say, one writes aB instead of aIB.

A function $a \in L^\infty(\mathbb{R})$ is called a Fourier multiplier on $L^p(\mathbb{R}, w)$ if the mapping $f \mapsto F^{-1}aFf$ maps $L^2(\mathbb{R}) \cap L^p(\mathbb{R}, w)$ into itself and

extends to a bounded operator of $L^p(\mathbb{R}, w)$ into itself. The latter operator is then usually denoted by $W^0(a)$. It is well known that the set $M^p(w)$ of all Fourier multipliers on $L^p(\mathbb{R}, w)$ is a Banach algebra under the norm

$$||a||_{p,w} = ||W^0(a)||_{\mathcal{L}(L^p(\mathbb{R},w))} ,$$

where $\mathcal{L}(X)$ stands for the Banach algebra of all bounded operators on a Banach space X.

One can show (see *e.g.* [5] and [12] for power weights) that $M^p(w)$ contains all functions $a \in L^\infty(\mathbb{R})$ with finite total variation $\mathrm{Var}\,(a)$ and that for such functions the estimate

$$||a||_{p,w} \le c_{p,w}\,(||a||_\infty + \mathrm{Var}\,(a))$$

holds, where $c_{p,w}$ is some constant independent of a.

Let $\dot{\mathbb{R}}$ be the compactification of $\mathbb{R}$ by one point at infinity. The closure in $M^p(w)$ of the set of all functions $a \in C(\dot{\mathbb{R}})$ with $\mathrm{Var}\,(a) < \infty$ is denoted by $C^p(w)$, and we let $PC^p(w)$ stand for the closure in $M^p(w)$ of the set of all piecewise continuous functions on $\mathbb{R}$ which have finite total variation and at most finitely many jumps. Clearly $C^p(w) \subset PC^p(w)$. It can be shown (see [14] and [12, Proposition 12.2] for power weights) that $PC^p(w)$ is continuously embedded into $L^\infty(\mathbb{R})$. This implies that $C^p(w) \subset C(\dot{\mathbb{R}})$ and $PC^p(w) \subset PC$, where PC refers to the C^*-subalgebra of $L^\infty(\mathbb{R})$ consisting of all functions a which possess finite one-sided limits $a(\xi \pm 0)$ at every point $\xi \in \dot{\mathbb{R}}$. Moreover, a function $a \in C^p(w)$ (respectively, $a \in PC^p(w)$) is invertible in $C^p(w)$ (respectively, $PC^p(w)$) if and only if it is invertible in $L^\infty(\mathbb{R})$ (see [12] for the case of power weights).

3.2. Wiener-Hopf integral operators.

The Wiener-Hopf operator $W(a)$ generated by a function $a \in M^p(w)$ (its so-called symbol) is the compression of $W^0(a)$ to the positive half-line $\mathbb{R}_+ = (0, \infty)$, *i.e.* $W(a)$ is the bounded operator on $L^p(\mathbb{R}_+, w)$ $(= L^p(\mathbb{R}_+, w|\mathbb{R}_+))$ acting by the rule $f \mapsto (W^0(a)f)|\mathbb{R}_+$. Let χ_+ be the characteristic function of $\mathbb{R}_+$. The space $L^p(\mathbb{R}_+, w)$ may be identified with $\chi_+ L^p(\mathbb{R}, w)$ and consequently, we may also think of $W(a)$ as the operator $\chi_+ W^0(a)|\mathrm{Im}\ \chi_+ I$.

Our aim is to describe the spectrum of $W(a)$ on $L^p(\mathbb{R}_+, w)$ if a belongs to $PC^p(w)$. The proof of Proposition 2.8 of [5] for $w \equiv 1$ along

with the arguments used in the proof of Proposition 1.6 of [14] for power weights can be easily modified to show that if $a \in PC^p(w) \cap GL^\infty(\mathbb{R})$, then $W(a)$ is invertible on $L^p(\mathbb{R}_+, w)$ if and only if $W(a)$ is Fredholm of index zero on $L^p(\mathbb{R}_+, w)$. We shall prove below that $a \in GL^\infty(\mathbb{R})$ whenever $a \in PC^p(w)$ and $W(a)$ is Fredholm. Thus, in order to identify the spectrum of $W(a)$ we are again left with finding a Fredholm criterion and an index formula.

We finally remark that if $a \in C^p(w)$, then $W(a)$ is Fredholm on $L^p(\mathbb{R}_+, w)$ exactly if $a(\xi) \neq 0$ for all $\xi \in \dot{\mathbb{R}}$, in which case the index of $W(a)$ equals minus the winding number of the naturally oriented curve $a(\dot{\mathbb{R}})$ about the origin (see [2], [4], [8] and [12] for power weights).

3.3. Singular integral operators.

The connection between Toeplitz operators on $H^p(\mathbb{R}, w)$ and Wiener-Hopf operators on $L^p(\mathbb{R}_+, w)$ is established by singular integral operators on $L^p(\mathbb{R}, w)$, *i.e.* operators of the form $bI + cS = bI + cW^0(\sigma)$ or, slightly more generally, of the form

$$\lambda_\eta^{-1}(bI + cS)\lambda_\eta I = bI + cW^0(\sigma_\eta)\,,$$

where $\lambda_\eta(x) = e^{i\eta x}$ for $x, \eta \in \mathbb{R}$, $\sigma(\xi) = -\operatorname{sgn}\xi$ for $\xi \in \mathbb{R}$, and $\sigma_\eta(\xi) = -\operatorname{sgn}(\xi - \eta)$ for $\xi, \eta \in \mathbb{R}$.

Let χ_- and χ_+ be the characteristic functions of $(-\infty, 0)$ and $(0, +\infty)$, respectively. We have

$$(I + \chi_+ W^0(a)\chi_- I)(\chi_- I + \chi_+ W^0(a)\chi_+ I) = \chi_- I + \chi_+ W^0(a)\,,$$

and since $I + \chi_+ W^0(a)\chi_- I$ has the inverse $I - \chi_+ W^0(a)\chi_- I$, it follows that the Wiener-Hopf operator $W(a) = \chi_+ W^0(a)|\operatorname{Im}\chi_+ I$ is Fredholm on $L^p(\mathbb{R}_+, w)$ if and only if so is the operator $\chi_- I + \chi_+ W^0(a)$ on $L^p(\mathbb{R}, w)$.

In the next chapters we shall use localization techniques to reduce the study of $\chi_- I + \chi_+ W^0(a)$ to the investigation of the operators

$$\chi_- I + \chi_+ W^0(a(\eta - 0)\chi_\eta^- + a(\eta + 0)\chi_\eta^+)\,, \qquad \eta \in \mathbb{R}\,,$$

and

$$\chi_- I + \chi_+ W^0(a(-\infty)\chi_0^- + a(+\infty)\chi_0^+)\,,$$

where χ_η^- and χ_η^+ are, respectively, the characteristic functions of $(-\infty,\eta)$ and $(\eta,+\infty)$. But if $\alpha,\beta\in\mathbb{C}$, then

$$\begin{aligned}\chi_-I+\chi_+W^0(\alpha\chi_\eta^-+\beta\chi_\eta^+) &= \chi_-I+\chi_+W^0\Big(\alpha\frac{1+\sigma_\eta}{2}+\beta\frac{1-\sigma_\eta}{2}\Big)\\ &= \Big(\chi_-+\frac{\alpha+\beta}{2}\chi_+\Big)I+\chi_+\frac{\alpha-\beta}{2}W^0(\sigma_\eta)\\ &= bI+cW^0(\sigma_\eta)\,,\end{aligned}$$

and since $W^0(\sigma_\eta)=\lambda_\eta^{-1}S\lambda_\eta I$ and $\lambda_\eta^{\pm1}I$ are isomorphisms, we arrive at the operators

$$\begin{aligned}bI+cS &= b(P+Q)+c(P-Q)=(b+c)P+(b-c)Q\\ &= (b-c)\Big(\frac{b+c}{b-c}P+Q\Big)=(b-c)(dP+Q)\,.\end{aligned}$$

Because now $(dP+Q)=(PdP+Q)(I+QdP)$ and $I+QdP$ has the inverse $I-QdP$, we are finally led to operators of the form $Pd|\mathrm{Im}\ P$ with $d\in PC$. But the latter operators are just the Toeplitz operators $T(b)$ on $\mathrm{Im}\ P=H^p(\mathbb{R},w)$ we have already studied in Chapter 2.

For further reference we summarize part of the preceding reasoning.

Lemma 3.4. *Let $b,c\in PC$ and $\eta\in\mathbb{R}$, and suppose $b-c\in GL^\infty(\mathbb{R})$. Then $bI+cW^0(\sigma_\eta)$ is Fredholm on $L^p(\mathbb{R},w)$ if and only if $T\big((b+c)/(b-c)\big)$ is Fredholm on $H^p(\mathbb{R},w)$.*

4. Local singular integral operators.

4.1. Preliminaries.

To carry out the program sketched in Section 3.3 we make use of the local principle of Gohberg and Krupnik [8] (see also [1], [2], [5] and [12]).

Let $\mathfrak{A}$ be a Banach algebra with identity element. A bounded subset $\mathfrak{M}\subset\mathfrak{A}$ is called a localizing class if $0\notin\mathfrak{M}$ and for every two elements $B_1,B_2\in\mathfrak{M}$ there exists a third element $B\in\mathfrak{M}$ such that $B_jB=BB_j=B$ for $j=1,2$. A family $\{\mathfrak{M}_\tau\}_{\tau\in T}$ of localizing classes is said to be covering if for every choice $\{B_\tau\}_{\tau\in T}$ of elements $B_\tau\in\mathfrak{M}_\tau$ there exist finitely many $\tau_1,\dots,\tau_m$ such that $B_{\tau_1}+\cdots+B_{\tau_m}$ is invertible in $\mathfrak{A}$.

Let now $\{\mathfrak{M}_\tau\}_{\tau\in T}$ be a covering family of localizing classes in $\mathfrak{A}$ and put $\mathfrak{B} = \bigcup\{\mathfrak{M}_\tau : \tau \in T\}$. The commutant $\operatorname{Com}\mathfrak{B}$ is a closed subalgebra of $\mathfrak{A}$. Fot $\tau \in T$, define

$$\mathfrak{Z}_\tau = \{A \in \operatorname{Com}\mathfrak{B} : \inf_{B\in\mathfrak{M}_\tau} ||AB|| = \inf_{B\in\mathfrak{M}_\tau} ||BA|| = 0\}\,.$$

One can easily show that $\mathfrak{Z}_\tau$ is a closed proper two-sided ideal of $\operatorname{Com}\mathfrak{B}$. Finally, for $A \in \operatorname{Com}\mathfrak{B}$, denote by A_τ the coset of the quotient algebra $\operatorname{Com}\mathfrak{B}/\mathfrak{Z}_\tau$ containing A.

Theorem 4.2. (Local principle of Gohberg and Krupnik, [8]). *With the notation introduced in Section* 4.1, *an element* $A \in \operatorname{Com}\mathfrak{B}$ *is invertible in* $\mathfrak{A}$ *if and only if* A_τ *is invertible in* $\operatorname{Com}\mathfrak{B}/\mathfrak{Z}_\tau$ *for every* $\tau \in T$.

4.3. The algebra $\mathfrak{A}$.

We apply Theorem 4.2 to the Calkin algebra $\mathfrak{A} = \mathcal{L}/\mathcal{K}$, where $\mathcal{L}$ is the Banach algebra of all bounded operators on $L^p(\mathbb{R}_+, w)$ and $\mathcal{K}$ stands for the ideal of all compact operators on $L^p(\mathbb{R}_+, w)$.

4.4. Localizing classes in $\mathfrak{A}$.

We are interested in "localizing" operators of the form $bI+cW^0(a)$, where $b, c \in PC$ and $a \in PC^p(w)$. In order to "localize" the coefficients b and c at $y \in \dot{\mathbb{R}}$ and the symbol a at $\eta \in \dot{\mathbb{R}}$, we consider the following candidates $\mathfrak{M}_{y,\eta}$ for localizing classes in $\mathfrak{A} = \mathcal{L}/\mathcal{K}$: the set $\mathfrak{M}_{y,\eta}$ consists of all cosets of the form $vW^0(u)+\mathcal{K}$ such that $v, u \in C(\dot{\mathbb{R}})$ are piecewise linear with finite total variation and v (respectively, u) is identically 1 in some open neighborhood of y (respectively, η) and identically 0 outside some other open neighborhood of y (respectively, η). One can show as in [5] (for $w \equiv 1$) or in [12] (for power weights) that $\mathfrak{M}_{y,\eta}$ coincides with $\mathcal{K}$ and thus the zero in $\mathcal{L}/\mathcal{K}$ whenever $y \in \mathbb{R}$ and $\eta \in \mathbb{R}$. However, if

$$(y,\eta) \in (\dot{\mathbb{R}} \times \dot{\mathbb{R}})\backslash(\mathbb{R}\times\mathbb{R}) = (\mathbb{R}\times\{\infty\}) \cup (\{\infty\}\times\mathbb{R}) \cup \{(\infty,\infty)\} = T\,,$$

then $\mathfrak{M}_{y,\eta}$ is indeed a localizing class in $\mathcal{L}/\mathcal{K}$ (again see [5] or [12] for power weights).

To check whether $\{\mathfrak{M}_{(y,\eta)}\}_{(y,\eta)\in T}$ is a covering family, *i.e.* whether $\sum_{j=1}^{m} u_j W^0(v_j)$ is Fredholm provided $\sum u_j \geq 1$ and $\sum v_j \geq 1$, and to decide whether the cosets we are interested in, namely

$$bI + cW^0(a) + \mathcal{K}, \qquad b, c \in PC,\ a \in PC^p(w),$$

belong to $\operatorname{Com}\mathfrak{B}$ a good piece of work must be done. For $w \equiv 1$ all this is done in [5], and for power weights a detailed exposition of these things is in [12] (see also [14]). It is not difficult to convince oneself that the arguments of [5] and [12] extend to arbitrary Muckenhoupt weights and thus show that $\{\mathfrak{M}_{(y,\eta)}\}_{(y,\eta)\in T}$ is a covering family of localizing classes in $\mathcal{L}/\mathcal{K}$ and that all the cosets mentioned above are in $\operatorname{Com}\mathfrak{B}$.

4.5. Localization in $\mathfrak{A}$.

For $b, c \in PC, a \in PC^p(w)$ and $(y,\eta) \in T$ we put

$$[bI + cW^0(a)]^{\pi}_{y,\eta} = [bI + cW^0(a) + \mathcal{K}]_{(y,\eta)}.$$

One can show (as in [4] and [12]) that

$$[bI + cW^0(a)]^{\pi}_{y,\eta} = [b_y I + c_y W^0(a_\eta)]^{\pi}_{y,\eta}$$

if $b_y,\ c_y \in PC$ and $a_\eta \in PC^p(w)$ are any function such that $b - b_y, c - c_y$ are continuous at y and $a - a_\eta$ is continuous at η. Hence, instead with an operator $bI + cW^0(a)$, one has to deal with the in general simpler operators $b_y I + c_y W^0(a_\eta)$; the price for this reduction is that invertibility in $\mathcal{L}/\mathcal{K}$ is replaced by invertibility in $\operatorname{Com}\mathfrak{B}/\mathfrak{Z}_{(y,\eta)}$.

Our main concern in this chapter is the invertibility of the elements ("local singular integral operators") $[bI + cW^0(\sigma_\zeta)]^{\pi}_{y,\eta}$ in $\operatorname{Com}\mathfrak{B}/\mathfrak{Z}_{(y,\eta)}$.

Lemma 4.6. *Let $b, c \in PC$, assume that $b - c$ and $b + c$ are both invertible in $L^\infty(\mathbb{R})$ (and thus in PC), and put $d = d_{b,c} = (b+c)/(b-c)$. Suppose further that $y, \eta, \zeta \in \mathbb{R}$. Then*

(1) *$[bI + cW^0(\sigma_\zeta)]^{\pi}_{y,\infty}$ is invertible if and only if*

$$0 \notin \mathcal{H}(d(y-0), d(y+0); \nu_y^-(p,w), \nu_y^+(p,w));$$

(2) *$[bI + cW^0(\sigma_\zeta)]^{\pi}_{\infty,\infty}$ is invertible;*

(3) $[bI + cW^0(\sigma_\zeta)]^\pi_{\infty,\eta}$ *is invertible if* $\eta \neq \zeta$;

(4) $[bI + cW^0(\sigma_\zeta)]^\pi_{\infty,\zeta}$ *is invertible if and only if*

$$0 \notin \mathcal{H}(d(+\infty), d(-\infty); \nu^-_\infty(p,w), \nu^+_\infty(p,w)) .$$

PROOF. (1) We have

$$[bI + cW^0(\sigma_\zeta)]^\pi_{y,\infty} = [b_y I + c_y W^0(\sigma_\zeta)]^\pi_{y,\infty} ,$$

where $b_y, c_y \in PC$ are any functions such that $b_y(y \pm 0) = b(y \pm 0)$ and $c_y(y \pm 0) = c(y \pm 0)$. The functions b_y and c_y may be chosen to be continuous on $\dot{\mathbb{R}}\backslash\{y\}$ and to satisfy $b_y \pm c_y \in GPC$ and $d_{b_y,c_y}(x) \neq 0$ for all $x \in \dot{\mathbb{R}}\backslash\{y\}$.

Suppose first that 0 does not belong to the horn $\mathcal{H} = \mathcal{H}(\cdots)$. We then infer from Lemma 3.4 and Theorem 2.10(1) that $b_y I + c_y W^0(\sigma_\zeta)$ is Fredholm, which, by Theorem 4.2, implies that $[b_y I + c_y W^0(\sigma_\zeta)]^\pi_{y,\infty}$ is all the more invertible.

Now suppose $0 \in \mathcal{H}$ and, contrary to what we want, assume $[b_y I + c_y W^0(\sigma_\zeta)]^\pi_{y,\infty}$ is invertible. For $x \in \dot{\mathbb{R}}\backslash\{y\}$ we have

$$[b_y I + c_y W^0(\sigma_\zeta)]^\pi_{x,\infty} = [a_y(x) I + b_y(x) W^0(\sigma_\zeta)]^\pi_{x,\infty} ,$$

and since the operator $a_y(x)I + b_y(x)W^0(\sigma_\zeta)$ (having constant coefficients) is Fredholm by Lemma 3.4 and Theorem 2.10(1) (for constant symbols), $[b_y I + c_y W^0(\sigma_\zeta)]^\pi_{x,\infty}$ must be invertible due to Theorem 4.2. Finally, if η is any point of $\mathbb{R}$, then

$$[b_y I + c_y W^0(\sigma_\zeta)]^\pi_{\infty,\eta} = [b_y(\infty) I + c_y(\infty) W^0(\sigma_\zeta)]^\pi_{\infty,\eta} ,$$

and combining Lemma 3.4, Theorem 2.10(1) (for constant symbols) and the "only if" part of Theorem 4.2 we conclude again that $[b_y I + c_y W^0(\sigma_\zeta)]^\pi_{\infty,\eta}$ is invertible

Hence it turns out that $[b_y I + c_y W^0(\sigma_\zeta)]^\pi_{x,\eta}$ is invertible for all $(x,\eta) \in T$, and so the "if" portion of Theorem 4.2 gives that $b_y I + c_y W^0(\sigma_\zeta)$ is Fredholm. This however contradicts Lemma 3.4 and Theorem 2.10(1), since $0 \in \mathcal{H}$.

(2) The element $[bI + cW^0(\sigma_\zeta)]^\pi_{\infty,\infty}$ is equal to

$$\begin{aligned} g = [&b(-\infty)\chi_- I + b(+\infty)\chi_+ I \\ &+ (c(-\infty)\chi_- + c(+\infty)\chi_+) W^0(\chi_- - \chi_+)]^\pi_{\infty,\infty} \end{aligned}$$

and hence belongs to the closed subalgebra $\mathfrak{C}$ of $\operatorname{Com}\mathfrak{B}/\mathfrak{Z}_{(\infty,\infty)}$ generated by

$$e = [I]^{\pi}_{\infty,\infty}, \quad r = [\chi_+ I]^{\pi}_{\infty,\infty}, \quad s = [W^0(\chi_+)]^{\pi}_{\infty,\infty}.$$

(notice that $\chi_+ = 1 - \chi_+$). Because $\phi W^0(\psi) - W^0(\psi)\phi I$ is compact on $L^p(\mathbb{R}, w)$ whenever $\phi \in C(\bar{\mathbb{R}})$ and $\psi \in C(\bar{\mathbb{R}}) \cap PC^p(w)$ (see *e.g.* [12, p. 93] for power weights) and there are such ϕ and ψ with

$$r = [\phi I]^{\pi}_{\infty,\infty}, \qquad s = [W^0(\psi)]^{\pi}_{\infty,\infty},$$

it follows that $\mathfrak{C}$ is commutative and that $r^2 = r$ and $s^2 = s$. Let M denote the maximal ideal space of $\mathfrak{C}$ and let $\Gamma : \mathfrak{C} \to C(M)$ stand for the Gelfand transform. The spectra of the idempotents r and s are subsets of $\{0,1\}$. For $j, k \in \{0,1\}$, put

$$M_{jk} = \{m \in M : (\Gamma r)(m) = j, \ (\Gamma s)(m) = k\}.$$

So $M = M_{00} \cup M_{01} \cup M_{10} \cup M_{11}$, and if $m \in M_{jk}$, then

$$(\Gamma g)(m) = b(-\infty)(1-j) + b(+\infty)j + (c(-\infty)(1-j) + c(+\infty)j)(1-k-k),$$

which is one of the four numbers

$$b(-\infty) \pm c(-\infty), \qquad b(+\infty) \pm c(+\infty).$$

Since $b \pm c \in GL^{\infty}(\mathbb{R})$, we obtain that g is invertible in $\mathfrak{C}$ and thus all the more in $\operatorname{Com}\mathfrak{B}/\mathfrak{Z}_{(\infty,\infty)}$.

(3) We now have

$$[bI + cW^0(\sigma_\zeta)]^{\pi}_{\infty,\eta} = [bI + \sigma_\zeta(\eta)cI]^{\pi}_{\infty,\eta},$$

and since $b + \sigma_\zeta(\eta)c$ is either $b - c$ or $b + c$, the multiplication operator $(b + \sigma_\zeta(\eta)c)I$ is invertible on $L^p(\mathbb{R}, w)$.

(4) Because

$$[bI + cW^0(\sigma_\zeta)]^{\pi}_{\infty,\zeta} = [b_\infty I + c_\infty W^0(\sigma_\zeta)]^{\pi}_{\infty,\zeta}$$

for any functions $b_\infty, c_\infty \in C(\bar{\mathbb{R}})$ such that

$$b_\infty(\pm\infty) = b(\pm\infty), \ c_\infty(\pm\infty) = c(\pm\infty), \ b_\infty \pm c_\infty \in GPC, \ d_{b_\infty, c_\infty} \neq 0,$$

for all $x \in \mathbb{R}$, it suffices to prove that $[b_\infty I + c_\infty W^0(\sigma_\zeta)]^\pi_{\infty,\zeta}$ is invertible if and only if 0 is not in the horn $\mathcal{H} \stackrel{\text{def}}{=} \mathcal{H}(\cdots)$.

If $0 \notin \mathcal{H}$, then $b_\infty I + c_\infty W^0(\sigma_\zeta)$ is Fredholm by Lemma 3.4 and Theorem 2.10(1) and hence $[b_\infty I + c_\infty W^0(\sigma_\zeta)]^\pi_{\infty,\zeta}$ is invertible by Theorem 4.2.

Conversely, assume $0 \in \mathcal{H}$ but $[b_\infty I + c_\infty W^0(\sigma_\zeta)]^\pi_{\infty,\zeta}$ is invertible. If $y \in \mathbb{R}$, then

$$[b_\infty I + c_\infty W^0(\sigma_\zeta)]^\pi_{y,\infty} = [b_\infty(y) + c_\infty(y) W^0(\sigma_\zeta)]^\pi_{y,\infty} ,$$

which is invertible by Lemma 3.4, Theorem 2.10(1) (with constant symbols), and Theorem 4.2. In case $\eta \in \dot{\mathbb{R}}\backslash\{\zeta\}$, we know that $[b_\infty I + c_\infty W^0(\sigma_\eta)]^\pi_{\infty,\eta}$ is invertible from the parts (2) and (3) we have already proved. Thus, $[b_\infty I + C_\infty W^0(\sigma_\zeta)]^\pi_{\infty,\eta}$ is invertible for all $(y,\eta) \in T$. From Theorem 4.2 we so infer that $b_\infty I + c_\infty W^0(\sigma_\zeta)$ is Fredholm, which contradicts Theorem 2.10(1), since $0 \in \mathcal{H}$.

5. Wiener-Hopf integral operators.

Lemma 5.1. *If $a \in PC^p(\omega) \cap GL^\infty(\mathbb{R})$ and $y, \eta \in \mathbb{R}$, then:*

(1) $[\chi_- I + \chi_+ W^0(a)]^\pi_{y,\infty}$ *is invertible if $y \neq 0$;*

(2) $[\chi_- I + \chi_+ W^0(a)]^\pi_{0,\infty}$ *is invertible if and only if*

$$0 \notin \mathcal{H}(a(+\infty), a(-\infty); \nu_0^-(p,w), \nu_0^+(p,w)) ;$$

(3) $[\chi_- I + \chi_+ W^0(a)]^\pi_{\infty,\infty}$ *is invertible;*

(4) $[\chi_- I + \chi_+ W^0(a)]^\pi_{\infty,\eta}$ *is invertible if and only if*

$$0 \notin \mathcal{H}(a(\eta - 0), a(\eta + 0); \nu_\infty^-(p,w), \nu_\infty^+(p,w)) .$$

PROOF. (1) The element $[\chi_- I + \chi_+ W^0(a)]^\pi_{y,\infty}$ is equal to $[I]^\pi_{y,\infty}$ for $y < 0$ and equal to

$$\begin{aligned} &[W^0(a(-\infty)\chi_- + a(+\infty)\chi_+)]^\pi_{y,\infty} \\ &\quad = \Big[\frac{a(-\infty) + a(+\infty)}{2} I + \frac{a(-\infty) - a(+\infty)}{2} W^0(\sigma)\Big]^\pi_{y,\infty} \\ &\quad = [bI + cW^0(\sigma)]^\pi_{y,\infty} \end{aligned}$$

for $y > 0$. It is clear that $[I]^{\pi}_{y,\infty}$ is invertible, and since

$$\frac{b+c}{b-c} = \frac{a(-\infty)}{a(+\infty)} \neq 0\,,$$

we deduce the invertility of $[\chi_- I + \chi_+ W^0(a)]^{\pi}_{y,\infty}$ from Lemma 4.6(1) (with constant b and c).

(2) The coset $[\chi_- I + \chi_+ W^0(a)]^{\pi}_{0,\infty}$ equals

$$\begin{aligned}
&[\chi_- I + \chi_+ W^0(a(-\infty)\chi_- + a(+\infty)\chi_+)]^{\pi}_{0,\infty} \\
&\quad = \Big[\chi_- I + \frac{a(-\infty)+a(+\infty)}{2} I + \frac{a(-\infty)-a(+\infty)}{2} W^0(\sigma)\Big]^{\pi}_{0,\infty} \\
&\quad = [bI + cW^0(\sigma)]^{\pi}_{0,\infty}\,.
\end{aligned}$$

We have

$$\Big(\frac{b+c}{b-c}\Big)(x) = 1 \quad \text{for } x < 0$$

and

$$\Big(\frac{b+c}{b-c}\Big)(x) = \frac{a(-\infty)}{a(+\infty)} \neq 0 \quad \text{for } x > 0\,.$$

Hence, Lemma 4.6(1) implies that $[\chi_- I + \chi_+ W^0(a)]^{\pi}_{0,\infty}$ is invertible if and only if

$$0 \notin \mathcal{H}\Big(1, \frac{a(-\infty)}{a(+\infty)}; \nu_0^-(p,w), \nu_0^+(p,w)\Big)\,,$$

which happens if and only if

$$0 \notin \mathcal{H}(a(+\infty), a(-\infty); \nu_0^-(p,w), \nu_0^+(p,w))\,.$$

(3) Because $[\chi_- I + \chi_+ W^0(a)]^{\pi}_{\infty,\infty}$ equals

$$[\chi_- I + \chi_+ W^0(a(-\infty)\chi_- + a(+\infty)\chi_+)]^{\pi}_{\infty,\infty} = [bI + cW^0(\sigma)]^{\pi}_{\infty,\infty}\,,$$

the assertion is immediate from Lemma 4.6(2).

(4) As in Section 3.3, let χ_η^- and χ_η^+ be the characteristic functions of $(-\infty, \eta)$ and $(\eta, +\infty)$, respectively. Then

$$\begin{aligned}
[\chi_- I + \chi_+ W^0(a)]^{\pi}_{\infty,\eta} &= [\chi_- I + \chi_+ W^0(a(\eta-0)\chi_\eta^- + a(\eta+0)\chi_\eta^+)]^{\pi}_{\infty,\eta} \\
&= \Big[\chi_- I + \frac{a(\eta-0)+a(\eta+0)}{2} I + \frac{a(\eta-0)-a(\eta+0)}{2} W^0(\sigma_\eta)\Big]^{\pi}_{\infty,\eta}
\end{aligned}$$

$$= [bI + cW^0(\sigma_\eta)]^\pi_{\infty,\eta} ,$$

and since

$$\frac{b+c}{b-c}(-\infty) = 1 , \qquad \frac{b+c}{b-c}(+\infty) = \frac{a(\eta-0)}{a(\eta+0)} ,$$

we obtain from Lemma 4.6(4) that $[\chi_- I + \chi_+ W^0(a)]^\pi_{\infty,\eta}$ is invertible if and only if

$$0 \notin \mathcal{H}\Big(\frac{a(\eta-0)}{a(\eta+0)}, 1; \nu^-_\infty(p,w), \nu^+_\infty(p,w)\Big) ,$$

which is equivalent to the condition that

$$0 \notin \mathcal{H}(a(\eta-0), a(\eta+0); \nu^-_\infty(p,w), \nu^+_\infty(p,w)) .$$

Here now is our main result

Theorem 5.2. *Let $w \in A_p$, $a \in PC^p(w)$, and define $\nu^\pm_0(p,w)$, $\nu^\pm_\infty(p,w)$ by Theorem* 2.10(1). *Then the essential spectrum of $W(a)$ on $L^p(\mathbb{R}_+, w)$ is*

$$a^{p,w} = \Big(\bigcup_{\eta\in\mathbb{R}} \mathcal{H}(a(\eta-0), a(\eta+0); \nu^-_\infty(p,w), \nu^+_\infty(p,w)) \Big) \bigcup \mathcal{H}(a(+\infty), a(-\infty); \nu^-_0(p,w), \nu^+_0(p,w)) .$$

If $0 \notin a^{p,w}$ then the index of $W(a)$ on $L^p(\mathbb{R}_+, w)$ equals $-\operatorname{wind} a^{p,w}$.

PROOF. We first show that the essential range of a is a subset of the essential spectrum of $W(a)$.

Let $\eta \in \dot{\mathbb{R}}$ be a point at which a is continuous and assume

$$W(a) - a(\eta)I = W(a - a(\eta))$$

is Fredholm on $L^p(\mathbb{R}_+, w)$. Then, by Section 3.3, the operator $\chi_- I + \chi_+ W^0(a - a(\eta))$ is also Fredholm on $L^p(\mathbb{R}, w)$ and consequently, by the "only if" part of Theorem 4.2,

$$[\chi_- I + \chi_+ W^0(a - a(\eta))]^\pi_{\infty,\eta} = [\chi_- I]^\pi_{\infty,\eta}$$

is invertible. This, however, is impossible because

$$[\chi_+ I]^\pi_{\infty,\eta}[\chi_- I]^\pi_{\infty,\eta} = [0]^\pi_{\infty,\eta}\,, \quad [\chi_+ I]^\pi_{\infty,\eta} \neq [0]^\pi_{\infty,\eta}\,.$$

Thus, the essential spectrum of $W(a)$ contains the values of a at all points at which it is continuous. Since these values are dense in the essential range of a and the essential spectrum of $W(a)$ is closed, it follows that the whole essential range is a subset of the essential spectrum.

We are now left with showing that if $a \in PC^p(w) \cap GL^\infty(\mathbb{R})$, then $\chi_- I + \chi_+ W^0(a)$ is Fredholm on $L^p(\mathbb{R}, w)$ if and only if $0 \notin a^{p,w}$; the index formula then follows by a standard homotopy argument from the case of continuous symbols.

But Theorem 4.2 in conjunction with Lemma 5.1 implies at once that if $a \in PC^p(w) \cap GL^\infty(\mathbb{R})$, then the Fredholmness of $W(a)$ is equivalent to the condition that $0 \notin a^{p,w}$.

5.3. CONCLUDING REMARK. Lemma 4.6 can also be used to gain interesting information about the Fredholmness of operators of the form

$$A = \sum_{j=1}^m b_j W^0(a_j)\,,$$

i.e. pseudodifferential operators with symbols

$$\sum_{j=1}^m b_j(x)a_j(\xi) \qquad b_j \in PC, a_j \in PC^p(w)$$

on $L^p(\mathbb{R}, w)$ (for the case $w \equiv 1$ see [4] and for power weights see [12] and [14]). We shall devote more space to this question in a forthcoming paper.

References.

[1] Böttcher, A., Krupnik, N., and Silbermann, B., A general look at local principles with special emphasis on the norm computation aspect. *Integral Equations Operator Theory* **11** (1988), 455-479.

[2] Böttcher, A. and Silbermann, B., *Analysis of Toeplitz operators.* Springer-Verlag, 1990.

[3] Böttcher, A. and Spitkovsky, I. M., Toeplitz operators with *PQC* symbols on weighted Hardy spaces. *J. Funct. Anal.* **97** (1991), 194-214.

[4] Duduchava, R., Wiener-Hopf integral operators with discontinuous symbols. *Soviet Math. Dokl.* **14** (1973), 1001-1005.

[5] Duduchava, R., *Integral equations with fixed singularities.* Teubner-Verlag, 1979.

[6] García-Cuerva, J. and Rubio de Francia, J. L., *Weighted norm inequalities and related topics.* North-Holland, 1985.

[7] Garnett, J. B., *Bounded analytic functions.* Academic-Press, 1981.

[8] Gohberg, I. and Krupnik, N., *Introduction to the theory of one-dimensional singular integral operators.* Shtiintsa, Kishinev, 1973 (Russian), (an extended English translation, containing also the result of [15], was edited by Birkhäuser Verlag in 1992).

[9] Hunt, R., Muckenhoupt, B. and Wheeden, R., Weighted norm inequalities for the conjugate function and Hilbert transform. *Trans. Amer. Math. Soc.* **176** (1973), 227-251.

[10] Koosis, P., *Introduction to H^p spaces.* Cambridge Univ. Press, 1980.

[11] Muckenhoupt, B., Weighted norm inequalities for the Hardy maximal function. *Trans. Amer. Math. Soc.* **165** (1972), 227-251.

[12] Roch, S. and Silbermann, B., *Algebras of convolution operators and their image in the Calkin algebra.* Report R-Math-05/90, Akad. Wiss. DDR, 1990.

[13] Sarason, D., *Function theory on the unit circle.* Virginia Polytechnic Inst. and State Univ., 1978.

[14] Schneider, R., Integral equations with piecewise continuous coefficients in L^p spaces with weight. *J. Integral Equations* **9** (1985), 135-152.

[15] Spitkovsky, I., Singular integral operators with PC symbols on the spaces with general weights. *J. Funct. Anal.* **105** (1992), 129-143.

[16] Widom, H., Singular integral equations in L^p. *Trans. Amer. Math. Soc.* **97** (1960), 131-160.

Recibido: 5 de marzo de 1.992

A. Böttcher
Fachbereich Mathematik
Technische Universität Chemnitz
O-9010 Chemnitz, GERMANY

I. M. Spitkovsky
Department of Mathematics
The College of William and Mary, Williamsburg
Virginia 23187-8795, U.S.A.

REVISTA MATEMÁTICA IBEROAMERICANA
VOL. 9, N.° 2, 1993

Interpolation between H^p Spaces and non-commutative generalizations, II

Gilles Pisier

Abstract. We continue an investigation started in a preceding paper. We discuss the classical results of Carleson connecting Carleson measures with the $\bar{\partial}$-equation in a slightly more abstract framework than usual. We also consider a more recent result of Peter Jones which shows the existence of a solution of the $\bar{\partial}$-equation, which satisfies simultaneously a good L_∞ estimate and a good L_1 estimate. This appears as a special case of our main result which can be stated as follows: Let $(\Omega, \mathcal{A}, \mu)$ be any measure space. Consider a bounded operator $u : H^1 \to L_1(\mu)$. Assume that on one hand u admits an extension $u_1 : L^1 \to L_1(\mu)$ bounded with norm C_1, and on the other hand that u admits an extension $u_\infty : L^\infty \to L_\infty(\mu)$ bounded with norm C_∞. Then u admits an extension $\tilde{u}$ which is bounded simultaneously from L^1 into $L_1(\mu)$ and from L^∞ into $L_\infty(\mu)$ and satisfies

$$\|\tilde{u}\colon\ L_\infty \to L_\infty(\mu)\| \le C\, C_\infty$$
$$\|\tilde{u}\colon\ L_1 \to L_1(\mu)\| \le C\, C_1$$

where C is a numerical constant.

Introduction.

We will denote by D the open unit disc of the complex plane, by $\mathbb{T}$ the unit circle and by m the normalized Lebesgue measure on $\mathbb{T}$. Let $0 < p \leq \infty$. We will denote simply by L_p the space $L_p(\mathbb{T}, m)$ and by H^p the classical Hardy space of analytic functions on D. It is well known that H^p can be identified with a closed subspace of L_p, namely the closure in L_p (for $p = \infty$ we must take the weak*-closure) of the linear span of the functions $\{e^{int} : n \geq 0\}$. More generally, when B is a Banach space, we denote by $L_p(B)$ the usual space of Bochner-p-integrable B-valued functions on $(\mathbb{T}, m)$, so that when $p < \infty$, $L_p \otimes B$ is dense in $L_p(B)$. We denote by $H^p(B)$ (and simply H^p if B is one dimensional) the Hardy space of B-valued analytic functions f such that $\sup_{r<1}(\int \|f(rz)\|^p \, dm(z))^{1/p} < \infty$. We denote

$$\|f\|_{H^p(B)} = \sup_{r<1} \left(\int \|f(rz)\|^p \, dm(z) \right)^{1/p}.$$

We refer to [G] and [GR] for more information on H^p-spaces and to [BS] and [BL] for more on real and complex interpolation.

We recall that a finite positive measure μ on D is called a Carleson measure if there is a constant C such that for any $r > 0$ and any real number θ, we have

$$\mu(\{z \in D : |z| > 1 - r, \ |\arg(z) - \theta| < r\}) \leq C\, r\,.$$

We will denote by $\|\mu\|_C$ the smallest constant C for which this holds. Carleson (see [G]) proved that, for each $0 < p < \infty$, this norm $\|\mu\|_C$ is equivalent to the smallest constant C' such that

$$\int |f|^p \, d\mu \leq C' \, \|f\|_{H^p}^p \,, \qquad \text{for all } f \in H^p \,. \tag{0.1}$$

Moreover, he proved that, for any $p > 1$ there is a constant A_p such that any harmonic function v on D admitting radial limits in $L_p(\mathbb{T}, m)$ satisfies

$$\int_D |v|^p d\mu \leq A_p \, \|\mu\|_C \int_{\mathbb{T}} |v|^p \, dm\,. \tag{0.2}$$

We observe in passing that a simple inner outer factorisation shows that if (0.1) holds for some $p > 0$ then it also holds for all $p > 0$ with the same constant.

It was observed a few years ago (by J. Bourgain [B], and also, I believe, by J. García-Cuerva) that Carleson's result extends to the Banach space valued case. More precisely, there is a numerical constant K such that, for any Banach space B, we have

$$\int \|f\|^p \, d\mu \leq K \, \|\mu\|_C \, \|f\|^p_{H^p(B)} \,, \tag{0.3}$$

for all $p > 0$ and $f \in H^p(B)$. Since any separable Banach space is isometric to a subspace of ℓ_∞, this reduces to the following fact. For any sequence $\{f_n : n \geq 1\}$ in H^p, we have

$$\int \sup_n |f_n|^p \, d\mu \leq K \, \|\mu\|_C \int \sup_n |f_n|^p \, dm \,. \tag{0.4}$$

This can also be deduced from the scalar case using a simple factorisation argument. Indeed, let F be the outer function such that $|F| = \sup_n |f_n|$ on the circle. Note that by the maximum principle we have $|F| \geq \sup_n |f_n|$ inside D, hence (0.1) implies

$$\int \sup_n |f_n|^p \, d\mu \leq \int |F|^p \, d\mu \leq C' \int |F|^p \, dm = \int \sup_n |f_n|^p \, dm \,.$$

This establishes (0.4) (and hence also (0.3)).

We wish to make a connection between Carleson measures and the following result due to Mireille Lévy [L]

Theorem 0.1. *Let S be any subspace of L_1 and let $u : S \to L_1(\mu)$ be an operator. Let C be a fixed constant. Then the following are equivalent:*

i) *For any sequence $\{f_n : n \geq 1\}$ in S, we have*

$$\int \sup_n |u(f_n)| \, d\mu \leq C \int \sup_n |f_n| \, dm \,.$$

ii) *The operator u admits an extension $\widetilde{u} : L_1 \to L_1(\mu)$ such that $\|\widetilde{u}\| \leq C$.*

PROOF. This theorem is a consequence of the Hahn Banach theorem in the same style as in the proof of Theorem 1 below. We merely sketch the proof of (i) implies (ii). Assume (i). Let $V \subset L_\infty(\mu)$ be the linear span of the simple functions (*i.e.* a function in V is a linear combination

of disjointly supported indicators). Consider the space $S \otimes V$ equipped with the norm induced by the space $L_1(m; L_\infty(\mu))$. Let $w = \sum_1^n \varphi_i \otimes f_i$ with $\varphi_i \in V$, $f_i \in S$. We will write

$$\langle u, w \rangle = \sum \langle \varphi_i, u f_i \rangle .$$

Then (i) equivalently means that for all such w

$$|\langle u, w \rangle| \leq C \, \|w\|_{L_1(m; L_\infty(\mu))} .$$

By the Hahn Banach Theorem, the linear form $w \to \langle u, w \rangle$ admits an extension of norm $\leq C$ on the whole of $L_1(m; L_\infty(\mu))$. This yields an extension of u from L_1 to $L_\infty(\mu)^* = L_1(\mu)^{**}$, with norm $\leq C$. Finally composing with the classical norm one projection from $L_1(\mu)^{**}$ to $L_1(\mu)$, we obtain (ii).

In particular, we obtain as a consequence the following (known) fact which we wish to emphasize for later use.

Proposition 0.2. *Let μ be a Carleson measure on D, then there is a bounded operator $T : L_1 \to L_1(\mu)$ such that $T(e^{int}) = z^n$ for all $n \geq 0$, or equivalently such that T induces the identity on H^1.*

PROOF. We simply apply Lévy's Theorem to H^1 viewed as a subspace of L_1, and to the operator $u : H^1 \to L_1(\mu)$ defined by $u(f) = f$. By (0.1) we have $\|u\| \leq K \, \|\mu\|_C$, but moreover by (0.4) and Lévy's Theorem there is an operator $T : L_1 \to L_1(\mu)$ extending u and with $\|T\| \leq K \, \|\mu\|_C$. This proves the proposition.

Allthough we have not seen this proposition stated explicitly, it is undoubtedly known to specialists (see the remarks below on the operator T^*). Of course, for $p > 1$ there is no problem, since in that case the inequality (0.2) shows that the operator of harmonic extension (given by the Poisson integral) is bounded from L_p into $L_p(\mu)$ and of course it induces the identity on H^p. However this same operator is well known to be unbounded if $p = 1$. The adjoint of the operator T appearing in Proposition 0.2 solves the $\bar{\partial}$-equation in the sense that for any φ in $L_\infty(\mu)$ the function $G = T^*(\varphi)$ satisfies $\|G\|_{L_\infty(m)} \leq \|T\| \, \|\varphi\|_\infty$ together with

$$\int f G \, dm = \int f \varphi \, d\mu , \qquad \text{for all } f \in H^1 ,$$

and by well known ideas of Hörmander [H] this means equivalently that $G\,dm$ is the boundary value (in the sense of [H]) of a distribution g on $\bar{D}$ such that $\bar{\partial} g = \varphi\mu$. In conclusion, we have

$$\bar{\partial} g = \varphi\mu \qquad \text{and} \qquad \|G\|_{L_\infty(m)} \leq K \|\mu\|_C \|\varphi\|_\infty .$$

This is precisely the basic L_∞-estimate for the $\bar{\partial}$-equation proved by Carleson to solve the corona problem, (*cf.* [G, Theorem 8.1.1, p. 320]). More recently, P. Jones [J] proved a refinement of this result by producing an explicit kernel which plays the role of the operator T^* in the above. He proved that one can produce a solution g of the equation $\bar{\partial} g = \varphi\mu$ which depends linearly on φ with a boundary value G satisfying simultaneously

$$\|G\|_{L_\infty(m)} \leq K \|\mu\|_C \|\varphi\|_\infty \qquad \text{and} \qquad \|G\|_{L_1(m)} \leq K \int |\varphi|\, d\mu ,$$

where K is a numerical constant. (Jones [J] mentions that A. Uchiyama found a different proof of this. A similar proof, using weights, was later found by S. Semmes.) Taking into account the previous remarks, our Theorem 1 below gives at the same time a different proof and a generalization of this theorem of Jones.

Our previous paper [P] contains simple direct proofs of several consequences of Jones' result for interpolation spaces between H^p-spaces. We will use similar ideas in this paper.

Let us recall here the definition of the K_t functional which is fundamental for the real interpolation method. Let A_0, A_1 be a compatible couple of Banach (or quasi-Banach) spaces. For all $x \in A_0 + A_1$ and for all $t > 0$, we let

$$K_t(x, A_0, A_1) = \inf \left\{ \|x_0\|_{A_0} + t\|x_1\|_{A_1} : x = x_0 + x_1, x_0 \in A_0, x_1 \in A_1 \right\}.$$

Let $S_0 \subset A_0$, $S_1 \subset A_1$ be closed subspaces. As in [P], we will say that the couple (S_0, S_1) is K-closed (relative to (A_0, A_1)) if there is a constant C such that for all $t > 0$ and $x \in S_0 + S_1$,

$$K_t(x, S_0, S_1) \leq C\, K_t(x, A_0, A_1) .$$

Main results.

Theorem 1. *Let $(\Omega, \mathcal{A}, \mu)$ be an arbitrary measure space. Let $u : H^\infty \to L_\infty(\mu)$ be a bounded operator with norm $\|u\| = C_\infty$. Assume that u is also bounded as an operator from H^1 into $L_1(\mu)$, moreover assume that there is a constant C_1 such that for all finite sequences $x_1, \ldots, x_n$ in H^1 we have*

$$\int \sup_i |u(x_i)|\, d\mu \le C_1 \int \sup |x_i|\, dm\,.$$

Then there is an operator $\tilde{u} : L^\infty \to L_\infty(\mu)$ which is also bounded from L^1 into $L_1(\mu)$ such that

$$\|\tilde{u} : L_\infty \to L_\infty(\mu)\| \le C\, C_\infty$$
$$\|\tilde{u} : L_1 \to L_1(\mu)\| \le C\, C_1$$

where C is a numerical constant.

PROOF. Let w be arbitrary in $L_\infty(\mu) \otimes H^\infty$. We introduce on $L_\infty(\mu) \otimes L^\infty(m)$ the following two norms for all w in $L_\infty(\mu) \otimes L^\infty(m)$

$$\|w\|_0 = \int \|w(\omega, \cdot)\|_{L^\infty(dm)}\, d\mu(\omega)\,,$$
$$\|w\|_1 = \int \|w(\cdot, t)\|_{L^\infty(d\mu)}\, dm(t)\,.$$

Let A_0 and A_1 be the completions of $L_\infty(\mu) \otimes L^\infty(m)$ for these two norms. (Note that A_0 and A_1 are nothing but respectively $L_1(d\mu; L_\infty(dm))$ and $L_1(dm; L_\infty(d\mu))$.) Let S_0 and S_1 be the closures of $L_\infty(\mu) \otimes H^\infty$ in A_0 and A_1 respectively.

The completion of the proof is an easy aplication (via the Hahn-Banach theorem) of the following result which is proved further below

Lemma 2. *(S_0, S_1) is K-closed.*

Indeed, assuming the lemma proved for the moment, fix $t > 0$, and consider w in $L_\infty(\mu) \otimes H^\infty$, we have (for some numerical constant C)

$$K_t(w, S_0, S_1) \le C\, K_t(w, A_0, A_1)\,, \qquad \text{for all } t > 0\,.$$

Recall that we denote by $V \subset L_\infty(\mu)$ the dense subspace of functions taking only finitely many values. Let $w = \sum_1^n \varphi_i \otimes f_i$ with $\varphi_i \in V$, $f_i \in H^\infty$. We will write, for every operator $u : H^1 \to L_1(\mu)$,

$$\langle u, w\rangle = \sum \langle \varphi_i, u f_i\rangle .$$

Clearly

$$\left\|\sum \varphi_i \otimes u(f_i)\right\|_{L^1_\mu(L^\infty_m)} \le C_\infty \|w\|_0 \tag{1}$$

and

$$\left\|\sum \varphi_i \otimes u(f_i)\right\|_{L^1_m(L^\infty_\mu)} \le C_1 \|w\|_1 \tag{2}$$

Moreover, by completion, we can extend (1) (respectively (2)) to the case when w is in S_0 (respectively, S_1). Hence, if $w = w_0 + w_1$ with $w_0 \in S_0$, $w_1 \in S_1$ we have by (1) and (2)

$$\begin{aligned} |\langle u, w\rangle| = \left|\sum \langle \varphi_i, u(f_i)\rangle\right| &\le C_\infty \|w_0\|_0 + C_1 \|w_1\|_1 \\ &\le C_\infty K_s(w, S_0, S_1) \\ &\le C\, C_\infty K_s(w, A_0, A_1) \end{aligned}$$

where $s = C_1(C_\infty)^{-1}$. By Hahn-Banach, there is a linear form ξ on $A_0 + A_1$ such that

$$\xi(w) = \langle u, w\rangle \qquad \text{for all } w \in S_0 + S_1 \tag{3}$$

and

$$|\xi(w)| \le C\, C_\infty K_s(w, A_0, A_1) \qquad \text{for all } w \in A_0 + A_1 .$$

Clearly this implies

$$|\xi(w)| \le C\, C_\infty \|w\|_0 , \qquad \text{for all } \omega \in A_0 , \tag{4}$$

$$|\xi(w)| \le C\, C_\infty s \|w\|_1 , \qquad \text{for all } \omega \in A_1 ,$$

$$\le C\, C_1 \|w\|_1 . \tag{5}$$

Now (4) implies for all $\varphi \in L_\infty(\mu)$ and for all $f \in L_\infty(dm)$

$$|\langle \xi, \varphi \otimes f\rangle| \le C\, C_\infty \|\varphi\|_1 \|f\|_\infty . \tag{6}$$

Define $\tilde{u} : L_\infty \to L_1(\mu)^* = L_\infty(\mu)$ as $\langle \tilde{u}(f), \varphi \rangle = \langle \xi, \varphi \otimes f \rangle$. Then, (6) implies

$$\|\tilde{u}(f)\|_{L_\infty(\mu)} \leq C\, C_\infty \|f\|_\infty ,$$

while (5) implies

$$|\langle \xi, \varphi \otimes f \rangle| \leq C\, C_1 \|\varphi\|_\infty \|f\|_1 ,$$

hence $\|\tilde{u}(f)\|_1 \leq C\, C_1 \|f\|_1$. Finally (3) implies that for all $f \in H^\infty$ and for all $\varphi \in L^\infty(\mu)$

$$\langle \tilde{u}(f), \varphi \rangle = \langle \varphi, u(f) \rangle$$

so that $\tilde{u}|_{H^\infty} = u$.

PROOF OF LEMMA 2. We start by reducing this lemma to the case when Ω is a finite set or equivalently, in case the σ-algebra $\mathcal{A}$ is generated by finitely many atoms, with a fixed constant independent of the number of atoms. Indeed, let V be the union of all spaces $L_\infty(\Omega, \mathcal{B}, \mu)$ over all the subalgebras $\mathcal{B} \subset \mathcal{A}$ which are generated by finitely many atoms. Assume the lemma known in that case with a fixed constant C independent of the number of atoms. It follows that for any w in $H^\infty \otimes V$ we have

$$K_t(w, S_0, S_1) \leq C\, K_t(w, A_0, A_1), \qquad \text{for all } t > 0 .$$

Since $H^\infty \otimes V$ is dense in $S_0 + S_1$, this is enough to imply Lemma 2.

Now, if $(\Omega, \mathcal{B}, \mu)$ is finitely atomic as above we argue exactly as in Section 1 in [P] using the simple (so-called) "square/dual/square" argument, as formalized in Lemma 3.2 in [P]. We want to treat by the same argument the pair

$$H^1(L_\infty(\mu)) \subset L^1(L_\infty(\mu)),$$
$$L_1(\mu; H^\infty) \subset L_1(\mu; L^\infty).$$

Taking square roots, the problem reduces to prove the following couple is K-closed

$$H^2(L_\infty(\mu)) \subset L^2(L_\infty(\mu)),$$
$$L_2(\mu; H^\infty) \subset L_2(\mu; L_\infty),$$

provided we can check that

$$(7) \qquad H^2(L_\infty(\mu)) \cdot L_2(\mu; H^\infty) \subset (H^1(L_\infty(\mu)), L_1(\mu; H^\infty))_{\frac{1}{2}\infty} .$$

We will check this auxiliary fact below. By duality and by Proposition 0.1 in [P] , we can reduce to checking the K-closedness for the couple

$$H^2(L_1(\mu)) \subset L^2(L_1(\mu)),$$
$$L_2(\mu; H^1) \subset L_2(\mu; L_1).$$

Taking square roots one more time this reduces to prove that the following couple is K-closed

$$\begin{cases} H^4(L_2(\mu)) \subset L^4(L_2(\mu)), \\ L_4(\mu; H^2) \subset L_4(\mu; L_2), \end{cases}$$

provided we have

$$(8) \qquad H^4(L_2(\mu)) \cdot L_4(\mu; H^2) \subset (H^2(L_1(\mu)), L_2(\mu; H^1))_{\frac{1}{2}\infty} .$$

But this last couple is trivially K-closed (with a fixed constant independent of $(\Omega, \mathcal{B}, \mu)$) because, by Marcel Riesz' Theorem, there is a simultaneously bounded projection

$$L_4(L_2(\mu)) \to H^4(L_2(\mu)),$$
$$L_4(\mu; L_2) \to L_4(\mu; H^2).$$

It remains to check the inclusions (7) and (8). We first check (7). By Jones' Theorem (see the beginning of Section 3 and Remark 1.12 in [P])

$$(9) \qquad H^2(L_\infty(\mu)) = (H^1(L_\infty(\mu)), H^\infty(L_\infty(\mu)))_{\frac{1}{2}2}$$

also by an entirely classical result (*cf.* [BL, p. 109])

$$(10) \qquad L_2(\mu; H^\infty) = (L_\infty(\mu; H^\infty), L_1(\mu; H^\infty))_{\frac{1}{2}2} .$$

By the bilinear interpolation theorem (*cf.* [BL, p. 76]) the two obvious inclusions

$$H^1(L_\infty(\mu)) \cdot L_\infty(\mu; H^\infty) \subset H^1(L_\infty(\mu)),$$
$$H^\infty(L_\infty(\mu)) \cdot L_1(\mu; H^\infty) \subset L_1(\mu; H^\infty),$$

(note that $H^\infty(L_\infty(\mu)) = L_\infty(\mu; H^\infty)$), imply that

$$(H^1(L_\infty(\mu)), H^\infty(L_\infty(\mu)))_{\frac{1}{2}2} \cdot (L_\infty(\mu; H^\infty), L_1(\mu; H^\infty))_{\frac{1}{2}2} \subset (H^1(L_\infty(\mu)), L_1(\mu; H^\infty))_{\frac{1}{2}\infty} .$$

Therefore, by (9) and (10), this proves (7). We now check (8). We will first prove an analogous result but with the inverses of all indices translated by $1/r$. More precisely, let $2 < r < \infty$, let p, r' be defined by the relations $1/2 = 1/r + 1/p$ and $1 = 1/r + 1/r'$. We will first check

$$H^{2p}(L_{2r'}(\mu)) \cdot L_{2p}(\mu; H^{2r'}) \subset (H^p(L_{r'}(\mu)), L_p(\mu; H^{r'}))_{\frac{1}{2}\infty} . \tag{11}$$

Indeed, we have

$$\begin{aligned} H^{2p}(L_{2r'}(\mu)) \cdot L_{2p}(\mu; H^{2r'}) &\subset L^{2p}(L_{2r'}(\mu)) \cdot L_{2p}(\mu; L^{2r'}) \\ &\subset (L^p(L_{r'}(\mu)), L_p(\mu; L^{r'}))_{\frac{1}{2}} . \end{aligned} \tag{12}$$

The last inclusion follows from a classical result on the complex interpolation of Banach lattices, (*cf.* [C, p. 125]). But now, since all indices appearing are between 1 and infinity, the orthogonal projection from L_2 onto H^2 defines an operator bounded simultaneously from $L^p(L_{r'}(\mu))$ into $H^p(L_{r'}(\mu))$ and from $L_p(\mu; L^{r'})$ into $L_p(\mu; H^{r'})$, hence also bounded from

$$(L^p(L_{r'}(\mu)), L_p(\mu; L^{r'}))_{\frac{1}{2}} \quad \text{into} \quad (H^p(L_{r'}(\mu)), L_p(\mu; H^{r'}))_{\frac{1}{2}} .$$

Since the latter space is included into $(H^p(L_{r'}(\mu)), L_p(\mu; H^{r'}))_{\frac{1}{2},\infty}$, (*cf.* [BL, p. 102]) we obtain the announced result (11).

Then, we use the easy fact that any element g in the unit ball of $H^4(L_2(\mu))$ (respectively, h in the unit ball of $L_4(\mu; H^2)$) can be written as $g = Gg_1$ (respectively, $h = Hh_1$) with G and H in the unit ball of $H^{2r}(L_{2r}(\mu)) = L_{2r}(\mu; H^{2r})$ and with g_1 (respectively, h_1) in the unit ball of $H^{2p}(L_{2r'}(\mu))$ (respectively, $L_{2p}(\mu; H^{2r'})$). Then, by (11), there is a constant C such that

$$\|g_1 h_1\|_{(H^p(L_{r'}(\mu)), L_p(\mu; H^{r'}))_{\frac{1}{2}\infty}} \leq C .$$

Now, the product $M = GH$ is in the unit ball of $H^r(L_r(\mu)) = L_r(\mu; H^r)$, therefore the operator of multiplication by M is of norm 1 both from $H^p(L_{r'}(\mu))$ into $H^2(L_1(\mu))$ and from $L_p(\mu; H^{r'})$ into $L_2(\mu; H^1)$. By interpolation, multiplication by M also has norm 1 from

$$(H^p(L_{r'}(\mu)), L_p(\mu; H^{r'}))_{\frac{1}{2}\infty} \quad \text{into} \quad (H^2(L_1(\mu)), L_2(\mu; H^1))_{\frac{1}{2}\infty} .$$

Hence, we conclude that $gh = Mg_1h_1$ has norm at most C in the space $(H^2(L_1(\mu)), L_2(\mu; H^1))_{\frac{1}{2}\infty}$. This concludes the proof of (8).

References.

[BL] Bergh, J., Löfström, J., *Interpolation spaces, an introduction.* Springer-Verlag, 1976.

[BS] Bennett, C., Sharpley, R., *Interpolation of operators.* Academic Press, 1988.

[B] Bourgain, J., On the similarity problem for polynomially bounded operators on Hilbert space. *Israel J. Math.* **54** (1986), 227-241.

[C] Calderón, A., Intermediate spaces and interpolation. *Studia Math.* **24** (1964), 113-190.

[G] Garnett, J., *Bounded Analytic Functions.* Academic Press, 1981.

[GR] García-Cuerva, J., Rubio de Francia, J. L., *Weighted norm inequalities and related topics.* North Holland, 1985.

[H] Hörmander, L., Generators for some rings of analytic functions. *Bull. Amer. Math. Soc.* **73** (1967), 943-949.

[J] Jones, P., L^∞ estimates for the $\bar{\partial}$-problem in a half plane. *Acta Math.* **150** (1983), 137-152.

[L] Lévy. M., Prolongement d'un opérateur d'un sous-espace de $L^1(\mu)$ dans $L^1(\nu)$. Séminaire d'Analyse Fonctionnelle 1979-1980, exposé 5. École Polytechnique. Palaiseau.

[P] Pisier, G., Interpolation between H^p spaces and non-commutative generalizations, I. *Pacific J. Math.* **155** (1992), 341-368.

Recibido: 20 de marzo de 1.992

Gilles Pisier*
Texas A. and M. University
College Station, TX 77843, U.S.A.
and
Université Paris 6
Equipe d'Analyse
75252 Paris Cedex 05, FRANCE

* Supported in part by N.S.F. grant DMS 9003550

Revista Matemática Iberoamericana
Vol. 9, N.° 2, 1993

Isopérimétrie pour les groupes et les variétés

Thierry Coulhon et Laurent Saloff-Coste

Introduction.

Dans cet article, nous proposons une approche très directe de différentes inégalités isopérimétriques. Cette approche est inspirée de la preuve élégante que donne Robinson de la décroissance du noyau de la chaleur sur les groupes de Lie, [20]. Dans le cadre des groupes unimodulaires moyennables, nous montrons dans la Section 1 comment des inégalités isopérimétriques optimales se déduisent simplement de la croissance du volume. Dans le cas de la croissance polynômiale, nous retrouvons ainsi les résultats de [26] (voir aussi [22] et [16]). Dans le cas de la croissance superpolynômiale, nous prouvons une conjecture de Varopoulos, [28]. Ces résultats sont optimaux (Section 2). Ils permettent de traiter par discrétisation les revêtements galoisiens de variétés compactes (Section 4).

Sur une variété riemannienne, en l'absence de structure de groupe, la croissance du volume est une information insuffisante pour déterminer l'isopérimétrie. Nous dégageons dans la Section 3 des hypothèses géométriques (inégalité de Poincaré, régularité dans la croissance du volume) qui permettent de déduire une inégalité isopérimétrique d'une minoration du volume. Les résultats connus en courbure de Ricci positive ou nulle apparaissent, grâce aux travaux de Buser [3], comme un cas particulier. Notre approche est aussi bien adaptée pour traiter de l'isopérimétrie dans le cadre des graphes (Section 5). Ces résultats

positifs sont à rapprocher des résultats négatifs de [9] (voir aussi [7]).

1. Groupes.

Nous allons d'abord nous placer dans le cadre des groupes discrets finiment engendrés, parce qu'il est intéressant par lui-même, qu'il permet de ne pas s'embarrasser de difficultés techniques, et enfin parce qu'on y voit apparaître des phénomènes de croissance intermédiaire du volume (ni polynômiale ni exponentielle) qui n'existent pas pour les groupes de Lie.

Soit G un groupe discret infini finiment engendré et $\{g_1, \dots, g_k\}$ un système de générateurs de G. Pour $n \in \mathbb{N}^*$, notons $B(n) = \{x \in G : \ x = g_{i_1}^{\varepsilon_1} \cdots g_{i_n}^{\varepsilon_n},\ i_1, \dots, i_n \in \{1, \dots, k\},\ \varepsilon_j = 0, \pm 1\}$. Si Ω est un sous-ensemble de G, $|\Omega|$ désignera son cardinal, et nous définirons son bord par

$$\partial\Omega = \left\{x \in \Omega : \text{ il existe } \ i \in \{1, \dots, k\} \text{ et } \varepsilon = \pm 1 \text{ tel que } x g_i^{\varepsilon} \notin \Omega\right\}.$$

Si f est une fonction à support fini sur G, nous définirons son gradient par

$$\nabla f(x) = \sum_{y \in G,\ y \sim x} |f(x) - f(y)| \ ,$$

où $x \sim y$ signifie qu'il existe un générateur g_i tel que $y = x g_i^{\pm 1}$. La quantité $\|\nabla f\|_1$ (les normes L^p sont bien sûr prises par rapport à la mesure de comptage) vaut donc $\sum_{x,y \in G, x \sim y} |f(x) - f(y)|$. On a clairement

$$|\partial\Omega| \leq \frac{1}{2} \|\nabla 1_\Omega\|_1 \leq 2\,k\,|\partial\Omega| \ .$$

La fonction de croissance du volume de G est $V(n) = |B(n)|$. Elle peut être polynômiale, *i.e.* il existe $D \in \mathbb{N}^*$ et $C > 0$ tels que

$$C^{-1}\, n^D \leq V(n) \leq C\, n^D \, .$$

D'après un théorème de Gromov, ceci se produit si et seulement si G contient un sous-groupe nilpotent d'indice fini. Elle peut être exponentielle, *i.e.* $V(n) \geq C e^{cn}$, par exemple si G est résoluble sans avoir de sous-groupe nilpotent d'indice fini, ou bien s'il est non moyennable. Elle peut enfin être intermédiaire, *i.e.*

$$C_1\, e^{c_1 n^\alpha} \leq V(n) \leq C_2\, e^{c_2 n^\beta} \ , \qquad 0 < \alpha \leq \beta < 1 \, ,$$

cf. [14]. Nous aurons à considérer la fonction $\phi : \mathbb{R}_+ \longrightarrow \mathbb{N}^*$ définie par $\phi(\lambda) = \inf\{n \in \mathbb{N}^* : V(n) > \lambda\}$.

Nous allons démontrer le

Théorème 1. *G vérifie l'inégalité isopérimétrique*

$$\frac{|\Omega|}{\phi(2\,|\Omega|)} \leq 8k\,|\partial\Omega|\,, \qquad \textit{pour tout } \Omega \subset G\,.$$

EXEMPLES. Si $V(n) \geq Cn^D$, alors

$$|\Omega|^{(D-1)/D} \leq C\,|\partial\Omega|\,,$$

ce qui est bien connu, au moins sous l'hypothèse $V(n) \simeq n^D$ (*cf.* [26]); notons toutefois que notre preuve est beaucoup plus simple que les approches déjà disponibles. Si $V(n) \geq Ce^{cn^\alpha}$, $0 < \alpha \leq 1$, alors

$$|\Omega|\,(\log|\Omega|)^{-1/\alpha} \leq C\,|\partial\Omega|\,,$$

pour $|\Omega| \geq 2$, ce qui était conjecturé dans [28].

Le théorème se déduit de la

Proposition. *Pour tout $\lambda > 0$, pour toute fonction f à support fini sur G,*

$$|\{|f| \geq \lambda\}| \leq \frac{2}{\lambda}\,\phi\Big(\frac{2}{\lambda}\,\|f\|_1\Big)\,\|\nabla f\|_1\,.$$

Il suffit pour le voir d'appliquer la proposition à $f = 1_\Omega$ et $\lambda = 1$.

PREUVE DE LA PROPOSITION. Si f est une fonction sur G et $n \in \mathbb{N}^*$, définissons f_n par

$$f_n(x) = \frac{1}{V(n)} \sum_{y \in B(n)} f(xy)\,.$$

Ecrivons

$$\Big|\big\{\,|f| \geq \lambda\big\}\Big| \leq \Big|\big\{\,|f - f_n| \geq \frac{\lambda}{2}\big\}\Big| + \Big|\big\{\,|f_n| \geq \frac{\lambda}{2}\big\}\Big|\,.$$

Il est clair que

$$\|f_n\|_\infty \leq V(n)^{-1}\,\|f\|_1\,.$$

Si n_0 est le plus petit entier tel que

$$V(n_0)^{-1} \|f\|_1 < \frac{\lambda}{2},$$

autrement dit si $n_0 = \phi((2/\lambda)\|f\|_1)$, on a

$$\left|\{|f_{n_0}| \geq \frac{\lambda}{2}\}\right| = 0$$

et

$$\left|\{|f| \geq \lambda\}\right| \leq \left|\{|f - f_{n_0}| \geq \frac{\lambda}{2}\}\right| \leq \frac{2}{\lambda} \|f - f_{n_0}\|_1 .$$

La preuve de la proposition est alors terminée grâce au

Lemme. *Pour toute fonction f à support fini sur G,*

$$\|f - f_n\|_1 \leq n \|\nabla f\|_1 .$$

Ce lemme est bien connu et sa preuve est simple. Définissons f^y par $f^y(x) = f(xy)$. Il est clair que, si y est un des générateurs $g_1, \ldots, g_k$ ou un de leurs inverses, $\|f - f^y\|_1 \leq \|\nabla f\|_1$. Par inégalité triangulaire, $\|f - f^y\|_1 \leq n \|\nabla f\|_1$ si $y \in B(n)$. Finalement

$$\begin{aligned} \|f - f_n\|_1 &\leq \frac{1}{V(n)} \sum_{x \in G} \sum_{y \in B(n)} |f(x) - f(xy)| \\ &= \frac{1}{V(n)} \sum_{y \in B(n)} \|f - f^y\|_1 \\ &\leq n \|\nabla f\|_1 . \end{aligned}$$

Le Théorème 1 peut, dans l'esprit de [24, Section 6], être reformulé en termes de plongement d'un espace de Sobolev dans un espace d'Orlicz. Ainsi, si $V(n) \geq Ce^{cn^\alpha}$, l'espace des fonctions à support fini complété par rapport à la norme $|||f||| = \|\nabla f\|_1$ se plonge dans l'espace d'Orlicz L_ψ, et même l'espace d'Orlicz-Lorentz $L_{\psi,1}$, où $\psi(x) = x(-\log x)^\alpha$.

Nous allons maintenant envisager le cas des groupes de Lie. Soit G un groupe de Lie réel connexe unimodulaire non compact et U un voisinage compact de l'origine dans G. La fonction de croissance du

volume de G est donnée par $V(n) = |U^n|$, où $|\cdot|$ désigne maintenant la mesure de Haar. Guivarc'h a montré dans [15] que deux situations seulement peuvent se produire: soit il existe $C > 0$ et $D \in \mathbb{N}^*$ tels que $C^{-1}n^D \leq V(n) \leq Cn^D$, et l'on dit que G est à croissance polynômiale d'exposant D, soit il existe $c, C > 0$ tels que $V(n) \geq Ce^{cn}$, et G est dit à croissance exponentielle. On a alors le

Théorème 2. *Soit G un groupe de Lie de dimension d, muni d'une métrique riemannienne invariante à gauche. Si G est à croissance polynômiale d'exposant D, on a l'inégalité isopérimétrique*

$$\sup \left\{ |\Omega|^{(d-1)/d}, |\Omega|^{(D-1)/D} \right\} \leq C\,|\partial\Omega|\,,$$

pour tout ouvert Ω à bord régulier de G, si $d \leq D$, et

$$\inf \left\{ |\Omega|^{(d-1)/d}, |\Omega|^{(D-1)/D} \right\} \leq C\,|\partial\Omega|\,,$$

pour tout ouvert Ω à bord régulier de G, si $d > D$. Si G est à croissance exponentielle, on a

$$\sup \left\{ |\Omega|^{(d-1)/d}, |\Omega|\,(\log(|\Omega| + 2))^{-1} \right\} \leq C\,|\partial\Omega|\,,$$

pour tout ouvert Ω à bord régulier de G.

REMARQUE. Si G est non-moyennable, on a

$$\sup \left\{ |\Omega|^{(d-1)/d}, |\Omega| \right\} \leq C\,|\partial\Omega|\,.$$

La preuve est identique à celle du cas discret. On est simplement amené à considérer aussi des boules de petit rayon t, pour lesquelles $V(t) \approx t^d$. Nous aurions pu, au lieu de munir G d'une structure riemannienne, le munir plus généralement d'une famille de champs de vecteurs invariants à gauche vérifiant la condition de Hörmander, comme dans [26]; l'exposant d n'est plus en général la dimension topologique: il est donné par la croissance des boules de petit rayon pour la distance associée aux champs et dépend donc des champs choisis. Dans ce contexte, la notion de volume du bord d'un ensemble et sa relation avec le gradient sont moins claires. Il suffit, pour contourner cette difficulté, d'énoncer des inégalités de Sobolev plutôt que des inégalités isopérimétriques.

Les considérations précédentes peuvent finalement s'étendre au cadre des groupes localement compacts à génération compacte, qui contient à la fois celui des groupes discrets finiment engendrés et celui des

groupes de Lie connexes. On retrouve et on généralise ainsi les résultats de [16, Section 7].

Une autre approche de l'isopérimétrie, précédemment développée par Varopoulos ([26], voir aussi [7, Section 2]), consiste à utiliser le semi-groupe de la chaleur dans le cas continu ou des puissances de convolution de mesures dans le cas discret. Ce type d'argument nécessite une estimation non-triviale des dérivées spatiales du noyau de la chaleur, *cf.* [22] (ou bien de leur analogue pour les puissances de convolution, *cf.* [16]). Ce qui nous permet ici d'éviter cette estimation, c'est l'idée, transposée de Robinson [21, Chap. IV, Section 2.a], d'opérer directement sur les convolutions par des fonctions caractéristiques de boules de rayon variable, plutôt que sur le semi-groupe de la chaleur (ou les puissances de convolution).

2. Décroissance des puissances de convolution et problèmes d'optimalité.

Nous nous plaçons à nouveau dans le cadre des groupes discrets. En faisant le lien avec les questions de décroissance des puissances de convolutions de probabilités traitées dans [29], nous allons vérifier que les inégalités isopérimétriques obtenues à la Section 1 sont optimales.

2.1. Groupes à croissance polynômiale.

Supposons que $V(n) \geq Cn^D$. Le Théorème 1 nous donne l'inégalité isopérimétrique $|\Omega|^{1/D} \leq C\,|\partial\Omega|^{1/(D-1)}$. Réciproquement, cette dernière a pour conséquence

$$V(n)^{(D-1)/D} \leq C\big(V(n+1) - V(n)\big)\,,$$

qui redonne $V(n) \geq Cn^D$. L'isopérimétrie et la minoration du volume correspondante sont donc équivalentes. On sait que chacune de ces deux propriétés entraîne à son tour le fait que, pour toute probabilité μ sur G à support générateur, $\mu^{(k)}(e) = O(k^{-D/2})$. On peut voir que l'isopérimétrie entraîne la décroissance des puissances de convolution en transitant par l'inégalité de Sobolev

$$\|f\|_{2D/(D-2)} \leq C\,\|\nabla f\|_2\ ,$$

puis en employant des arguments d'analyse fonctionnelle (*cf.* [26], [10] et [11]).

Mais on peut aussi partir directement de la minoration du volume et écrire, comme Robinson, et avec les notations de la Section 1,

$$\|f\|_2 \leq \|f - f_n\|_2 + \|f_n\|_2 \leq C\, n\, \|\nabla f\|_2 + C\, n^{-D/2}\, \|f\|_1 \ ,$$

et en déduire, en optimisant sur n,

$$\|f\|_2^{1+2/D} \leq C\, \|\nabla f\|_2\, \|f\|_1^{2/D} \ ,$$

qui suffit à nouveau à entraîner $\mu^{(k)}(e) = O(k^{-D/2})$ (*cf.* [4], ou [10, Théorème 5], ou encore [11, Proposition 2]).

Rappelons que, comme par ailleurs la décroissance des puissances de convolution, via l'inégalité de Sobolev L^2 à laquelle elle équivaut, entraîne la minoration du volume, il y a une équivalence complète entre $V(n) \geq n^D$, $|\Omega|^{1/D} \leq C\, |\partial\Omega|^{1/(D-1)}$ et $\mu^{(k)}(e) = O(k^{-D/2})$.

2.2. Groupes à croissance super-polynômiale.

Supposons $V(n) \geq Ce^{cn^\alpha}$. L'inégalité isopérimétrique obtenue, $|\Omega|\,(\log(|\Omega| + 2))^{-1/\alpha} \leq C\, |\partial\Omega|$, ne donne cette fois en retour que

$$V(n) \geq C\, e^{cn^{\alpha/(\alpha+1)}}$$

En revanche, partant de $V(n) \geq Ce^{cn^\alpha}$, on peut écrire

$$\|f\|_2 \leq C n\, \|\nabla f\|_2 + C\, e^{-cn^\alpha/2}\, \|f\|_1 \ ,$$

et en déduire que, pour μ probabilité à support générateur,

$$\mu^{(k)}(e) = O(e^{-ck^{\alpha/(\alpha+2)}})$$

(les détails sont donnés dans [29]); on retrouve ainsi les résultats de [28] (voir aussi [16]). Ces résultats sont optimaux, puisque sur tout groupe polycyclique à croissance exponentielle, c'est-à-dire tel que $\alpha = 1$, et pour toute probabilité ν à support fini et générateur sur ce groupe, $\nu^{(k)}(e) \approx O(e^{-ck^{1/3}})$, *cf.* [1].

Même si on ne peut pas le voir cette fois à travers des arguments de volume, l'inégalité isopérimétrique fournie par le Théorème 1 est

elle aussi optimale, au moins pour $\alpha = 1$: une inégalité sensiblement meilleure, du type

$$|\Omega| \, (\log(|\Omega| + 2))^{-1+\varepsilon} \leq C \, |\partial\Omega|$$

entraînerait une décroissance trop forte des puissances de convolution, en $e^{-ck^{1/(3-2\varepsilon)}}$ (voir [28, Section 5]).

Notons enfin que l'inégalité isopérimétrique extrême $|\Omega| \leq C \, |\partial\Omega|$ entraîne, elle, directement la croissance exponentielle $V(n) \geq e^{cn}$; mais elle ne vaut que sur les groupes non-moyennables. Sur ces groupes, et sur ces groupes seulement, les puissances de convolution de probabilités décroissent d'ailleurs en e^{-cn}.

3. Variétés.

Soit M une variété riemannienne complète connexe. Pour Ω un sous-ensemble mesurable de M, $|\Omega|$ désignera son volume riemannien, et s'il est suffisamment régulier, $|\partial\Omega|$ sera la mesure superficielle de son bord; $B(x,r)$ dénotera la boule riemannienne de centre x et de rayon r, et l'on notera $V(x,r) = |B(x,r)|$. Pour $f \in C_0^\infty(M)$, nous poserons

$$f_r(x) = \frac{1}{V(x,r)} \int_{B(x,r)} f(y) \, dy \, .$$

Supposons que $V(x,r) \geq V(r)$, pour toute $r > 0$, et pour toute $x \in M$, où V est une fonction strictement positive croissante; soit ϕ la fonction réciproque de V, définie par

$$\phi(\lambda) = \inf\{r \in \mathbb{R}_+^* \, : \, V(r) \geq \lambda\} \, .$$

Pour montrer, en utilisant les idées de la Section 1, qu'il existe $C > 0$ tel que, pour tout sous-ensemble Ω de M à bord régulier, $|\Omega| \, / \phi(C \, |\Omega|) \leq C \, |\partial\Omega|$, il suffit de supposer

$$(*) \qquad \|f - f_r\|_1 \leq C \, r \, \|\nabla f\|_1 \, , \quad \forall r > 0, \quad \forall f \in C_0^\infty(M) \, .$$

On écrit en effet à nouveau, pour $x \in M$, $r > 0$ et $f \in C_0^\infty(M)$,

$$\left|\{\, |f| \geq \lambda\}\right| \leq \left|\{\, |f - f_r| \geq \frac{\lambda}{2}\}\right| + \left|\{\, |f_r| > \frac{\lambda}{2}\}\right| .$$

Comme $|f_r(x)| \leq V(x,r)^{-1} \|f\|_1$, on a bien

$$\|f_r\|_\infty \leq V^{-1}(r) \|f\|_1 \ . \tag{1}$$

D'autre part,

$$\left|\{ |f - f_r| \geq \frac{\lambda}{2}\}\right| \leq \frac{2}{\lambda} \|f - f_r\|_1 \ .$$

On a alors, d'après $(*)$,

$$\left|\{ |f - f_r| \geq \frac{\lambda}{2}\}\right| \leq \frac{2Cr}{\lambda} \|\nabla f\|_1 \ . \tag{2}$$

Comme dans le cas des groupes, on déduit des estimations (1) et (2)

$$|\{|f| \geq \lambda\}| \leq \frac{2C}{\lambda} \phi(\frac{2}{\lambda} \|f\|_1) \|\nabla f\|_1 \ ,$$

d'où l'inégalité isopérimétrique annoncée.

Nous allons maintenant considérer des propriétés géométriques qui suffisent à entraîner la propriété $(*)$. Nous dirons que M vérifie l'inégalité de Poincaré à l'échelle sur les boules si

$$\int_{B(x,r)} |f(y) - f_r(x)| \, dy \leq Cr \int_{B(x,2r)} |\nabla f(y)| \, dy \, , \tag{P}$$

pour tout $x \in M$, $r > 0$, et que M a la propriété de doublement du volume si la quantité $V(x,2r)/V(x,r)$ est bornée indépendamment de $x \in M$ et de $r > 0$. On sait qu'une variété à courbure de Ricci positive ou nulle vérifie (P) (*cf.* [3]) et a la propriété de doublement du volume ([6, Proposition 4.1]). La propriété (P) peut être traduite en termes de constantes de Cheeger relatives aux boules de rayon r (*cf.* [18, p. 232]). Notons qu'une variété possédant la propriété de doublement du volume et à courbure de Ricci minorée est nécessairement à croissance polynômiale: il existe $D > 0$ tel que $V(x,r) \leq C r^D$, pour tout $x \in M$ et $r > 0$.

Lemme. *Si M vérifie* (P) *et a la propriété de doublement du volume, alors la propriété* $(*)$ *est vérifiée:*

$$\|f - f_r\|_1 \leq Cr \|\nabla f\|_1 \ ,$$

pour tout $r > 0$ et $f \in C_0^\infty(M)$.

PREUVE DU LEMME. Considérons dans M une famille maximale $(x_i)_{i\in I}$ de points à distance mutuelle supérieure ou égale à r. On a $M = \bigcup_{i\in I} B(x_i, r)$, et les boules $B(x_i, r/2)$ sont deux à deux disjointes.

La propriété de doublement du volume a classiquement pour conséquence le fait que le nombre de boules $B(x_i, 2r)$ dans lesquelles se trouve un point quelconque de M est borné. Soit en effet $x \in M$ et $r > 0$. On a, si l'on note $I_r(x) = \{i \in I : x \in B(x_i, 2r)\}$,

$$V(x,4r) \geq \sum_{I_r(x)} V\Big(x_i, \frac{r}{2}\Big) \geq K^{-4} \sum_{I_r(x)} V(x_i, 8r)$$
$$\geq K^{-4} \operatorname{Card} I_r(x)\, V(x,4r)$$

(dans ce qui précède, K est bien sûr la constante de doublement du volume; cette ligne de raisonnement est recopiée dans [17]).

On a alors

$$\begin{aligned}\|f - f_r\|_1 &\leq \sum_{i\in I} \int_{B(x_i,r)} |f(x) - f_r(x)|\, dx \\ &\leq \sum_{i\in I} \int_{B(x_i,r)} |f(x) - f_r(x_i)|\, dx \\ &\quad + \sum_{i\in I} \int_{B(x_i,r)} |f_r(x_i) - f_{2r}(x_i)|\, dx \\ &\quad + \sum_{i\in I} \int_{B(x_i,r)} |f_r(x) - f_{2r}(x_i)|\, dx\,.\end{aligned}$$

L'inégalité (P) donne pour le premier terme

$$\sum_{i\in I} \int_{B(x_i,r)} |f(x) - f_r(x_i)|\, dx \leq C\, r \sum_{i\in I} \int_{B(x_i,2r)} |\nabla f(x)|\, dx$$
$$\leq C\, K^4\, r\, \|\nabla f\|_1\ .$$

Le second terme vaut

$$\begin{aligned}V(x_i,r)\, |f_r(x_i) - f_{2r}(x_i)| &\leq \int_{B(x_i,r)} |f(y) - f_{2r}(x_i)|\, dy \\ &\leq \int_{B(x_i,2r)} |f(y) - f_{2r}(x_i)|\, dy \\ &\leq 2\, C\, r \int_{B(x_i,4r)} |\nabla f(x)|\, dx\,.\end{aligned}$$

Pour traiter le troisième terme, on écrit

$$\begin{aligned}\sum_{i\in I}\int_{B(x_i,r)} &|f_r(x)-f_{2r}(x_i)|\,dx \\ &\le \int_{B(x_i,r)}\frac{1}{V(x,r)}\int_{B(x,r)}|f(y)-f_{2r}(x_i)|\,dy\,dx \\ &\le \int_{B(x_i,r)}\frac{1}{V(x,r)}\int_{B(x_i,2r)}|f(y)-f_{2r}(x_i)|\,dy\,dx \\ &= \int_{B(x_i,2r)}|f(y)-f_{2r}(x_i)|\,dy\int_{B(x_i,r)}\frac{1}{V(x,r)}\,dx\,.\end{aligned}$$

Le premier facteur se majore à nouveau grâce à (P), et le second en remarquant que $B(x,4r)\supset B(x_i,r)$, et en utilisant le doublement du volume. Ceci termine la démonstration du lemme.

Nous avons démontré le

Théorème 3. *Si M vérifie* (P), *a la propriété de doublement du volume et si $V(r)=\inf_{x\in M}V(x,r)>0$ pour tout $r>0$, alors M vérifie l'inégalité isopérimétrique:*

il existe $C>0$ tel que, pour tout sous-ensemble Ω de M à bord régulier,

$$\frac{|\Omega|}{\phi(C\,|\Omega|)}\le C\,|\partial\Omega|\ ,$$

où ϕ est la fonction réciproque de V.

EXEMPLE. Soit M une variété riemannienne à courbure de Ricci positive ou nulle de dimension d. Les théorèmes de comparaison donnent

$$\frac{V(x,r)}{V(x,s)}\le\left(\frac{r}{s}\right)^d,\qquad \text{pour tout } x\in M \quad\text{et}\quad 0<s\le r\,.$$

En particulier, en faisant tendre s vers 0 dans l'inégalité précédente, on obtient $V(x,r)\le Cr^d, \forall x\in M, \forall r\ge 1$. Supposons que $V(x,r)\ge Cr^D$, pour r grand, avec $D\le d$. Le Théorème 4 permet d'affirmer que

$$\inf\left\{\,|\Omega|^{(d-1)/d}, |\Omega|^{(D-1)/D}\,\right\}\le C\,|\partial\Omega|\ ,$$

pour tout Ω ouvert à bord régulier de M. Ceci généralise le résultat pour $D=d$ annoncé dans [24, Section 5].

REMARQUE. Soit M une variété riemannienne de dimension d à courbure de Ricci minorée et telle que $V(x,1) \geq c > 0$ pour tout $x \in M$ (cette dernière condition est vérifiée par exemple si le rayon d'injectivité de M est strictement positif). Alors, avec les notations habituelles, nos techniques permettent de montrer que M vérifie l'inégalité isopérimétrique

$$\text{(IL)} \qquad |\Omega|^{(d-1)/d} \leq C\,|\partial\Omega| \qquad \text{pour tout } \Omega \text{ tel que } |\Omega| \leq 1\,.$$

On retrouve ainsi une inégalité de Sobolev locale contenue dans [27], voir aussi [23].

Dans le cas des variétés comme dans celui des groupes, une approche de l'isopérimétrie via le semi-groupe de la chaleur est possible ([27], voir aussi [7]). Les estimations des dérivées spatiales du noyau de la chaleur qu'elle nécessite sont alors fournies par les inégalités de Harnack paraboliques démontrées par Li et Yau dans [18]. Le théorème ci-dessus peut donc être vu comme un moyen de leur substituer les inégalités de Poincaré. Dans cette direction, voir aussi [23].

4. Revêtements.

Soit N une variété riemannienne compacte de dimension d et M un revêtement galoisien de N, de groupe Γ: Γ agit proprement et librement sur M par isométries et $M/\Gamma = N$. Le groupe Γ est finiment engendré. Réciproquement, tout groupe finiment engendré peut apparaître comme le groupe d'un revêtement co-compact. Soit V_1 la fonction de croissance du volume de Γ au sens du Section 1. Milnor a remarqué dans [20] qu'il existe $C > 0$ tel que, pour tout $x \in M$, pour tout $n \in \mathbb{N}^*$,

$$C^{-1}\, V_1(C^{-1}n) \leq V(x,n) \leq C\, V_1(Cn)\,.$$

Si M est à courbure de Ricci positive ou nulle, Γ est à croissance polynômiale de degré au plus d et si M est à courbure sectionnelle majorée par un nombre strictement négatif, Γ est à croissance exponentielle. Nous allons montrer que, comme dans le cas des groupes, l'isopérimétrie sur M est contrôlée par la croissance du volume.

Théorème 4. *Si M est un revêtement galoisien d'une variété compacte de dimension d et si V_1 est la fonction de croissance du volume du groupe du revêtement, on a les inégalités isopérimétriques*

$$\sup\left\{ |\Omega|^{(d-1)/d}, |\Omega|^{(D-1)/D} \right\} \leq C\,|\partial\Omega|\,, \quad \textit{si } V_1(n) \simeq n^D \textit{ et } d \leq D\,,$$
$$\inf\left\{ |\Omega|^{(d-1)/d}, |\Omega|^{(D-1)/D} \right\} \leq C\,|\partial\Omega|\,, \quad \textit{si } V_1(n) \simeq n^D \textit{ et } d > D\,,$$
$$\sup\left\{ |\Omega|^{(d-1)/d}, |\Omega|\,(\log(|\Omega|+2))^{-1/\alpha} \right\} \leq C\,|\partial\Omega|\,,$$

si $V_1(n) \geq e^{cn^\alpha}$, avec $0 < \alpha \leq 1$.

REMARQUE. Si Γ est non-moyennable, on montre

$$\sup\left\{ |\Omega|^{(d-1)/d}, |\Omega| \right\} \leq C\,|\partial\Omega|\,,$$

ce qui raffine le résultat de [2].

PREUVE DU THÉORÈME. Identifions Γ à une fibre Γx_0, pour x_0 fixé dans N; Γ est alors un discrétisé de M au sens de [17]. Nous reprenons la démarche et les notations de [8, Section III]: $\varepsilon > 0$ est tel que $d(x,y) \geq \varepsilon$, pour tout $x, y \in \Gamma, x \neq y$, et $d(x,\Gamma) \leq \varepsilon$, pour tout $x \in M$; $(\varphi_x)_{x\in X}$ est une partition de l'unité convenablement subordonnée aux boules $B(x, 2\varepsilon)$.

Soit maintenant une fonction $\psi \in C_0^\infty(M)$; $\tilde{\psi}$ est la fonction sur Γ définie par

$$\tilde{\psi}(x) = \frac{1}{|B(x,\varepsilon)|} \int_{B(x,\varepsilon)} \psi\, d\xi\,,$$

et S l'opérateur de $C_0^\infty(M)$ dans lui-même défini par

$$S\psi = \sum_{x\in X} \tilde{\psi}(x)\varphi_x\,.$$

Si ϕ_1 est définie par $\phi_1(\lambda) = \inf\{n \in \mathbb{N}^* : V_1(n) > \lambda\}$, la proposition de la Section 1 donne, pour tout $\lambda > 0$ et pour tout $\psi \in C_0^\infty(M)$,

$$|\{|\tilde{\psi}| \geq \lambda\}| \leq \frac{2}{\lambda}\,\phi_1\Big(\frac{2}{\lambda}\|\tilde{\psi}\|_1\Big)\,\|\nabla\tilde{\psi}\|_1\,.$$

Maintenant il est clair que $\|\tilde{\psi}\|_1 \leq C\,\|\psi\|_1$ et que $\|\nabla\tilde{\psi}\|_1 \leq C\,\|\nabla\psi\|_1$ (suivant le contexte, ∇ désigne le gradient riemannien sur M ou bien

le gradient discret sur Γ et les normes $\|\cdot\|_1$ sont prises par rapport à la mesure riemannienne sur M ou bien la mesure de comptage sur Γ; l'inégalité sur les gradients utilise une inégalité de Poincaré sur les boules de rayon ε, voir [8, Section III, Lemme 3]).

Soit N tel que, pour tout $y \in M$, il y ait au plus N fonctions φ_x non nulles en y. On a

$$|\{|S\psi| \geq N\lambda\}| \leq C\,|\{|\tilde{\psi}| > \lambda\}|$$

(ici, $|\cdot|$ désigne soit la mesure riemannienne, soit la mesure de comptage). En effet, si $K = \{x \in \Gamma : |\tilde{\psi}(x)| > \lambda\}$, notons $\tilde{K} = \{y \in M : \varphi_x(y) > 0 \text{ pour au moins un } x \in K\}$. Sur $\tilde{K}^c$, on a $|S\psi| \leq N\lambda$, donc

$$|\{|S\psi| \geq N\lambda\}| \leq |\tilde{K}| \leq |K|\,|B(x, 2\varepsilon)|\ .$$

Finalement,

$$|\{|S\psi| \geq \lambda\}| \leq \frac{C}{\lambda}\,\phi_1\Big(\frac{C}{\lambda}\,\|\psi\|_1\Big)\,\|\nabla\psi\|_1\ .$$

D'autre part, on remarque comme dans [8, Lemme 3] que $\|\psi - S\psi\|_1 \leq C\,\|\nabla\psi\|_1$, donc a fortiori que

$$|\{|\psi - S\psi| \geq \lambda\}| \leq \frac{C}{\lambda}\,\|\nabla\psi\|_1\ .$$

Il en résulte que

$$|\{|\psi| \geq \lambda\}| \leq \frac{C'}{\lambda}\,\phi_1\Big(\frac{C}{\lambda}\,\|\psi\|_1\Big)\,\|\nabla\psi\|_1$$

(on se souviendra que $\phi_1 \geq 1$). Jointe à l'inégalité isopérimétrique locale (IL) de la Section 3, cette inégalité donne le théorème.

5. Graphes.

Soit X un graphe dénombrable et connexe, tel que tout sommet ait au plus N voisins. Si x et y sont deux points de X, on note $d(x, y)$ le nombre de pas du plus court chemin joignant x à y; $x \sim y$ signifie que x et y sont voisins. Si Ω est un ensemble de sommets de X, $|\Omega|$ désignera le cardinal de Ω, et $\partial\Omega$ l'ensemble des points de Ω qui sont voisins d'au moins un point de Ω^c. Pour $x \in X$ et $n \in \mathbb{N}^*$, nous noterons $B(x, n) =$

$\{y \in X : d(x,y) \leq n\}$ et $V(x,n)$ le cardinal de $B(x,n)$. Si f est une fonction à support fini sur X, son gradient sera défini par $\nabla f(x) = \sum_{y \in X, y \sim x} |f(x) - f(y)|$. Comme à la Section 1, on a clairement

$$|\partial\Omega| \leq \frac{1}{2} \|\nabla 1_\Omega\|_1 \leq N\,|\partial\Omega| \,.$$

La moyenne de f sur $B(x,n)$ sera notée $f_n(x)$:

$$f_n(x) = \frac{1}{V(x,n)} \sum_{y \in B(x,n)} f(y) \,.$$

Soit $V(n) = \inf_{x \in X} V(x,n)$ et, pour $\lambda \in \mathbb{R}_+$, $\phi(\lambda) = \inf\{n \in \mathbb{N}^* : V(n) > \lambda\}$. Si la propriété

$$(*) \qquad \|f - f_n\|_1 \leq C\,n\,\|\nabla f\|_1 \,, \qquad \text{pour tout } \; n \in \mathbb{N}^* \,,$$

et pour toute fonction à support fini sur X, est vérifiée, on montre comme précédemment que

$$\frac{|\Omega|}{\phi(C\,|\Omega|)} \leq C\,|\partial\Omega| \,, \qquad \text{pour tout } \; \Omega \subset X \,.$$

En particulier, on a l'analogue du Théorème 3: nous dirons que X vérifie (P) si

$$\sum_{y \in B(x,n)} |f(y) - f_n(x)| \leq C\,n \sum_{y \in B(x,2n)} |\nabla f(y)| \,,$$

pour tout $x \in X$ et $n \in \mathbb{N}^*$, et que X a la propriété de doublement du volume si la quantité $V(x,2\,n)/V(x,n)$ est bornée indépendamment de $x \in X$ et de $n \in \mathbb{N}^*$. La propriété (P) s'exprime aussi en termes d'une constante de Cheeger relative aux boules et, jointe à la propriété de doublement du volume, elle entraîne $(*)$ comme à la Section 3.

Théorème 5. *Si X vérifie (P) et a la propriété de doublement du volume, alors X vérifie l'inégalité isopérimétrique:*

$$\textit{il existe } \; C > 0 \; \textit{tel que, pour tout } \; \Omega \subset X \,, \qquad \frac{|\Omega|}{\phi(C\,|\Omega|)} \leq C\,|\partial\Omega| \,.$$

Nous allons maintenant recourir, pour obtenir (P), à des idées développées par Diaconis et Stroock dans [13]. Soit $x \in X$ et $n \in \mathbb{N}^*$;

pour chaque couple de points y, z dans $B(x,n)$, choisissons un chemin minimisant $\gamma_{y,z}$ joignant y et z. Soit $\Gamma_{x,n} = \{\gamma_{y,z} : y,z \in B(x,n)\}$; il est facile de voir que les éléments de $\Gamma_{x,n}$ ne sortent pas de $B(x,2n)$. Si e est une arête orientée du graphe X, nous noterons e_+ son origine et e_- son extrémité. On a alors, pour $y,z \in B(x,n)$,

$$|f(y) - f(z)| \leq \sum_{e \in \gamma_{y,z}} |f(e_+) - f(e_-)|,$$

et donc

$$\sum_{y,z \in B(x,n)} |f(y) - f(z)| \leq \sum_{y,z \in B(x,n)} \sum_{e \in \gamma_{y,z}} |f(e_+) - f(e_-)|.$$

Mais le membre de gauche de l'inégalité précédente majore

$$V(x,n) \sum_{y \in B(x,n)} |f(y) - f_n(x)|,$$

et le membre de droite est majoré par

$$\sum_{e \in B(x,2n)} \#\{\gamma \in \Gamma_{x,n} : e \in \gamma\} \, |f(e_+) - f(e_-)|$$

(par $e \in B(x,2n)$, nous entendons $e_+, e_- \in B(x,2n)$). Si l'on pose

$$K(x,n) = V(x,n)^{-1} \max_{e \in B(x,2n)} \#\{\gamma \in \Gamma_{x,n} : e \in \gamma\},$$

on a donc

$$\sum_{y \in B(x,n)} |f(y) - f_n(x)| \leq K(x,n) \sum_{y \in B(x,2n)} |\nabla f(y)|.$$

L'inégalité (P) est donc vérifiée dès que $K(x,n) \leq C\,n$. Il est tentant d'appeler graphe à courbure positive ou nulle un graphe tel qu'il existe un choix des $\Gamma_{x,n}$ pour lequel $K(x,n)$ vérifie la majoration demandée. Nous ne connaissons pas, en dehors du cas des graphes de Cayley, de méthode générale pour prouver qu'un graphe est à courbure positive au sens précédent.

En fait, il est naturel d'appliquer les idées ci-dessus directement à la propriété $(*)$, ce qui nous permettra d'éviter l'hypothèse de doublement du volume. Soit $x \in X$ et $n \in \mathbb{N}^*$; pour chaque y dans

$B(x,n)$, choisissons un chemin minimisant $\gamma_{x,y}$ joignant y à x. Notons $\Gamma'_{x,n} = \{\gamma_{x,y} : y \in B(x,n)\}$.

Ecrivons

$$|f(x) - f_n(x)| \leq V(x,n)^{-1} \sum_{y \in B(x,n)} |f(x) - f(y)|$$
$$\leq V(x,n)^{-1} \sum_{\gamma \in \Gamma'_{x,n}} \sum_{e \in \gamma} |f(e_+) - f(e_-)| ,$$

et

$$\sum_{x \in X} |f(x) - f_n(x)| \leq \sum_{x \in X} V(x,n)^{-1} \sum_{y \in B(x,n)} |f(x) - f(y)|$$
$$\leq \sum_{x \in X} V(x,n)^{-1} \sum_{\gamma \in \Gamma'_{x,n}} \sum_{e \in \gamma} |f(e_+) - f(e_-)|$$
$$\leq \sum_{e \in E} \Big(\sum_{x \in B(e_+,n)} V(x,n)^{-1} \#\{\gamma \in \Gamma'_{x,n} | e \in \gamma\} \Big)$$
$$\cdot |f(e_+) - f(e_-)|$$
$$\leq K(n) \|\nabla f\|_1 ,$$

où E désigne l'ensemble des arêtes orientées de X et

$$K(n) = \max_{e \in E} \sum_{x \in X} V(x,n)^{-1} \#\{\gamma \in \Gamma'_{x,n} : e \in \gamma\} .$$

Nous avons démontré le

Théorème 6. *Si $K(n) \leq Cn$, alors:*

il existe $C > 0$ tel que, pour tout $\Omega \subset X$, $\dfrac{|\Omega|}{\phi(C\,|\Omega|)} \leq C\,|\partial\Omega| .$

REMARQUES. L'énoncé ci-dessus contient le Théorème 1. En effet, suivant les calculs de [12], on peut montrer que, lorsque X est le graphe de Cayley d'un groupe finiment engendré, on a $K(n) \leq Cn$, (et aussi d'ailleurs $K(x,n) \leq Cn$ si le groupe est à croissance polynômiale).

Une estimation du type $K(n) \leq Cn^\alpha$, $\alpha \geq 1$ entraîne

$$|\Omega| / \phi^\alpha(C\,|\Omega|) \leq C\,|\partial\Omega| .$$

Nous n'avons pas utilisé dans ce qui précède toute la souplesse de la méthode: en particulier, il suffit d'établir la propriété (*) en remplaçant les f_n par des moyennes sur des familles d'ensembles plus générales que des boules. Le lecteur intéressé construira facilement des graphes où (P) est fausse, mais qui peuvent être traités par ce moyen.

Nous allons maintenant laisser de côté les questions d'isopérimétrie et reprendre les idées précédentes, mais au niveau L^2, dans le but d'obtenir des informations sur la décroissance des marches aléatoires directement en termes de croissance du volume. Nous noterons $|\gamma|$ la longueur d'un chemin γ.

On a, par Cauchy-Schwarz,

$$|f(y) - f(z)|^2 \leq |\gamma_{y,z}| \sum_{e \in \gamma_{y,z}} |f(e_+) - f(e_-)|^2 ,$$

pour $y, z \in B(x,n)$. Il en résulte que

$$\sum_{y,z \in B(x,n)} |f(x) - f(y)|^2 \leq \sum_{y,z \in B(x,n)} |\gamma_{y,z}| \sum_{e \in \gamma_{y,z}} |f(e_+) - f(e_-)|^2 .$$

Le membre de gauche majore

$$V(x,n) \sum_{y \in B(x,n)} |f(y) - f_n(x)|^2 ,$$

et le membre de droite est majoré par

$$\sum_{e \in B(x,2n)} \sum_{\{\gamma \in \Gamma_{x,n} \,:\, e \in \gamma\}} |\gamma| \, |f(e_+) - f(e_-)|^2 .$$

Si l'on pose

$$K_2(x,n) = \max_{e \in B(x,2n)} \sum_{\{\gamma \in \Gamma_{x,n} \,:\, e \in \gamma\}} |\gamma| ,$$

on a donc

$$\sum_{y \in B(x,n)} |f(y) - f_n(x)|^2 \leq \frac{K_2(x,n)}{V(x,n)} \sum_{y \in B(x,2n)} |\nabla f(y)|^2 .$$

Mais $K_2(x,n)$ est toujours majoré par $2n\,V^2(x,n)$. Supposons maintenant que

$$C^{-1} n^D \leq V(x,n) \leq C n^D .$$

On a, d'après la majoration de $K_2(x,n)$,

$$\sum_{y\in B(x,n)} |f(y)-f_n(x)|^2 \le C\,n^{D+1} \sum_{y\in B(x,2n)} |\nabla f(y)|^2 .$$

Il est facile d'en déduire, puisque X a en particulier la propriété de doublement du volume, que

$$\|f-f_n\|_2 \le C\,n^{(D+1)/2}\,\|\nabla f\|_2 , \qquad \text{pour tout } n\in\mathbb{N}^*,$$

et pour tout fonction f à support fini sur X. Suivons maintenant la ligne de raisonnement esquissée dans le deuxième alinéa de la Section 2.2 ci-dessus

$$\|f\|_2 \le \|f-f_n\|_2 + \|f_n\|_2 \le C\,n^{(D+1)/2}\,\|\nabla f\|_2 + C\,n^{-D/2}\,\|f\|_1 ,$$

donc, en optimisant sur n,

$$\|f\|_2^{1+(D+1)/D} \le C\ \|\nabla f\|_2\ \|f\|_1^{(D+1)/D}$$

(on a utilisé cette fois la minoration du volume). En utilisant les méthodes de [4] ou [11], on en déduit une estimation sur la marche aléatoire standard sur X.

Théorème 7. *Soit X un graphe tel que $C^{-1}\,n^D \le V(x,n) \le C\,n^D$. Soit p_k le noyau de la marche aléatoire standard sur X à l'instant k. Alors*

$$\sup_{x,y\in X} p_k(x,y) = O(k^{-D/(D+1)}).$$

Ceci précise le résultat de Varopoulos, valable pour tout graphe infini, suivant lequel

$$\sup_{x,y\in X} p_k(x,y) = O(k^{-1/2})$$

(voir [4] et [11]). D'un autre côté, il est exclu d'obtenir en général mieux que

$$\sup_{x,y\in X} p_k(x,y) = O(k^{-1}),$$

ne serait-ce que parce que X peut être récurrent. Par discrétisation ([8, Théorème 3]), on peut déduire des considérations précédentes

Théorème 8. *Soit M une variété riemannienne à courbure de Ricci minorée, et telle que $C^{-1} r^D \leq V(x,r) \leq C r^D$, pour tout $r \geq 1$ et pour tout $x \in M$. Alors*

$$\sup_{x,y \in M} p_t(x,y) = O(t^{-D/(D+1)}), \quad t \to +\infty,$$

où p_t est le noyau de la chaleur sur M.

On rapprochera cet énoncé du Théorème 3 de [5].

Remerciement. Les auteurs tiennent à remercier Nicholas Varopoulos, qui, après avoir ouvert la voie, a attiré leur attention à maintes reprises sur le problème de l'isopérimétrie optimale pour les groupes à croissance super-polynômiale.

Bibliographie.

[1] Alexopoulos, G., A lower estimate for central probabilities on polycyclic groups, prépublication.

[2] Brooks, R., The fundamental group and the spectrum of the Laplacian. *Comm. Math. Helvetici* **56** (1981), 581-598.

[3] Buser, P., A note on the isoperimetric constant. *Ann. Sci. Ecole Norm. Sup.* **15** (1982), 213-230.

[4] Carlen, E., Kusuoka, S., Stroock, D., Upper bounds for symmetric Markov transition functions. *Ann. Inst. H. Poincaré, Probabilités et Statistique* **23** (1987), 245-287.

[5] Chavel, I., Feldman, E., Modified isoperimetric constants and large time heat diffusion in Riemannian manifolds. *Duke Math. J.* **64** (1991).

[6] Cheeger, J., Gromov, M., Taylor, M., Finite propagation speed, kernel estimates for functions of the Laplace operator and the geometry of complete Riemannian manifolds. *J. Differential Geom.* **17** (1982), 15-53.

[7] Coulhon, T., Sobolev inequalities on graphs and on manifolds, in *Harmonic Analysis and Discrete Potential Theory*, pp. 207-214, Picardello éd., Plenum Press, 1992.

[8] Coulhon, T., Noyau de la chaleur et discrétisation d'une variété riemannienne, à paraître dans *Isr. J. Math.*

[9] Coulhon, T., Ledoux, M. Isopérimétrie, décroissance du noyau de la chaleur et transformations de Riesz: un contre-exemple, à paraître dans *Arkiv. Mat.*

[10] Coulhon, T., Saloff-Coste, L., Puissances d'un opérateur régularisant. *Ann. Inst. H. Poincaré, Probabilités et Statistique* **26** (1990), 419-436.

[11] Coulhon, T., Saloff-Coste, L., Marches aléatoires non symétriques sur les groupes unimodulaires. *C. R. Acad. Sci. Paris* **310** (1990), 627-630.

[12] Diaconis, P., Saloff-Coste, L., Comparison theorems for reversible Markov chains, à paraître dans *Ann. Appl. Probab.*

[13] Diaconis, P., Stroock, D., Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Probab.* **1** (1991), 39-61.

[14] Grigorchuk, R., Degrees of growth of finitely generated groups and the theory of invariant means. *Math. U.S.S.R. Izvestiya* **25** (1985).

[15] Guivarc'h, Y., Croissance du volume et périodes des fonctions harmoniques. *Bull. Soc. Math. France* **101** (1973), 333-379.

[16] Hebisch, W., Saloff-Coste, L., Gaussian estimates for Markov chains and random walks on groups, à paraître dans *Ann Probab.*

[17] Kanai, M., Rough isometries and combinatorial approximations of geometries of non-compact riemannian manifolds. *J. Math. Soc. Japan* **37** (1985), 391-413.

[18] Li, P., Yau, S. Estimates of eigenvalues of a compact riemannian manifold. *Proc. Symp. Pure Math.* **36** (1980), 205-239.

[19] Li, P., Yau, S., On the parabolic kernel of the Schrödinger operator. *Acta Math.* **156** (1986), 153-201.

[20] Milnor, J., A note on curvature and fundamental group. *J. Differential Geom.* **2** (1968), 1-7.

[21] Robinson, D., *Elliptic operators and Lie groups.* Oxford Univ. Press, 1991.

[22] Saloff-Coste, L., Analyse sur les groupes de Lie à croissance polynômiale. *Arkiv Math.* **28** (1990), 315-331.

[23] Saloff-Coste, L., A note on Poincaré, Sobolev and Harnack inequalities. *Duke Math. J.* **65** (1992), 27-38.

[24] Varopoulos, N., Hardy-Littlewood theory for semigroups. *J. Funct. Anal.* **63** (1985), 240-260.

[25] Varopoulos, N., Random walks and Brownian motion on manifolds. *Symp. Math.* **XXIX** (1987), 97-109.

[26] Varopoulos, N., Analysis on Lie Groups. *J. Funct. Anal.* **76** (1988), 346-410.

[27] Varopoulos, N., Small time gaussian estimates of heat diffusion kernels. Part I: the semigroup technique. *Bull. Sc. Math.* **113** (1989), 253-277.

[28] Varopoulos, N., Groups of superpolynomial growth, in *Proceedings of the I.C.M. satellite conference on Harmonic Analysis*, Springer Verlag, 1991.

[29] Varopoulos, N., Saloff-Coste, L., Coulhon, T., *Analysis and Geometry on Groups*, Cambridge Tracts in Mathematics **100**, 1992.

Recibido: 10 de marzo de 1.992

Thierry Coulhon*
Département de Mathématiques
Université de Cergy-Pontoise
95806 Cergy Pontoise Cedex, FRANCE

and

Laurent Saloff-Coste
CNRS, Laboratoire Analyse Complexe et Géométrie
Université Paris VI
75252 Paris Cedex 05, FRANCE

* Le présent travail a été réalisé pendant un détachement du premier auteur au CNRS.

REVISTA MATEMÁTICA IBEROAMERICANA
VOL. 9, N.° 2, 1993

Initial traces of solutions to a one-phase Stefan problem in an infinite strip

D. Andreucci and M. K. Korten

Introduction.

The main result of this paper is an integral estimate valid for non-negative solutions (with no reference to initial data) $u \in L^1_{\mathrm{loc}}(\mathbb{R}^n \times (0,T))$ to

$$u_t - \Delta(u-1)_+ = 0\,, \qquad \text{in } \mathcal{D}'(\mathbb{R}^n \times (0,T))\,, \tag{0.1}$$

for $T > 0$, $n \geq 1$. Equation (0.1) is a formulation of a one-phase Stefan problem: in this connection u is the enthalpy, $(u-1)_+$ the temperature, and $u = 1$ the critical temperature of change of phase. Our estimate may be written in the form

$$\int_{\mathbb{R}^n} u(x,t) e^{-|x|^2/(2(T-t))}\, dx \leq C\,, \qquad 0 < t < T\,, \tag{0.2}$$

where C depends on n, T, t, u but it stays bounded as $t \to 0$.

Inequalities of this kind are well known in the case of diffusion equations

$$u_t - \Delta u^m = 0\,, \qquad m \geq 1\,, \qquad \text{in } \mathcal{D}'(\mathbb{R}^n \times (0,T))\,. \tag{0.3}$$

When $m = 1$ and (0.3) reduces to the standard heat equation, it goes back to Tichonov [18], Täcklind [17] and Widder [19] that the representation formula

$$(0.4)\quad u(x,t) = (4\pi)^{-n/2}(t-t_0)^{-n/2}\int_{\mathbb{R}^n} u(\xi,t_0)\, e^{-|\xi-x|^2/(4(t-t_0))}\, d\xi\ ,$$

$0 < t_0 < t < T$, $x \in \mathbb{R}^n$ actually holds for all nonnegative solutions to (0.3) with $m = 1$ (see also [14] and [3] for extensions to more general equations).

In the case of the porous media equation, *i.e.* (0.3) with $m > 1$, it was proved in [4] that

$$(0.5)\quad \begin{aligned} &\rho^{-(n+2/(m-1))}\int_{B_\rho(0)} u(x,t)dx \\ &\qquad \le c\left[T^{-1/(m-1)} + T^{n/2}\,\frac{u(0,3T/4)^{1+(m-1)n/2}}{\rho^{n+2/(m-1)}}\right], \end{aligned}$$

$0 < t < T/4$, is satisfied by all nonnegative solutions (see also [9] and [2] for extensions to more general equations).

The first consequence of estimates like (0.2), (0.4)-(0.5) is the fact that the growth of u as $|x| \to \infty$ cannot be arbitrary: indeed it must satisfy the restriction imposed by the corresponding inequality. We remark that the growth allowed by (0.2) is the same as the one given in (0.4) (*i.e.* roughly speaking, $u(\cdot,t) \sim e^{C(t)|x|^2}$ as $|x| \to \infty$) though the representation formula (0.4) obviously cannot hold for solutions to (0.1), and though the property of infinite speed of propagation does not hold for (0.1), contrarily to (0.3) for $m = 1$.

It can be easily shown that the growth predicted by (0.2) is actually optimal (see Section 2); in Section 3 we prove that solutions to (0.1) exist corresponding to arbitrary nonnegative locally integrable initial data satisfying (0.2). A by-product of this existence result is that the growth condition $u(x,t) \sim e^{C(t)|x|^2}$, $|x| \to \infty$, $t > 0$, is fulfilled in a pointwise sense (rather than in an integral sense as in (0.2): see Section 3).

A second consequence of (0.2) is the existence of a trace of u for $t = 0$ (the "initial trace"). This trace is -in general- a Radon measure and it is taken in the appropriate sense. It follows from the results proved here that the initial trace to a solution u to (0.1) belongs to $\mathcal{G}_{1/2T}$, where for $C > 0$ we define

$$(0.6)\qquad \mathcal{G}_C = \left\{\mu \text{ Radon measure in } \mathbb{R} : \int_{\mathbb{R}^n} e^{-C|x|^2} d\mu < +\infty\right\}.$$

The initial trace is actually unique (see Section 3). The technique of proof of (0.2) relies on a suitable procedure of approximation of u by compactly supported (in the space variables) solutions to (0.1) and on the use of the fundamental solution to the heat equation.

Another essential ingredient is the continuity of $(u-1)_+$: this follows from the results in [8], which in turn, may be applied since $u \in L^\infty_{\mathrm{loc}}(\mathbb{R}^n \times (0,T))$ (*cf.* [13]) and $(u-1)_+ \in W^{1,1}_{2,\mathrm{loc}}(\mathbb{R}^n \times (0,T))$ (see Section 2).

Also, a comparison result valid for solutions to (0.1) belonging to $\mathcal{G}_C$ is employed extensively; this follows from a generalisation of the results in [7] (see Remark 2.2).

We remark that the results found here carry over to more general equations of the type

$$u_t - L(u-1)_+ = 0\,,$$

where L is a linear elliptic operator, $\left(\dfrac{\partial}{\partial t} - L\right) z = 0$ having a fundamental solution (provided the solution u satisfies the local regularity assumptions quoted above). This follows from the proofs.

It is also clear that the assumption that u be nonnegative, can be relaxed to $u \geq -c$, $c > 0$. Indeed $v = u + c \geq 0$ fulfils

$$v_t - \Delta(v - c - 1)_+ = 0\,.$$

The paper is organised as follows: in Section 1 we recall some known facts and establish some further regularity results. In Section 2 we prove the inequality (0.2). In Section 3 we prove some consequences of inequality (0.2); namely, existence and uniqueness of the initial trace, and, conversely, existence and uniqueness of a nonnegative solution to the Cauchy problem for (0.1) taking an initial datum $u_0 \in L^1_{\mathrm{loc}}(\mathbb{R}^n) \cap \mathcal{G}_C$.

1. Regularity results.

In this section we summarize some known facts about integrability and local boundedness of solutions, and we also prove some regularity results, which we will need in Section 2.

The following sequence of results is obtained in [13] by following the methods of [10]. For future reference we will state them as a theorem.

Theorem 1.1. *Let $0 \le u \in L^1_{\mathrm{loc}}(\mathbb{R}^n \times (0,T))$ be a solution in the sense of $\mathcal{D}'(\mathbb{R}^n \times (0,T))$ of*

$$u_t = \Delta(u-1)_+ \,, \tag{1.1}$$

i.e.

$$\int_{\mathbb{R}^n} \int_0^T (u\varphi_t + (u-1)_+ \Delta\varphi)\, dx\, dt = 0\,,$$

for every $\varphi \in \mathcal{D}(\mathbb{R}^n \times (0,T))$. Then

i) *For any smooth bounded domain $D \subset \mathbb{R}^n$ and $0 < a < b < T$, there exist nonnegative measures ν_a, ν_b on D and μ on $\partial D \times [a,b]$ such that*

$$\begin{aligned} \int_D \psi(x,b) d\nu_b = &\int_D \psi(x,a)\, d\nu_a \\ &+ \iint_{D\times(a,b)} \left((u-1)_+ \Delta\psi + u\frac{\partial\psi}{\partial t}\right) dx\, dt \\ &+ \iint_{\partial D\times[a,b]} \frac{\partial\psi}{\partial n} d\mu \,, \end{aligned} \tag{1.2}$$

for any $\psi \in C_0^\infty(\mathbb{R}^n \times (0,T))$ such that $\psi\big|_{\partial D\times[a,b]} = 0$.
Here $\dfrac{\partial}{\partial n}$ denotes differentiation with respect to the inward unit normal to ∂D.

ii) *For a.e. t, $d\nu_t = u(x,t)\,dx$, $0 < t < T$. We remark that by considering a countable sequence of domains $\{D_m\}$ invading $\mathbb{R}^n$, ν_t may be taken independent of the domain D.*

iii) $u \in L^2_{\mathrm{loc}}(\mathbb{R}^n \times (0,T))$.

iv) *If for $\bar t > 0$ and $E \subset \mathbb{R}^n$ measurable, $|E| > 0$, $u(y,\bar t) < 1$ a.e. in E, then $u(x,t) \le u(x,\bar t)$ for a.e. $x \in E$ and $0 < t < \bar t$ (this was found independently also in* [1]).

v) (Comparison) *If $0 \le v \in L^1_{\mathrm{loc}}(\mathbb{R}^n \times (0,T))$ is a solution to* (1.1) *such that the traces of u and v on the parabolic boundary of $D \times (\bar t, t)$ (D a smooth bounded domain in $\mathbb{R}^n$, $\bar t > 0$) are ordered, the same order holds for u and v a.e. in $D \times (\bar t, t)$.*

vi) $(u-1)_+$ *satisfies*

$$\Delta(u-1)_+ - \frac{\partial}{\partial t}(u-1)_+ \ge 0 \qquad \text{in } \mathcal{D}'(\mathbb{R}^n \times (0,T))\,.$$

vii) $u \in L^{\infty}_{\mathrm{loc}}(\mathbb{R}^n \times (0,T))$.

REMARK. In fact Theorem 1.1 applies to local solutions of (1.1) defined in $\Omega \times (0,T)$, $\Omega \subset \mathbb{R}^n$, with the obvious modifications.

Lemma 1.2. *For any solution $0 \le u$ to* (1.1) *in the sense made precise in Theorem* 1.1, $(u-1)_+$ *belongs to* $W_2^{1,1}(K)$, *for any compact* $K \subset \mathbb{R}^n \times (0,T)$.

PROOF. For the sake of notational simplicity, we assume that u is a solution to (1.1) in $\mathbb{R}^n \times (0, T+\delta)$, $\delta > 0$. Defining u and $(u-1)_+$ as zero for $t < 0$, we let

$$u_m(x,t) = \int_{-\infty}^{\infty} \int_{\mathbb{R}^n} u(y,s)\rho_m(x-y)\tau_m(t-s)\,dy\,ds\,,$$

and analogously we set $w_m = (u-1)_+ * \rho_m\tau_m$, where ρ_m and τ_m are the usual (compactly supported) mollifiers. Therefore we have by (1.1)

$$\frac{\partial}{\partial t}u_m(x,t) - \Delta w_m(x,t) = 0\,, \qquad \text{in } B_R \times (t_0,T)\,, \tag{1.3}$$

where $R > 0$ and $T > t_0 > 0$ are arbitrarily fixed, and m is large enough.

Let $\alpha(s) = (s-1)_+$, and $\alpha_m(s)$ a smooth regularization of α such that $\alpha_m(s) = \alpha(s)$ for $s \ge 1 + 1/m$ and $\alpha_m'(s) > 0$ for $s \ge 0$. Define v_m as the solution to

$$\begin{cases} v_t = \Delta\alpha_m(v) & \text{in } B_R \times (t_0,T)\,, \\ v(x,t_0) = u_m(x,t_0) & \text{for } x \in B_R\,, \\ \alpha_m(v) = w_m & \text{on } \partial B_R \times (t_0,T)\,. \end{cases} \tag{1.4}$$

By standard calculations (see *e.g.* [6]) it follows that for any $K \subset B_R \times (t_0,T)$, K compact,

$$\begin{aligned} \|v_m\|_{L^{\infty}(B_R \times (t_0,T))} &\le C(M)\,, \\ \|\nabla\alpha_m(v_m)\|_{L^2(K)} + \|\alpha_m(v_m)_t\|_{L^2(K)} &\le C(M,K)\,, \end{aligned}$$

where $M = \|u\|_{L^{\infty}(B_{2R} \times (t_0/2,T))} < \infty$.

Therefore, choosing a subsequence of $\{v_m\}$, which we denote by $\{v_m\}$ again, we may prove that a $v \in L^\infty(B_R \times (t_0, T))$ exists such that

$$
\begin{aligned}
v_m &\to v && \text{weakly in } L^2(B_R \times (t_0,T)) , \\
\alpha_m(v_m) &\to (v-1)_+ && \text{a.e. in } B_R \times (t_0,T) , \\
\nabla\alpha_m(v_m) &\to \nabla(v-1)_+ && \text{weakly in } L^2_{\mathrm{loc}}(B_R \times (t_0,T)) , \\
\frac{\partial}{\partial t}\alpha_m(v_m) &\to \frac{\partial}{\partial t}(v-1)_+ && \text{weakly in } L^2_{\mathrm{loc}}(B_R \times (t_0,T)) .
\end{aligned} \tag{1.5}
$$

Next we show that $(u-1)_+ \equiv (v-1)_+$ in $B_R \times (t_0, T)$. We follow [16], and introduce the smooth function

$$
\psi_m(x,t) = \int_t^T (\alpha_m(v_m) - w_m)(x,\tau)\, d\tau .
$$

We subtract the first equation of (1.4) from (1.3), multiply by ψ_m and integrate by parts over $B_R \times (t_0, T)$, finding

$$
\begin{aligned}
&\int_{t_0}^T \int_{B_R} (v_m - u_m)(\alpha_m(v_m) - w_m)\, dx\, dt \\
&\qquad = -\int_{t_0}^T \int_{B_R} \nabla(\alpha_m(v_m) - w_m) \\
&\qquad\qquad \cdot \int_t^T \nabla(\alpha_m(v_m) - w_m)(x,\tau)\, d\tau\, dx\, dt \\
&\qquad = \frac{1}{2}\int_{t_0}^T \int_{B_R} \frac{\partial}{\partial t} \Big| \int_t^T \nabla(\alpha_m(v_m) - w_m) d\tau \Big|^2 dx\, dt \\
&\qquad = -\frac{1}{2}\int_{B_R} \Big| \int_{t_0}^T \nabla(\alpha_m(v_m) - w_m) d\tau \Big|^2 dx \le 0 .
\end{aligned} \tag{1.6}
$$

Then we let $m \to \infty$ in (1.6) to get, taking into account (1.5),

$$
\int_{t_0}^T \int_{B_R} (v-u)((v-1)_+ - (u-1)_+)\, dx\, d\tau \le 0 . \tag{1.7}
$$

Since

$$
(v-u)((v-1)_+ - (u-1)_+) \ge [(v-1)_+ - (u-1)_+]^2 \ge 0 ,
$$

(1.7) implies $(u-1)_+ \equiv (v-1)_+$, and due to (1.5), the proof is completed.

REMARK. It follows from the proof above that Lemma 1.2 applies to any nonnegative distributional solution of the Stefan problem (1.1), defined in a cylinder $B_R \times (0,T)$, since such solutions are locally bounded, according to Theorem 1.1.

Corollary 1.3. *Let* $0 \le u \in L^1_{\mathrm{loc}}(\mathbb{R}^n \times (0,T))$ *be a (distributional) solution to* (1.1). *Then* $(u-1)_+$ *is continuous.*

Corollary 1.3 is a consequence of the results of [8] (which apply to solutions with first derivatives in L^2) and of Lemma 1.2, and it holds in fact for local solutions as pointed out in the remark above.

A useful consequence of the regularity results given above is the following

Lemma 1.4. *Let* u *be as in Corollary* 1.3. *For all* $R > 0$, $0 < \varepsilon < T/2$,

$$\operatorname*{ess\,sup}_{|t-t_0|<\delta;\ 0<\varepsilon<t_0\,,\,t<T-\varepsilon} \int_{B_R} |u(x,t) - u(x,t_0)|\,dx \to 0 \qquad \text{as} \quad \delta \downarrow 0\,.$$

PROOF. We define $h = u - (u-1)_+$; because of Theorem 1.1. iv) we have for a.e. $t_0, t \in (0,T)$, $t > t_0$,

$$\begin{aligned}\int_{B_R} |u(x,t) - u(x,t_0)|\,dx &\le \int_{B_R} |(u-1)_+(x,t) - (u-1)_+(x,t_0)|\,dx \\ &\quad + \int_{B_R} (h(x,t) - h(x,t_0))\,dx \\ &\equiv A(t_0,t) + B(t_0,t)\,.\end{aligned}$$

Note that $A(\cdot,\cdot)$ is continuous, owing to Corollary 1.3; then $A(t_0,t) \to 0$ as $t \downarrow t_0$. Also

$$B(t_0,t) \le \int_{B_{2R}} \psi(x)(h(x,t) - h(x,t_0))\,dx\,,$$

for a nonnegative $\psi \in C_0^\infty(\mathbb{R}^n)$, $\psi(x) \equiv 1$ for $|x| \le R$ and $\psi(x) \equiv 0$ for $|x| > 2R$. Employing (1.2) and the local integrability of u, the claim follows.

REMARK. In the following we say that a family $\{\mu_\varepsilon\}_{\varepsilon>0}$ of Radon measures belongs *uniformly* to $\mathcal{G}_c$ if

$$\int_{\mathbb{R}^n} e^{-c|x|^2} d\mu_\varepsilon \le M < \infty, \qquad \text{for all } \varepsilon > 0,$$

for a constant M independent of ε.

2. The main estimate.

This section will be devoted to the proof of our main result (see (2.1) below), enabling us to identify the growth at infinity of any (distributional, $L^1_{\mathrm{loc}}(\mathbb{R}^n \times (0,T))$) nonnegative solution to (1.1). Therefore we can identify the natural class to which such solutions belong, without any a priori requirement on initial values. We want to single out that, as a consequence of this result, even though compactly supported solutions to (1.1) propagate with finite speed, solutions of (1.1) cannot grow at infinity faster than solutions to the heat equation.

Theorem 2.1. *Let $0 \le u \in L^1_{\mathrm{loc}}(\mathbb{R}^n \times (0,T))$, be a solution to (1.1) in the sense of $\mathcal{D}'(\mathbb{R}^n \times (0,T))$. Then for any $0 < \beta < 1$, there exists a constant $M = M(u, n, T, \beta)$ such that*

$$\int u(x,t) e^{-|x|^2/(4\beta(T-t))}\, dx \le M, \tag{2.1}$$

for any $0 < t < T/2$.

Moreover, this inequality is optimal, i.e. there exist solutions actually exhibiting the maximal growth allowed by (2.1).

REMARK. The constant M in (2.1) actually may be assumed to depend on u only through a bound for $|u|$ over $B_1(x_1) \times (T/2, 3T/4)$, through $|x_1|$, and $u(x_1, 3T/4)$. Here $x_1 \in \mathbb{R}^n$ is chosen such that $u(x_1, 3T/4) > 1$. This follows from the proof below, and from the results in [8] on the modulus of continuity of $(u-1)_+$.

PROOF. For the sake of notational simplicity, we may assume that u is defined in $\mathbb{R}^n \times (0, T+\lambda)$, for some $\lambda > 0$. Choose $0 < t_0 < T$. For any $\rho \ge 1$, define v_ρ as the (compactly supported) solution to

$$\begin{cases} \dfrac{\partial}{\partial t} v_\rho - \Delta (v_\rho - 1)_+ = 0, & \text{in } \mathcal{D}'(\mathbb{R}^n \times (t_0, T+\lambda)), \\ \| v_\rho(\cdot, t) - u(\cdot, t_0) \chi_{B_\rho}(\cdot) \|_{L^1(\mathbb{R}^n)} \to 0, & t \downarrow t_0. \end{cases} \tag{2.2}$$

The existence of v_ρ is provided *e.g.* by semigroup arguments (see [5] and references given therein).

We remark that as a consequence of Theorem 1.1.v),

$$\rho_1 \geq \rho_2 \text{ implies } v_{\rho_1} \geq v_{\rho_2}\,, \qquad \text{a.e. in } \mathbb{R}^n \times (t_0, T+\lambda)\,, \tag{2.3}$$

$$v_\rho \leq u\,, \qquad \text{a.e. in } \mathbb{R}^n \times (t_0, T+\lambda)\,. \tag{2.4}$$

Hence, due to monotonicity (implied by (2.3)) and local boundedness (implied by (2.4) and vii) of Theorem 1.1) there exists

$$u[t_0] = \lim_{\rho\to\infty} v_\rho \qquad \text{in } \mathbb{R}^n \times (t_0, T+\lambda)\,,$$

and $u[t_0]$ solves

$$w_t - \Delta(w-1)_+ = 0 \qquad \text{in } \mathcal{D}'(\mathbb{R}^n \times (t_0, T+\lambda))\,, \tag{2.5.a}$$

$$w(x,t) \to u(x,t_0)\,, \qquad \text{in } L^1_{\mathrm{loc}}(\mathbb{R}^n)\,, \qquad \text{as } t \downarrow t_0\,, \tag{2.5.b}$$

$$u[t_0] \leq u\,, \qquad \text{a.e. in } \mathbb{R}^n \times (t_0, T+\lambda)\,. \tag{2.6}$$

Equation (2.5.a) is obviously implied by the definition of $u[t_0]$. The convergence of $u[t_0](\cdot,t) \to u(\cdot,t_0)$ in $L^1_{\mathrm{loc}}(\mathbb{R}^n)$ follows by

$$v_\rho(x,t) - u(x,t_0) \leq u[t_0](x,t) - u(x,t_0) \leq u(x,t) - u(x,t_0)\,,$$

for all $\rho \geq 1$, when we take into account Lemma 1.3 and (2.2). Alternatively, the second of (2.5) may be derived subtracting the weak formulations of (2.2) and (2.5.a), and using again $u[t_0] \geq v_\rho$, with a suitable choice of the test functions.

Assume first that a point $P_0 \equiv (x_0, T)$ exists such that $u[t_0](P_0) > 1$: then we may find an $\varepsilon > 0$ such that

$$u[t_0](x,t) \geq 1 + 2\varepsilon\,, \quad \text{for all } (x,t) \in B_\varepsilon(x_0) \times (T-\varepsilon\,,\ T)\,. \tag{2.7}$$

Here we are using the continuity of $(u[t_0](x,t) - 1)_+$, provided by [8] and Lemma 1.2. We may assume that for $\rho \geq \rho_0$ large enough,

$$v_\rho(x,t) \geq 1 + \varepsilon\,, \quad \text{for all } (x,t) \in B_\varepsilon(x_0) \times (T-\varepsilon\,,\ T)\,. \tag{2.8}$$

Then, since $v_\rho(\cdot,t)$ is compactly supported in $\mathbb{R}^n$, we have for all $t_0 \leq t < T$, $0 < \delta < T - t$,

$$\begin{aligned} 0 &= \int_{\mathbb{R}^n} v_\rho\,\eta\,dx\Big|_t^{T-\delta} - \int_{\mathbb{R}^n}\int_t^{T-\delta} (v_\rho\eta_\tau + (v_\rho - 1)_+\Delta\eta)\,dx\,d\tau \\ &= \int_{\mathbb{R}^n} v_\rho\,\eta\,dx\Big|_t^{T-\delta} - \int_{\mathbb{R}^n}\int_t^{T-\delta} (v_\rho - (v_\rho-1)_+)\,\eta_\tau\,dx\,d\tau\,, \end{aligned} \tag{2.9}$$

where we have defined

$$\eta(x,\tau) = \frac{1}{(4\pi)^{n/2}} \frac{1}{(T-\tau)^{n/2}} e^{-|x-x_0|^2/(4(T-\tau))} ,$$

so that $\eta_\tau + \Delta\eta = 0$, $\tau < T$.

Note that since

$$v_\rho - (v_\rho - 1)_+ = (v_\rho - (v_\rho - 1)_+ - 1) + 1 = -(1 - v_\rho)_+ + 1 ,$$

it holds

$$\begin{aligned} -\int_{\mathbb{R}^n} \int_t^{T-\delta} & (v_\rho - (v_\rho - 1)_+)\, \eta_\tau \, dx \, d\tau \\ &= \int_{\mathbb{R}^n} \int_t^{T-\delta} (1 - v_\rho)_+ \, \eta_\tau - \int_{\mathbb{R}^n} \eta(x,\tau)\, dx \Big|_t^{T-\delta} \\ &= \int_{\mathbb{R}^n} \int_t^{T-\delta} (1 - v_\rho)_+ \, \eta_\tau \\ &\le \iint_{A_\varepsilon} |\eta_\tau| \, dx \, d\tau , \end{aligned}$$

where

$$A_\varepsilon = \Big(\mathbb{R}^n \times (0,T)\Big) \setminus \Big(B_\varepsilon(x_0) \times (T-\varepsilon, T)\Big) .$$

The last integral is majorized by a constant depending on T, n, and ε only; since ε depends in turn on t_0 (through $u[t_0]$), we denote this constant $c_1(\varepsilon(t_0))$. Thus, letting $\delta \to 0$ in (2.9), we have for $t_0 < t < T$

$$\begin{aligned} & v_\rho(x_0,T) + c_1(\varepsilon(t_0)) \\ & \qquad \ge (4\pi)^{-n/2} (T-t)^{-n/2} \int_{\mathbb{R}^n} v_\rho(x,t)\, e^{-|x-x_0|^2/(4(T-t)} \, dx , \end{aligned} \tag{2.10}$$

implying, owing to (2.6)

$$\begin{aligned} & (4\pi)^{n/2} [u(x_0,T) + c_1(\varepsilon(t_0))] \\ & \qquad \ge (T-t)^{-n/2} \int_{\mathbb{R}^n} u[t_0](x,t) e^{-|x-x_0|^2/(4(T-t))} \, dx , \end{aligned} \tag{2.11}$$

for $t_0 < t < T$.

Due to Theorem 1.1. iv), estimate (2.11) is trivial if $u[t_0](x,T) \le 1$ for all $x \in \mathbb{R}^n$. Of course, the same conclusions hold for any time level

$t_1 \in (t_0, T)$ and the relative $u[t_1]$, but the constant c_1 in (2.11), as well as the point x_0, are a priori different from the ones found above.

Next we show that actually the same x_0 and ε as those employed in estimating $u[t_0](x,t)$ may be used in estimating $u[t_1](x,t)$, $t_1 < t < T$. Note that, because of (2.6),

$$u[t_0](x,t_1) \le u(x,t_1) = u[t_1](x,t_1)\,,$$

for a.e. $x \in \mathbb{R}^n$.

Taking into account (2.11) and the analogous inequality valid for $u[t_1]$, $t_1 \le t < T$, for all $T - t_1 > \sigma > 0$ we may find constants $\gamma = \gamma(\sigma)$, $M = M(\sigma, t_0, t_1)$ such that

$$\sup_{t_1 \le t \le T-\sigma} \int_{\mathbb{R}^n} \left(u[t_0](x,t) + u[t_1](x,t)\right) e^{-\gamma|x|^2} dx \le M\,. \tag{2.12}$$

Therefore a comparison principle may be applied ([7] and Remark 2.2), giving

$$u[t_0] \le u[t_1] \qquad \text{a.e. in } \mathbb{R}^n \times (t_1, T)$$

(indeed σ in (2.12) may be chosen arbitrarily small). Hence

$$u[t_1](x,t) \ge 1 + 2\varepsilon\,, \quad \text{for all } (x,t) \in B_\varepsilon(x_0) \times (T-\varepsilon, T)\,, \tag{2.13}$$

where x_0 and ε are the same as in (2.7). We may now repeat the estimation of $u(\cdot, t_1) = u[t_1](\cdot, t_1)$ with this choice of x_0 and ε and, taking into account the arbitrariness of $t_1 \in (t_0, T)$, we get

$$\begin{aligned} &(4\pi)^{n/2}\Big(u(x_0,T) + c_1(\varepsilon(t_0))\Big) \\ &\qquad \ge (T-t)^{-n/2} \int_{\mathbb{R}^n} u(x,t) e^{-|x-x_0|^2/(4(T-t))}\, dx\,, \end{aligned} \tag{2.14}$$

for almost all $t_0 < t < T$. We still have to get rid of the dependence on t_0 of both x_0 and ε in the left hand side of (2.14). We reason as follows.

Estimate (2.14) implies that, for all $0 < \sigma < T - t_0$,

$$u(\cdot,t) \in \mathcal{G}_{1/2\sigma} \quad \text{uniformly for} \quad t_0 < t < T - \sigma\,.$$

Then, since

$$\|u[t_0](x, t+t_0) - u(x, t+t_0)\|_{L^1_{\mathrm{loc}}(\mathbb{R}^n)} \to 0\,, \qquad \text{as } t \downarrow 0\,,$$

the uniqueness result in [7] can be applied (see Remark 2.2), to find

$$u[t_0] = u \qquad \text{a.e. in } \mathbb{R}^n \times (t_0, T) .$$

Next we choose (x_1, T) and $\varepsilon_1 > 0$ such that for $t_0 < T$,

$$u[t_0](x,t) \equiv u(x,t) \geq 1 + 2\varepsilon_1 , \tag{2.15}$$

for all $(x,t) \in B_{\varepsilon_1}(x_1) \times (T - \varepsilon_1, T)$. We remark that the choice of x_1 and ε_1 does not depend on t_0. We may now repeat the arguments leading to (2.10) and (2.11), to find

$$\begin{aligned} &(4\pi)^{n/2}\big(u(x_1,T) + c_1(\varepsilon_1)\big) \\ &\qquad \geq (T - t_0)^{-n/2} \int_{\mathbb{R}^n} u(x,t_0) e^{-|x-x_1|^2/(4(T-t_0))}\, dx , \end{aligned} \tag{2.16}$$

for a.e. $0 < t_0 < T$. We note again that in (2.16) x_1, ε_1 may be chosen without any further constraint than (2.15). Inequality (2.1) follows easily from (2.16) (see [11, Lemma 4, p. 25]).

In order to show that (2.1) is optimal, just consider the Cauchy problem

$$\begin{cases} u_t = \Delta u & \text{in } \mathbb{R}^n \times (0,T) , \\ u(x,0) = e^{|x|^2} & \text{for } x \in \mathbb{R}^n . \end{cases}$$

It is well known that a solution u exists in $\mathbb{R}^n \times (0,T)$, $T = 1/4$. Since $u \geq 1$ in $\mathbb{R}^n \times (0,T)$, u may be seen as a solution to (1.1) in $\mathbb{R}^n \times (0,T)$, $u(\cdot,t) \sim e^{c(t)|x|^2}$ as $|x| \to \infty$.

REMARK 2.2. (Comparison in $\mathcal{G}_c$, personal communication of J.E. Bouillet). In order to adapt the proof of [7] to our situation it has to be shown that for suitable $0 < \tau < t < T$

$$\int_{R-1}^{R+1} \int_{\tau}^{t} \oint_{\partial B_{\tilde{R}}} f^r \nabla\phi \cdot \nu \, dS \, d\theta \, d\tilde{R} \to 0 , \qquad \text{as } R \to \infty ,$$

uniformly in $r > 0$, where

$$|\nabla\phi(x,t)| \leq k\, e^{-(R-R_1-1)^2/(8(t-\tau))} ,$$

$|x| = R > R_1 > 0$, $k > 0$ given,

$$\sup_{0<t<T} \int_{\mathbb{R}^n} |f(x,t)|\, e^{-c|x|^2}\, dx < \infty , \tag{2.17}$$

and

$$f^r(x,t) = \int_{|x-z|<r} f(z,t)\rho_r(x-z)\,dz\,,$$

with ρ_r the standard (compactly supported) mollifier. But for R large enough, $0 < r < 1$, we have

$$\begin{aligned}
&\Big| \int_{R-1}^{R+1} \int_\tau^t \oint_{\partial B_{\tilde R}} f^r\, \nabla\phi\cdot\nu\; dS\, d\theta\, d\tilde R \Big| \\
&\qquad \le k \int_{R-1}^{R+1} \int_\tau^t \oint_{\partial B_{\tilde R}} \int |f(z,\theta)|\rho_r(x-z) \\
&\qquad\qquad\qquad \cdot e^{-(R-R_1-1)^2/(8(t-\tau))}\, dz\, dS_x\, d\theta\, d\tilde R \\
&\qquad \le k \int_\tau^t \int_{B_{R+1}\setminus B_{R-1}} \int |f(z,\theta)|\rho_r(x-z)\, e^{-|z|^2/(16(t-\tau))}\, dz\, dx d\theta \\
&\qquad \le k \int_\tau^t \int_{B_{R+2}\setminus B_{R-2}} |f(z,\theta)|\, e^{-|z|^2/(16(t-\tau))}\, dz\, d\theta \to 0\,,
\end{aligned}$$

as $R \to \infty$ if $1/(16(t-\tau)) > c$ due to (2.17).

3. Applications.

In this section we will derive some consequences of Theorem 2.1. Corollary 3.1 (existence of a unique initial trace in the class $\mathcal{G}_c$) is an extension of the result of [12], valid under the a priori assumption $u(\cdot,t) \in \mathcal{G}_c$, $T > t > 0$, to any nonnegative distributional solution $u \in L^1_{\rm loc}(\mathbb{R}^n \times (0,T))$ to (1.1). Corollary 3.2 extends the comparison result of [7] (uniqueness in $\mathcal{G}_c \cap L^\infty_{\rm loc}(\mathbb{R}^n \times (0,T))$) to distributional solutions belonging to $L^1_{\rm loc}(\mathbb{R}^n \times (0,T))$ (see also Remark 2.2). The proof of Theorem 3.4 (existence for the Cauchy problem when $u_0(x) \in \mathcal{G}_c \cap L^1_{\rm loc}(\mathbb{R}^n)$) yields as a by-product the fact that the growth at infinity "at most as $e^{c|x|^2}$" is actually pointwise.

Corollary 3.1. *For any nonnegative distributional solution $u \in L^1_{\rm loc}(\mathbb{R}^n \times (0,T))$, there exists a unique nonnegative locally finite measure $\mu \in \mathcal{G}_c$, $c = 1/(2T)$, such that*

$$\lim_{t\downarrow 0} \int u(x,t)\varphi(x)\,dx = \int \varphi(x)\,d\mu\,,$$

for all $\varphi \in C_0^\infty(\mathbb{R}^n)$.

PROOF. The proof of [12] for solutions in the class

$$\left\{u(x,t):\ \sup_{R>1}\frac{1}{|B_R|}\int_{B_R} u(x,t)\,e^{-c'|x|^2}\,dx < M\,,\quad 0<t<T\right\}$$

applies with only minor changes to the present situation. In fact the class $\{u(x,t):\ u(\cdot,t)\in\mathcal{G}_c\,,$ uniformly on $t\in(0,T)\}$ is contained in a functional class of the type above, for a suitable c'.

Corollary 3.2. *Let* $u,v\in L^1_{\mathrm{loc}}(\mathbb{R}^n\times(0,T))$ *be two nonnegative solutions (in* $\mathcal{D}'(\mathbb{R}^n\times(0,T))$*) to* (1.1) *such that*

i) $\|(u-v)_+(\cdot,t)\|_{L^1_{\mathrm{loc}}}\to 0$ *as* $t\downarrow 0$, *or*
ii) $n=1$, *and for every* $\varphi\in C_0(\mathbb{R})$,

$$\int_{\mathbb{R}}(u-v)_+(x,t)\,\varphi(x)\,dx\to 0\qquad \text{as}\quad t\downarrow 0\,.$$

Then $u(x,t)\le v(x,t)$ *a.e. in* $\mathbb{R}^n\times(0,T)$.

Corollary 3.2 follows as in [7], once we use the estimates provided by Theorem 2.1 and Remark 2.2.

Corollary 3.3. *Any nonnegative distributional solution* $u\in L^1_{\mathrm{loc}}(\mathbb{R}^n\times(0,T))$ *belongs in fact to* $L^\infty((0,T-\varepsilon):L^1_{\mathrm{loc}}(\mathbb{R}^n))$, *for all* $\varepsilon>0$.

Theorem 3.4. *Let* $0\le u_0\in L^1_{\mathrm{loc}}(\mathbb{R}^n)$ *be such that* $u_0\in\mathcal{G}_c$, $c>0$. *Then there exists a (unique) nonnegative solution to*

$$u_t=\Delta(u-1)_+\,,\qquad \text{in}\quad \mathcal{D}'(\mathbb{R}^n\times(0,T)) \tag{3.1.a}$$

$$\|u(x,t)-u_0(x)\|_{L^1_{\mathrm{loc}}(\mathbb{R}^n)}\to 0,\qquad t\downarrow 0\,, \tag{3.1.b}$$

with $T=1/(4c)$.

PROOF. Let

$$u_0^{(n)}(x)=\begin{cases}0 & \text{if } |x|\ge n\,,\\ u_0(x) & \text{if } |x|<n \quad\text{and } u_0(x)<n\,,\\ n & \text{if } |x|<n \quad\text{and } u_0(x)\ge n\,,\end{cases}$$

and $u^{(n)}(x,t)$ be the (semigroup) solution to (3.1.a) with initial datum $u_0^{(n)}$ (see [5] and references given therein).

It should be pointed out that $(u^{(n)}-1)_+ \le v$, where v is the solution to the heat equation with initial datum $(u_0-1)_+$. This can be shown by local comparison between $(u^{(n)}-1)_+$ (which is compactly supported) and the solution $v^{(n)}$ to the heat equation with initial datum $(u_0^{(n)}-1)_+$:

$$(u^{(n)}(x,t)-1)_+ \le v^{(n)}(x,t) \uparrow v(x,t)\,,$$

employing Theorem 1.1. vi). Since $\{u^{(n)}(x,t)\}$ is increasing in n and bounded (by $v+1$), there exists

$$u(x,t) = \lim_{n\to\infty} u^{(n)}(x,t)\,.$$

By Lebesgue's bounded convergence theorem, u is a solution in $\mathcal{D}'(\mathbb{R}^n \times (0,T))$ to (0.1).

The convergence $u(\cdot,t) \to u_0(\cdot)$ in the sense of measures can be proved subtracting the weak formulations of (2.1) for $u(x,t)$ and $u^{(n)}(x,t)$ with a suitable choice of the test function, and using the fact that

$$\|u^{(n)}(x,t) - u_0^{(n)}(x)\|_{L^1_{\mathrm{loc}}(\mathbb{R}^n)} \to 0\,, \qquad \text{as } t \downarrow 0\,.$$

Then relation (3.1.b) follows using $(u-1)_+ \le v$ and reasoning as in the proof of Lemma 1.4 and of (2.5.b). Finally uniqueness follows from Corollary 3.2.

Corollary 3.5. (of the proof of Theorem 3.4). *For a.e.* $(x,t),(y,s) \in \mathbb{R}^n \times (0,T)$, $0 < s_0 < s < t < T$,

$$u(y,s) \le (v[s_0](x,t)+1)\, e^{c\left(|x-y|^2/(t-s)+\log((t-s_0)/(s-s_0))+1\right)}\,,$$

where $v[s_0]$ *solves*

$$\begin{cases} v_t = \Delta v\,, \\ v(x,s_0) = (u(x,s_0)-1)_+\,. \end{cases}$$

Here $c = c(n)$.

PROOF. Apply Theorem 3.4 with $u_0(x) = u(x,s_0)$ in $\mathbb{R}^n \times (s_0,T)$ and combine it with the known inequality for solutions to the heat equation (see [15]).

REMARK 3.6. The existence result in Theorem 3.4 can be extended to the case of initial datum a Radon measure $\mu \in \mathcal{G}_C$; in this case relation (3.1.b) is replaced by

$$u(\cdot, t) \to \mu \qquad \text{as } t \downarrow 0 \text{ in the sense of measures.}$$

The proof follows the lines of the one given above for $u_0 \in L^1_{\text{loc}}(\mathbb{R}^n)$, employing truncation and regularization of μ.

Acknowledgments. We want to thank Prof. E. Di Benedetto for his helpful suggestions and advice. We also thank Prof. J.E. Bouillet for having pointed out to us the comparison result in Remark 2.2. We are indebted to the referee for some remarks about the presentation of the paper.

The present results were developed during a visit of the second author to the research group of Professors A. Fasano and M. Primicerio at the Dipartimento di Matematica "U. Dini" of the University of Florence. The second author is gratefully indebted to them and the Department "U. Dini" for their kind hospitality, and to the Departamento de Matemática, Facultad de Ciencias Exactas y Naturales of the University of Buenos Aires, for the corresponding leave of absence.

References.

[1] Andreucci, D., Behaviour of mushy regions under the action of a volumetric heat source. *Math. Methods Appl. Sci.* **16** (1993), 35-47.

[2] Andreucci, D., Di Benedetto, E., A new approach to initial traces in nonlinear filtration. *Ann. Inst. H. Poincaré. Analyse non linéaire* **7** (1990), 305-334.

[3] Aronson, D. G., Non-negative solutions of linear parabolic equations. *Ann. Scuola Norm. Sup. Pisa* **22** (1968), 607-694.

[4] Aronson, D. G., Caffarelli, L. A., The initial trace of a solution of the porous medium equation. *Trans. Amer. Math. Soc.* **280** (1983), 351-366.

[5] Bénilan, Ph., Crandall, M. G., The continuous dependence on φ of solutions of $u_t - \Delta\varphi(u) = 0$. *Indiana U. Math. J.* **30** (1981), 161-177.

[6] Bénilan, Ph., Crandall, M. G., Pierre, M., Solutions of the porous medium equation in $\mathbb{R}^N$ under optimal conditions on initial values. *Indiana U. Math. J.* **33** (1984), 51-87.

[7] Bouillet, J. E., Signed solutions to diffusion-heat conduction equations, in *Free Boundary Problems: Theory and Applications.* Proc. Int. Colloq., Irsee, Germany 1987, Vol II, Pitman Res. Notes Math. Series **186** (1990), 888-892.

[8] Di Benedetto, E., Continuity of weak solutions to certain singular parabolic equations. *Ann. Mat. Pura Appl.* **130** (1982), 131-176.

[9] Dahlberg, B. E. J., Kenig, C. E., Non-negative solutions of generalized porous medium equations. *Revista Mat. Iberoamericana* **2** (1986), 267-305.

[10] Dahlberg B. E. J., Kenig, C. E., Weak solutions of the porous medium equation. Preprint (1990).

[11] Friedman, A., *Partial differential equations of parabolic type.* Prentice Hall, 1964.

[12] Korten, M. K., Existencia de trazas de soluciones débiles no negativas de $u_t = \Delta(u-1)_+$. Communication to the XXXIX Reunión Anual de la UMA, Oct. 1989, to appear in *Revista de la Unión Mat. Argentina.*

[13] Korten, M. K., L^1_{loc} solutions to the Cauchy problem for $u_t = \Delta(u-1)_+$. Trabajos de Matemática, IAM, preprint 197, to appear.

[14] Krzyzánski, M., *Partial differential equations of second order.* PWN-Polish Scientific Publishers, Monografie Matematyczne, **53-54** (1971).

[15] Moser, J., A Harnack inequality for parabolic differential equations. *Comm. Pure Appl. Math.* **17** (1964), 101-134.

[16] Oleinik, O. A., Kalashnikov, A. S., Chzhou, Yui-Lin, The Cauchy problem and boundary value problems for equations of the type of non-stationary filtration. *Izv. Akad. Nauk USSR* **22** (1958), 667-704. (russian)

[17] Täcklind, S., Sur les classes quasianalytiques des solutions des équations aux derivées partielles du type parabolique. *Nova Acta Regiae Scoc. Sci. Upsal.* **10** (1936), 1-57.

[18] Tichonov, A., Théorèmes d'unicité pour l'équation de la chaleur. *Mat. Sbornik* **42** (1935), 199-216.

[19] Widder, D. V., Positive temperatures in an infinite rod. *Trans. Amer. Math. Soc.* **55** (1944), 85-95.

Recibido: 17 de diciembre 1.991

D. Andreucci
Dipartimento di Matematica "U. Dini"
Università degli Studi di Firenze
50134 Firenze, ITALIA

and

M. K. Korten
Departamento de Matemática
Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires
and
Instituto Argentino de Matemática (CONICET)
1055 Buenos Aires, ARGENTINA

Work performed within a research program sponsored by ECMI, Italian CNR, with the participation of ILLY'S CAFFE' S.p.A. (Triestre, Italy).

Revista Matemática Iberoamericana
Vol. 9, N.° 2, 1993

Ondelettes generalisées et fonctions d'échelle à support compact

Pierre-Gilles Lemarié-Rieusset

Résumé. On montre que lorsqu'une analyse multi-résolution de $L^2(\mathbb{R})$ de multiplicité d et de facteur de dilatation A (A entier ≥ 2) admet des fonctions d'échelle à support compact alors elle admet également des ondelettes à support compact. Inversement si $(\psi_{\varepsilon,j,k} = A^{j/2}\psi_\varepsilon(A^j x - k))$, $1 \leq \varepsilon \leq E$, $j,k \in \mathbb{Z}$, est une base hilbertienne de $L^2(\mathbb{R})$ avec les fonctions mères ψ_ε continues à support compact alors elle provient d'une analyse multi-résolution de facteur de dilatation A, de multiplicité $d = E/(A-1)$ et de fonctions d'échelle à support compact et de même régularité que les ondelettes ψ_ε. Ces résultats s'étendent aux cas de fonctions à localisation exponentielle et des ondelettes biorthogonales.

Abstract. We show that to any multi-resolution analysis of $L^2(\mathbb{R})$ with multiplicity d, dilation factor A (where A is an integer ≥ 2) and with compactly supported scaling functions we may associate compactly supported wavelets. Conversely, if $(\psi_{\varepsilon,j,k} = A^{j/2}\psi_\varepsilon(A^j x - k))$, $1 \leq \varepsilon \leq E$, $j,k \in \mathbb{Z}$, is a Hilbertian basis of $L^2(\mathbb{R})$ with continuous compactly supported mother functions ψ_ε, then it is provided by a multi-resolution analysis with dilation factor A, multiplicity $d = E/(A-1)$ and with compactly supported scaling functions (which have the same regularity as the wavelets ψ_ε). Those results can be extended to the cases of exponentially localized functions and of biorthogonal wavelets.

Introduction.

La théorie des *bases d'ondelettes* remonte à 1985 lorsque Y. Meyer se posa et résolut dans [14] et [16] le problème de construire des bases orthonormées de $L^2(\mathbb{R})$ de la forme $(\psi_{j,k})_{j,k\in\mathbb{Z}}$ avec

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k) \tag{1}$$

où ψ était régulière, localisée et oscillante. La fonction ψ était appelée la *mère* des ondelettes, puisqu'elle engendrait la base $(\psi_{j,k})$ par translations et dilatations dyadiques. Pour étendre la construction de Y. Meyer à $\mathbb{R}^n$, Y. Meyer et l'auteur introduisirent une seconde fonction φ, le *père* des ondelettes, telle que la famille

$$(\varphi(x-k))_{k\in\mathbb{Z}} \bigcup (\psi_{j,k})_{j\geq 0,\ k\in\mathbb{Z}}$$

forment une base orthonormée de $L^2(\mathbb{R})$.

En 1986, S. Mallat introduisit le concept *d'analyse multi-résolution* dans [15]. Une analyse multi-résolution de $L^2(\mathbb{R})$ est une suite de sous-espaces fermés $(V_j)_{j\in\mathbb{Z}}$ de $L^2(\mathbb{R})$ tels que

$$\text{(2.a)} \qquad V_j \subset V_{j+1}\ ,\ \bigcap_{j\in\mathbb{Z}} V_j = \{0\}\ ,\ \bigcup_{j\in\mathbb{Z}} V_j \text{ est dense dans } L^2(R)\ ,$$

$$\text{(2.b)} \qquad f(x) \in V_j \text{ si et seulement si } f(2x) \in V_{j+1}\ ,$$

$$\text{(2.c)} \qquad f(x) \in V_0 \text{ si et seulement si } f(x-1) \in V_0\ ,$$

$$\text{(2.d)} \qquad V_0 \text{ a une base orthonormée de la forme } \varphi(x-k),\ k\in\mathbb{Z}.$$

La fonction φ est alors appelée (la) *fonction d'échelle* de (V_j). Le lien avec les bases d'ondelettes est le suivant. Si $(\psi_{j,k})$ est une base d'ondelettes, on note W_j l'espace fermé de L^2 engendré par les $\psi_{j,k}$, $k\in\mathbb{Z}$, et $V_j = \oplus_{p\leq j-1} W_p$. Il est clair alors que (V_j) vérifie (2.a), (2.b) et (2.c); dire que (V_j) vérifie (2.d) revient à dire qu'il existe une fonction père associée à $(\psi_{j,k})$. Inversement si (V_j) est une analyse multi-résolution de fonction d'échelle φ, on note W_j le complémentaire orthogonal de V_j dans V_{j+1}. On montre alors que W_0 a une base orthonormée de la forme $\psi(x-k)$, $k\in\mathbb{Z}$; la fonction ψ est alors la mère d'une base d'ondelettes dont φ est le père.

L'inclusion $V_0 \subset V_1$ donne une *équation à deux échelles* sur φ

(3.a) $\varphi(x) = \sum_{k\in\mathbb{Z}} a_k \varphi(2x-k)$ avec $a_k = \langle \varphi, 2\varphi(2x-k)\rangle$

ou encore, en notant $\hat{\varphi}(\xi) = \int \varphi(x) e^{-ix\xi} dx$ la transformée de Fourier de φ,

(3.b) $$\hat{\varphi}(\xi) = m_0\Big(\frac{\xi}{2}\Big)\hat{\varphi}\Big(\frac{\xi}{2}\Big) \quad \text{avec} \quad m_0(\xi) = \frac{1}{2}\sum_{k\in\mathbb{Z}} a_k e^{-ik\xi}\,.$$

On obtient alors, si $\hat{\varphi}$ est continue,

$$\hat{\varphi}(\xi) = \hat{\varphi}(0)\prod_1^\infty m_0\Big(\frac{\xi}{2^j}\Big)\,.$$

La fonction ψ se déduit de la fonction φ qui se déduit elle-même de la fonction m_0. Cette réduction a permis à I. Daubechies de construire en 1987 dans [3] des bases d'ondelettes régulières à support compact.

Toute base d'ondelettes ne provient pas d'une analyse multi-résolution, comme le montrent des contre-exemples (voir par exemple [10]; l'idée de ces contre-exemples est due à Jean-Lin Journé). Cependant si la mère ψ est Höldérienne à support compact (ou à localisation exponentielle), l'auteur a démontré récemment qu'il existait alors une fonction-père φ elle-même à support compact (ou à localisation exponentielle) *cf.* [12] et [13].

La notion d'analyse multi-résolution se généralise de plusieurs manières. Le premier exemple que nous étudierons est le passage à la dimension n. Une analyse multi-résolution de $L^2(\mathbb{R}^n)$ est une suite de sous-espaces fermés $(V_j)_{j\in\mathbb{Z}}$ de $L^2(\mathbb{R}^n)$ tels que:

(4.a) $$V_j \subset V_{j+1}\,,\ \bigcap_{j\in\mathbb{Z}} V_j = \{0\}\,,\ \bigcup_{j\in\mathbb{Z}} V_j \text{ est dense dans } L^2(\mathbb{R}^n)\,,$$

(4.b) $f(x) \in V_j$ si et seulement si $f(2x) \in V_{j+1}$,

(4.c) $f(x) \in V_0$ si et seulement si $f(x-k) \in V_0$ pour tout $k \in \mathbb{Z}^n$,

(4.d) V_0 a une base orthonormée de la forme $\varphi(x-k)$, $k \in \mathbb{Z}^n$.

On définit de même W_j comme le complémentaire orthogonal de V_j dans V_{j+1}. W_0 a alors une base orthonormée de la forme

$$\psi_\varepsilon(x-k)\,, \qquad 1 \le \varepsilon \le 2^n - 1\,;\ k \in \mathbb{Z}^n\,.$$

Le problème de la construction des ψ_ε est assez compliqué quand les V_j ne proviennent pas d'une analyse multi-résolution $(V_j^{(1)})$ de $L^2(\mathbb{R})$ par tensorisation $(V_j = V_j^{(1)}\hat{\otimes}\cdots\hat{\otimes}V_j^{(1)})$. En 1987, K. H. Gröchenig a montré le résultat fondamental suivant, *cf.* [7], [17]: si φ est à décroissance rapide $(\forall\alpha \in \mathbb{N}^n,\ x^\alpha\varphi(x) \in L^2(\mathbb{R}^n))$ on peut choisir de même les ψ_ε. Ce théorème se démontre assez simplement et une analyse un peu plus détaillée de sa démonstration permet de conclure que si φ est à localisation exponentielle on peut choisir de même les ψ_ε et que si V_0 a une base de Riesz de la forme $g(x-k)$, $k \in \mathbb{Z}^n$, avec g à support compact il en va de même pour W_0 (avec $2^n - 1$ fonctions g_ε). Ce dernier résultat, redémontré par Micchelli et ses collaborateurs dans [9] en utilisant un théorème de géométrie algébrique, ne semble pas avoir été mis en lumière comme corollaire simple du théorème de Gröchenig ; nous consacrerons donc une annexe au théorème de Gröchenig et à ses corollaires.

Un problème qui reste ouvert et que la démonstration du théorème de Gröchenig ne permet pas de résoudre est le suivant: si φ est à support compact, peut-on choisir les ψ_ε à support compact ? L'orthonormalisation de la base de Riesz $g_\varepsilon(x-k)$, $k \in \mathbb{Z}^n$, $1 \leq \varepsilon \leq 2^n - 1$, détruit la compacité des supports des g_ε et ne donne comme information qu'une localisation exponentielle des ψ_ε.

Une autre généralisation est de changer le facteur de dilatation 2 dans (2b) par un facteur entier $A \geq 2$. Le cas $A = 3$ est brièvement discuté dans le dernier chapitre du livre d'I. Daubechies [4] qui explique les motivations d'une telle généralisation en traitement du signal. On peut aussi changer le nombre de fonctions d'échelle, *cf.* [6] et [8]: cela permet par exemple de généraliser les analyses multi-résolutions de fonctions splines à des fonctions polynômiales par morceaux plus générales ; les splines affines correspondent à une interpolation lagrangienne (avec l'interpolante $\Delta(x) = (1 - |x|)^+$) tandis que les polynômes par morceaux de degré 3 et de classe C^1 permettent une interpolation hermitienne (avec les interpolantes $\alpha(x) = (1 + 2|x|)(1 - |x|)^2\chi_{[-1,1]}$ et $\beta(x) = x(1 - |x|)^2\chi_{[-1,1]}$).

Une *analyse multi-résolution généralisée* de $L^2(\mathbb{R})$ de facteur de dilatation A (A entier ≥ 2) et de multiplicité d (d entier ≥ 1) est une suite de sous-espaces fermés $(V_j)_{j\in\mathbb{Z}}$ de $L^2(\mathbb{R})$ tels que

$$\text{(5.a)} \qquad V_j \subset V_{j+1}\ ,\ \bigcap_{j\in\mathbb{Z}} V_j = \{0\}\ ,\ \bigcup_{j\in\mathbb{Z}} V_j \text{ est dense dans } L^2(\mathbb{R})\ ,$$

$$\text{(5.b)} \qquad f(x) \in V_j \text{ si et seulement si } f(Ax) \in V_{j+1}\ ,$$

(5.c) $f(x) \in V_0$ si et seulement si $f(x-1) \in V_0$,

(5.d) V_0 a une base orthonormée de la forme $\varphi_\ell(x-k)$, $1 \leq \ell \leq d$, $k \in \mathbb{Z}$.

Les fonctions $\varphi_1, \ldots, \varphi_d$ sont appelées les fonctions d'échelle de (V_j). On note encore W_j le complémentaire orthogonal de V_j dans V_{j+1}. W_0 a une base orthonormée de la forme

$$\psi_\varepsilon(x-k), \qquad 1 \leq \varepsilon \leq (A-1)d;\ k \in \mathbb{Z}.$$

On peut facilement adapter le théorème de Gröchenig pour montrer que si V_0 admet des fonctions d'échelle φ_ℓ à décroissance rapide W_0 admet des ondelettes ψ_ε à décroissance rapide; de même si V_0 admet des fonctions d'échelle φ_ℓ à localisation exponentielle (ou une base de Riesz $g_\ell(x-k)$, $1 \leq \ell \leq d$, $k \in \mathbb{Z}$, avec g_ℓ à support compact) W_0 admet des ondelettes ψ_ε à localisation exponentielle (ou une base de Riesz $\gamma_\varepsilon(x-k)$, $1 \leq \varepsilon \leq (A-1)$, $k \in \mathbb{Z}$, avec γ_ε à support compact). Nous renvoyons encore une fois à l'annexe sur le théorème de Gröchenig pour de plus amples développements.

Une fois encore, le théorème de Gröchenig ne permet pas de conclure que si V_0 admet des fonctions d'échelles φ_ℓ à support compact on peut choisir les ondelettes ψ_ε elles-mêmes à support compact. Le but de cet article est de démontrer ce résultat (par une méthode totalement différente de celle de Gröchenig). Cette méthode permettra de montrer également qu'inversement toute base orthonormée de $L^2(\mathbb{R})$ de la forme

$$A^{j/2}\psi_\varepsilon(A^j x - k);\, 1 \leq \varepsilon \leq E\,,\ j,k \in \mathbb{Z} \text{ (avec } A \text{ entier} \geq 2)$$

où les ψ_ε sont continues à support compact provient d'une analyse multi-résolution généralisée de facteur de dilatation A et de multiplicité $d = E/(A-1)$ (de sorte que E doit être divisible par $A-1$) et de fonctions d'échelle à support compact ; de plus les fonctions d'échelle et les ondelettes ont la même régularité. Ce résultat s'étend à la localisation exponentielle.

Une dernière généralisation est l'analyse multi-résolution bi-orthogonale de A. Cohen, I. Daubechies et J. C. Feauveau, *cf.* [2] et [5]. Une analyse multi-résolution bi-orthogonale de $L^2(\mathbb{R})$ est la donnée de deux analyses multi-résolutions (V_j), (V_j^*) (vérifiant (2.a) à (2.d)) telles qu'elles admettent des fonctions d'échelle φ, φ^* Höldériennes et en dualité

$$\langle \varphi(x), \varphi^*(x-k) \rangle = \delta_{k,0}\,. \tag{6}$$

On définit alors W_j comme $W_j = V_{j+1} \cap (V_j^*)^\perp$ et de même $W_j^* = V_{j+1}^* \cap V_j^\perp$. Le projecteur oblique P_j de L^2 sur V_j parallèlement à $(V_j^*)^\perp$ a pour noyau

$$p_j(x,y) = 2^j \sum_{k\in\mathbb{Z}} \varphi(2^j x - k)\, \bar{\varphi}^*(2^j y - k) \tag{7}$$

et vérifie $P_j \circ P_{j+1} = P_{j+1} \circ P_j = P_j$; l'opérateur $Q_j = P_{j+1} - P_j$ est donc un projecteur, c'est le projecteur sur W_j parallèlement à $(W_j^*)^\perp$; de plus W_0 admet une base de Riesz $\psi(x-k)$, $k \in \mathbb{Z}$, et W_0^*, $\psi^*(x-k)$, $k \in \mathbb{Z}$, telles que $\langle \psi, \psi^*(x-k)\rangle = \delta_{k,0}$ et

$$Q_j f = 2^j \sum_{k\in\mathbb{Z}} \langle f, \psi^*(2^j x - k)\rangle\, \psi(2^j x - k)\,.$$

Si φ et φ^* sont à localisation exponentielle (respectivement à support compact), on peut choisir ψ et ψ^* à localisation exponentielle (respectivement, à support compact). Enfin les familles $2^{j/2}\psi(2^j x - k)$, $j,k \in \mathbb{Z}$ et $2^{j/2}\psi^*(2^j x - k)$, $j,k \in \mathbb{Z}$ forment un système de bases inconditionnelles de $L^2(\mathbb{R})$ qui sont bi-orthogonales

$$2^{(j+j')/2}\langle \psi(2^j x - k), \psi^*(2^{j'} x - k')\rangle = \delta_{j,j'}\ \delta_{k,k'}\ .$$

Nous montrerons qu'inversement si ψ et ψ^* sont Höldériennes, engendrent par translation-dilatations dyadiques des bases bi-orthogonales de $L^2(\mathbb{R})$ et sont à localisation exponentielle, alors elles proviennent d'une analyse multi-résolution bi-orthogonale avec des fonctions d'échelle duales φ, φ^* à localisation exponentielle. Lorsque ψ et ψ^* sont à support compact, φ et φ^* peuvent de même être choisies à support compact.

Le plan de l'article est alors le suivant:

1. Enoncé des résultats : ondelettes et fonctions d'échelles à support compact.
2. Projecteurs $\mathbb{Z}$-invariants, indices et bases hilbertiennes.
3. Le cas de la localisation exponentielle.
4. Ondelettes bi-orthogonales à localisation exponentielle.
5. Ondelettes bi-orthogonales à support compact.
6. Annexe: le Théorème de Gröchenig, ses corollaires et ses variantes.

NOTATIONS.

- On note L^2_{comp} l'espace des fonctions de L^2 à support compact. Si $\omega \in L^2_{\text{comp}}$ on note $\delta(\omega)$ le diamètre du support de ω, $\delta(\omega) = \sup\{|x-y| : x, y \in \operatorname{supp}\omega\}$.
- Pour $\alpha \geq 0$, on note C^α l'espace des fonctions de classe C^α: si α est entier, $f \in C^\alpha$ si f est α fois continûment dérivable et si ses dérivées sont bornées; si $\alpha = N + \rho$, N entier, $0 < \rho < 1$, $f \in C^\alpha$ si $f \in C^N$ et si

$$|||f^{(N)}|||_\rho < +\infty, \quad \text{où} \quad |||g|||_\rho = \sup_{x \neq y} \frac{|g(x) - g(y)|}{|x-y|^\rho}$$

  (module de continuité Höldérienne d'exposant ρ).
- De même on notera E_α l'espace des fonctions régulières (de classe α) à localisation exponentielle (ainsi que leurs dérivées): la définition précise de E_α est donnée au début de la Section 3.
- La transformée de Fourier $\hat{f}$ d'une fonction f est définie par

$$\hat{f}(\xi) = \int f(x)\, e^{-ix\xi}\, dx\,.$$

1. Enoncé des résultats: Ondelettes et fonctions d'échelle à support compact.

Le but de cet article est de montrer le théorème suivant

Théorème 1. *Soit A un entier ≥ 2. Alors*

i) *Si $(V_j)_{j\in\mathbb{Z}}$ est une analyse multi-résolution de facteur de dilatation A et de multiplicité d à fonctions d'échelle φ_ℓ $(1 \leq \ell \leq d)$ à support compact alors le complémentaire orthogonal W_0 de V_0 dans V_1 admet une base orthonormée de la forme*

$$(\psi_\varepsilon(x-k))_{1\leq\varepsilon\leq(A-1)d;\ k\in\mathbb{Z}}$$

où les ψ_ε sont à support compact. En particulier les fonctions

$$A^{j/2}\psi_\varepsilon(A^j x - k)\,, \qquad 1 \leq \varepsilon \leq (A-1)d;\ j,k \in \mathbb{Z}\,,$$

forment une base orthonormée de $L^2(\mathbb{R})$.

On a de plus

$$\delta(\psi_\varepsilon) \leq \max_{1\leq \ell \leq d} \delta(\varphi_\ell) \qquad \textit{pour} \quad 1 \leq \varepsilon \leq (A-1)d$$

et si les φ_ℓ sont de classe C^α pour un réel $\alpha \geq 0$ alors ψ_ε est également de classe C^α.

ii) *Inversement, si*

$$(A^{j/2}\psi_\varepsilon(A^j x - k))_{1\leq \varepsilon \leq E;\ k\in\mathbb{Z}}$$

est une base orthonormée de $L^2(\mathbb{R})$ avec les ψ_ε à support compact et continues, les $\psi_\varepsilon(x-k)$, $1 \leq \varepsilon \leq E$, $k \in \mathbb{Z}$, forment une base hilbertienne d'un espace W_0 associé à une analyse multi-résolution (V_j) de facteur de dilatation A, de multiplicité $d = E/(A-1)$ (en particulier E est un multiple de $A-1$) et de fonctions d'échelle φ_ℓ $(1 \leq \ell \leq d)$ à support compact et continues. De plus, on a

$$\delta(\varphi_\ell) \leq \max_{1\leq \varepsilon \leq E} \delta(\psi_\varepsilon)$$

et si les ψ_ε sont de classe C^α alors les φ_ℓ sont de classe C^α.

La démonstration repose sur l'analyse des projecteurs orthogonaux Q_0 sur W_0 pour le point i) et P_0 sur V_0 pour le point ii). Ces projecteurs admettent des noyaux

$$q_0(x,y) = -\sum_{\ell=1}^{d}\sum_{k\in\mathbb{Z}} \varphi_\ell(x-k)\,\bar{\varphi}_\ell(y-k) + A\sum_{\ell=1}^{d}\sum_{k\in\mathbb{Z}} \varphi_\ell(Ax-k)\,\bar{\varphi}(Ay-k)$$

et

$$p_0(x,y) = \sum_{j=-\infty}^{-1}\sum_{\varepsilon=1}^{E}\sum_{k\in\mathbb{Z}} A^j\,\psi_\varepsilon(A^j x-k)\,\bar{\psi}_\varepsilon(A^j y-k)$$

qui vérifient la propriété fondamentale suivante:

$$q_0(x,y) = 0\,, \qquad \text{si } |x-y| \geq \max_\ell \delta(\varphi_\ell)$$

et

$$p_0(x,y) = 0 \qquad \text{si } |x-y| \geq \max_\varepsilon \delta(\psi_\varepsilon)\,.$$

C'est évident pour q_0 ; pour p_0, il faut remarquer que $P_0 = \sum_{j=-\infty}^{-1} Q_j$ (en définissant $V_0 = \oplus_{j\leq -1} W_j$, W_j comme le sous-espace engendré par $A^{j/2}\psi_\varepsilon(A^j x - k)$, $(1 \leq \varepsilon \leq E,\ k \in \mathbb{Z})$, et Q_j le projecteur orthogonal sur W_j) mais que également $P_0 = I - \sum_{j=0}^{+\infty} Q_j$; or pour $j \geq 0$, $q_j(x,y)$ est nul pour $|x-y| \geq (1/A^j)\max_\varepsilon \delta(\psi_\varepsilon)$ et donc pour $|x-y| \geq \max_\varepsilon \delta(\psi_\varepsilon)$. Par ailleurs W_0 est invariant par translations entières dans le cas i) (puisque $W_0 = V_1 \cap V_0^\perp$ et que V_1 et V_0 sont invariants par translations entières) et V_0 est invariant par translations entières dans le cas ii) puisque $V_0^\perp = \oplus_{j\geq 0} W_j$ et que pour $j \geq 0$, W_j est invariant par translation entière.

On est donc amené à étudier les espaces invariants par translation entière dont le noyau du projecteur s'annule loin de la diagonale.

Théorème 2. *Soit $V \subset L^2(\mathbb{R})$ un sous-espace fermé et P son projecteur orthogonal (de noyau-distribution $p(x,y)$). On suppose que*

j) *$f(x) \in V$ si et seulement si $f(x-1) \in V$,*
jj) *$p(x,y) = 0$ si $|x-y| \geq M$.*

On suppose de plus l'hypothèse suivante

jjj) *Pour tout $a < b$, l'espace $V_{a,b} = \{\omega \in V :\ \mathrm{supp}\,\omega \subset [a,b]\}$ est de dimension finie.*

Alors (si $V \neq \{0\}$),

k) *Il existe $\omega \in V$ avec ω à support compact, $\delta(\omega) \leq M$, les $\omega(x-k)$ $(k \in \mathbb{Z})$ orthornomées.*
kk) *V admet une base orthonormée de la forme*

$$\omega_i(x-k), \qquad 1 \leq i \leq N\,;\ k \in \mathbb{Z},$$

avec ω_i à support compact et $\delta(\omega_i) \leq M$.

Le nombre N ne dépend que de V et pas du choix des fonctions de base ω_i. Il sera appelé l'indice de V sur $\mathbb{Z}$ et sera noté $N = \mathrm{Ind}_{\mathbb{Z}} V$.

Le Théorème 2 sera démontré dans la section suivante ainsi que la proposition ci-dessous qui donne un critère pour que la condition jjj) soit vérifiée.

Proposition 1. *Soit V un sous-espace de $L^2(\mathbb{R})$ vérifiant les conditions* j) *et* jj) *du Théorème* 2. *Alors V satisfait également la condition* iii) *si le noyau $p(x,y)$ est une fonction continue.*

Nous pouvons maintenant achever la démonstration du Théorème 1.

• pour le point i): W_0 vérifie de manière évidente la condition jjj) puisque $W_{0,a,b}$ est contenu dans l'espace engendré par les $\varphi_\ell(Ax-k)$ $(1 \le \ell \le d,\ k \in \mathbb{Z})$ tels que $\operatorname{supp}\varphi_\ell(Ax-k) \cap [a,b] \neq \emptyset$; ces fonctions sont en nombre fini et donc $W_{0,a,b}$ est de dimension finie. Le Théorème 2 nous indique donc que W_0 admet une base orthonormée à support compact

$$\psi_\varepsilon(x-k), \qquad 1 \le \varepsilon \le \operatorname{Ind}_{\mathbb{Z}} W_0\ ;\ k \in \mathbb{Z}.$$

Par ailleurs $\operatorname{Ind}_{\mathbb{Z}} V_1 = \operatorname{Ind}_{\mathbb{Z}} V_0 + \operatorname{Ind}_{\mathbb{Z}} W_0$ puisque $V_1 = V_0 \oplus W_0$ (la somme étant orthogonale); $\operatorname{Ind}_{\mathbb{Z}} V_0 = d$ par hypothèse et $\operatorname{Ind}_{\mathbb{Z}} V_1 = Ad$ (en prenant comme base les $\sqrt{A}\,\varphi_\ell(A(x-k)-r)$, $0 \le r \le A-1$, $k \in \mathbb{Z}$, $1 \le \ell \le d$) et donc $\operatorname{Ind}_{\mathbb{Z}} W_0 = (A-1)d$. Enfin tout élément à support compact de V_1 se décompose comme une combinaison linéaire finie des $\varphi_\ell(Ax-k)$ et a donc la même régularité que les fonctions φ_ℓ.

• pour le point ii): on vérifie que V_0 satisfait jjj) en montrant que p_0 vérifie la condition décrite dans la Proposition 1. En effet on a

$$p_0(x,y) = \sum_{j \le -1} \sum_{\varepsilon=1}^{E} \sum_{k \in \mathbb{Z}} A^j\, \psi_\varepsilon(A^j x - k)\, \bar{\psi}_\varepsilon(A^j y - k).$$

Or on vérifie facilement que $\psi_\varepsilon(X-k)\bar{\psi}_\varepsilon(Y-k)$ est nul pour X, Y fixés sauf pour un nombre fini d'indices (ε,k) et que ce nombre se majore indépendamment de X et Y par le nombre $M_0 = E(1+\max_\varepsilon \delta(\psi_\varepsilon))$. Si les ψ_ε sont de classe C^α, $\alpha = N+\rho$, alors: pour $0 \le n \le N$

$$\begin{aligned}\left|\frac{\partial^n}{\partial x^n} p_0\right| &\le \sum_{j \le -1} A^{j(1+n)} M_0 \sup_\varepsilon \|\psi_\varepsilon\|_\infty\, \|\psi_\varepsilon^{(N)}\|_\infty \\ &= \frac{1}{A^{n+1}-1}\, M_0 \sup_\varepsilon \|\psi_\varepsilon\|_\infty\, \|\psi_\varepsilon^{(N)}\|_\infty\end{aligned}$$

et

$$\begin{aligned}&\left|\frac{\partial^N}{\partial x^N} p_0(x,y) - \frac{\partial^N}{\partial x^N} p_0(x+h,y)\right| \\ &\qquad \le \sum_{j \le -1} A^{j(1+N+\rho)}\, |h|^\rho\, M_0 \sup_\varepsilon \|\psi_\varepsilon\|_\infty\, |||\psi_\varepsilon^{(N)}|||_\rho\end{aligned}$$

$$= \Big(\frac{1}{A^{1+\alpha}-1}\, M_0 \sup_{\varepsilon} \|\psi_\varepsilon\|_\infty \, |||\psi_\varepsilon^{(N)}|||_\rho\Big)|h|^\rho.$$

En particulier si les ψ_ε sont continues alors p_0 est continue. Grâce à la Proposition 1 et au Théorème 2, on en déduit que V_0 admet une base orthonormée à support compact $\varphi_\ell(x-k)$, $1 \le \ell \le \mathrm{Ind}_{\mathbb{Z}} V_0$, $k \in \mathbb{Z}$. Comme on a à nouveau $\mathrm{Ind}_{\mathbb{Z}} V_1 = \mathrm{Ind}_{\mathbb{Z}} V_0 + \mathrm{Ind}_{\mathbb{Z}} W_0$, on a $\mathrm{Ind}_{\mathbb{Z}} V_0 = E/(A-1)$, ce qui implique que E est divisible par $A-1$. Enfin la régularité des fonctions φ_ℓ provient de ce que

$$\frac{d^n}{dx^n}\varphi_\ell = \frac{d^n}{dx^n}\left(\int p_0(x,y)\varphi_\ell(y)\,dy\right) = \int \frac{\partial^n}{\partial x^n} p_0(x,y)\varphi_\ell(y)\,dy$$

de sorte que φ_ℓ est bien de classe C^α.

Le Théorème 1 est donc démontré (modulo la Proposition 1 et le Théorème 2).

2. Projecteurs $\mathbb{Z}$-invariants, indices et bases hilbertiennes.

Dans cette section nous démontrons le Théorème 2 et la Proposition 1.

DÉMONSTRATION DE LA PROPOSITION 1. Soit $(f_n)_{1\le n\le N}$ une famille orthonormée de $V_{[a,b]}$. Alors si $\theta \in L^2$, $\sum_n |\langle \theta, f_n\rangle|^2$ est la norme du projeté de θ sur $\mathrm{Vect}_n(f_n) \subset V$ et donc

$$\sum_n |\langle \theta, f_n\rangle|^2 \le \|P\theta\|^2 = \langle P\theta, \theta\rangle = \iint p(x,y)\theta(y)\bar{\theta}(x)\,dxdy\,.$$

On prend $\theta_\varepsilon = \theta\left((x-x_0)/\varepsilon\right)/\varepsilon$ et on fait tendre ε vers 0. Les f_n sont continues puisque

$$f_n(x) = \int_a^b p(x,y) f_n(y)\,dy\,.$$

On obtient donc

$$\sum_{n=1}^{N} |f_n(x_0)|^2 \le p(x_0,x_0) \tag{8.a}$$

et

$$(8.b) \qquad N = \int_a^b \sum |f_n(x_0)|^2 dx_0 \leq \int_a^b p(x_0, x_0)\, dx_0 \ .$$

Comme $p(x,y)$ est continue, $\int_a^b p(x_0,x_0)\,dx_0 < +\infty$ et donc $\dim V_{[a,b]} < +\infty$.

DÉMONSTRATION DU THÉORÈME 2. On va commencer par démontrer que pour tout $\varepsilon > 0$ il existe $\omega_\varepsilon \in V$ satisfaisant ω_ε à support compact, $\delta(\omega_\varepsilon) \leq M + \varepsilon$, les $\omega_\varepsilon(x-k)$ $(k \in \mathbb{Z})$ orthonormées et $\operatorname{supp}\omega_\varepsilon \subset [0, M+1]$. En effet, si $V \neq \{0\}$ il existe au moins une fonction $\Omega \in V$ à support compact non nulle (car d'après jj) les éléments de V à support compact sont denses dans V). On peut supposer que $\alpha = \inf \operatorname{supp}\Omega \in [0,1[$ quitte à translater Ω par un entier. Soit maintenant $\varepsilon > 0$ avec $\alpha + \varepsilon \leq 1$. On définit

$$H_\varepsilon = \left\{ f \in V : \ \operatorname{supp} f \subset [\alpha, +\infty[\,, \ \int_\alpha^{\alpha+\varepsilon} |f|^2\, dx = 1 \right\}$$

et

$$K_\varepsilon = \{ f \in H_\varepsilon : \ \operatorname{supp} f \subset [\alpha, \alpha + 2M + \varepsilon] \} .$$

H_ε est non vide puisque à une constante multiplicative près $\Omega \in H_\varepsilon$. On pose alors $\theta = \inf_{g \in H_\varepsilon} \|g\|_2$; on va montrer que cet infimum est atteint en une fonction ω_0 et que la fonction $\omega_\varepsilon = \omega_0 / \|\omega_0\|_2$ convient.

Pour cela, on décompose $g \in H_\varepsilon$ en

$$g = P(g\, \chi_{[\alpha, \alpha+M+\varepsilon]}) + P(g\, \chi_{[\alpha+M+\varepsilon, +\infty[}) = g_1 + g_2 \ .$$

La fonction g_2 est à support dans $[\alpha+\varepsilon, +\infty[$, de sorte que $g_1 = g - g_2$ vérifie $g_1 \in H_\varepsilon$, Par ailleurs

$$\operatorname{supp} g_1 \subset [\alpha, +\infty[\cap [\alpha - M, \alpha + 2M + \varepsilon]$$

de sorte que $g_1 \in K_\varepsilon$. Enfin

$$\|g_1\|_2 \leq \|g\, \chi_{[\alpha, \alpha+M+\varepsilon]}\|_2 \leq \|g\|_2$$

et il n'y a égalité entre $\|g_1\|_2$ et $\|g\|_2$ que si $\operatorname{supp} g \subset [\alpha, \alpha + M + \varepsilon]$. On en conclut que $\theta = \inf_{g \in K_\varepsilon} \|g\|_2$. Soit $g_0 \in K_\varepsilon$ fixé; on a

$$\theta = \inf \{ \|g\|_2 : \ g \in K_\varepsilon \ , \|g\|_2 \leq \|g_0\|_2 \} .$$

Or l'ensemble $\{g \in K_\varepsilon : \|g\|_2 \leq \|g_0\|_2\}$ est compact puisque c'est un fermé borné de $V_{\alpha,\alpha+2M+\varepsilon}$ qui est de dimension finie. L'infimum est donc atteint en une fonction ω_0 ; en particulier

$$\|\omega_0\|_2 \leq \|P(\omega_0 \chi_{[\alpha,\alpha+M+\varepsilon]})\|_2$$

et donc $\operatorname{supp}\omega_0 \subset [\alpha, \alpha+M+\varepsilon]$ et $\delta(\omega_0) \leq M+\varepsilon$. Par ailleurs si $k \geq 1$, $\omega_0(x-k)$ est à support dans $[\alpha+1, +\infty[$ et donc $\omega_0 + \lambda\omega_0(x-k) \in H_\varepsilon$ quel que soit $\lambda \in \mathbb{C}$; on obtient donc $\|\omega_0\|_2 \leq \|\omega_0 + \lambda\omega_0(x-k)\|_2$, ce qui entraîne $\langle \omega_0, \omega_0(x-k)\rangle = 0$ pour $k \geq 1$ et donc pour $k \neq 0$ puisque $\langle \omega_0, \omega_0(x+k)\rangle = \langle \omega_0(x-k), \omega_0\rangle$. Enfin $\|\omega_0\|_2 \neq 0$ puisque $\int_\alpha^{\alpha+\varepsilon} |\omega_0|^2 dx = 1$. La fonction $\omega_\varepsilon = \omega_0 / \|\omega_0\|_2$ vérifie donc bien les propriétés annoncées.

Or $\|\omega_\varepsilon\|_2 = 1$ et $\omega_\varepsilon \in V_{0,M+1}$ qui est de dimension finie; on peut donc extraire une sous-suite ω_{ε_k} (avec $\varepsilon_k \to 0$) convergente dans $V_{0,M+1}$ vers une fonction ω. Il est clair que ω est dans V, à support dans $[\alpha, \alpha+M]$ (et donc $\delta(\omega) \leq M$) et que

$$\langle \omega, \omega(x-k)\rangle = \lim_{\varepsilon_j \to 0} \langle \omega_{\varepsilon_j}, \omega_{\varepsilon_j}(x-k)\rangle = \delta_{k,0} .$$

On a donc démontré le premier point du Théorème 2.

L'existence de la base $\omega_i(x-k)$, $1 \leq i \leq N$, $k \in \mathbb{Z}$, se démontre alors facilement par récurrence sur $\dim V_{0,M+1}$. En effet, une fois trouvée la fonction ω décrite par le premier point du théorème, on note W l'espace engendré par les $\omega(x-k)$, $k \in \mathbb{Z}$, et $V' = V \cap W^\perp$. Alors V' vérifie les mêmes conditions que V : le projecteur P' de L^2 sur V' s'écrit $P'f = \int p'(x,y) f(y)\, dy$ avec

$$p'(x,y) = p(x,y) - \sum_{k\in\mathbb{Z}} \omega(x-k)\,\bar{\omega}(y-k) ,$$

on a bien $f \in V'$ si et seulement si $f(x-1) \in V'$ et $p'(x,y) = 0$ si $|x-y| \geq M$. De plus, on a $\dim V'_{a,b} \leq \dim V_{a,b}$ puisque $V' \subset V$. Enfin $\dim V'_{0,M+1} \leq \dim V_{0,M+1} - 1$ puisque $V'_{0,M+1} \subset V_{0,M+1}$, $\omega \in V_{0,M+1}$, $\omega \perp V'_{0,M+1}$. Si $\dim V_{0,M+1} = 1$, on voit que $V'_{0,M+1} = \{0\}$ mais cela implique que $V' = \{0\}$ (sinon on pourrait trouver d'après le point k) une fonction $\omega' \in V'_{0,M+1}$ avec $\|\omega'\|_2 = 1$) et donc $V = W$. La récurrence est alors immédiate.

Il ne reste plus à vérifier que le fait que l'indice ne dépend pas du choix de la base. Cela est bien connu et assez évident. Par exemple

on note Π l'opérateur défini sur $V_{\text{comp}} = \{f \in V : f \text{ est à support compact}\}$ par $\Pi f = \sum_{k\in\mathbb{Z}} f(x-k)$; alors si $\omega_i(x-k)$, $1 \leq i \leq N$, $k \in \mathbb{Z}$, est une base orthonormée de V avec les ω_i à support compact, les $\Pi\omega_i$, $1 \leq i \leq N$, sont une base orthonormée de $\Pi V_{\text{comp}} \subset L^2([0,1])$; on voit donc que N ne dépend pas du choix des ω_i puisque $N = \dim \Pi V_{\text{comp}}$.

Le Théorème 2 est donc démontré.

3. Le cas de la localisation exponentielle.

Pour décrire la localisation exponentielle, nous introduisons les espaces E_α , $\alpha \geq 0$, comme suit

i) $f \in E_0$ s'il existe $C, D > 0$ tels que

$$\forall x \in \mathbb{R}, \qquad |f(x)| \leq C^{\,-D|x|} ,$$

ii) $f \in E_\rho$, $0 < \rho < 1$, si $f \in E_0$ et s'il existe $C', D' > 0$ tels que

$$\forall x \in \mathbb{R},\ \forall h \in [-1,1], \quad |f(x) - f(x+h)| \leq C' e^{-D'|x|} |h|^\rho ,$$

iii) $f \in E_{N+\rho}$, $N \in \mathbb{N}$, $0 \leq \rho < 1$, si $f, f', \ldots, f^{(N)} \in E_0$ et si $f^{(N)} \in E_\rho$.

Pour $f \in E_0$, on note $\Pi f = \sum_{k\in\mathbb{Z}} f(x-k)$. On a la propriété fondamentale suivante: si la famille $f_i(x-k)$, $1 \leq i \leq$, $k \in \mathbb{Z}$, est orthonormée dans $L^2(\mathbb{R})$ avec $f_i \in E_0$ alors la famille Πf_i, $1 \leq i \leq L$, est orthonormée dans $L^2([0,1])$: en effet

$$\int_0^1 \Pi f_i \, \overline{\Pi f_j} \, dx = \int_{-\infty}^{+\infty} f_i \, \overline{\Pi f_j} \, dx = \sum_k \langle f_i, f_j(x-k) \rangle .$$

Le Théorème 2 et la Proposition 1 se transcrivent alors en

Théorème 2-bis. *Soit $V \subset L^2(\mathbb{R})$ un sous-espace fermé tel que $f(x) \in V$ si et seulement si $f(x-1) \in V$. Alors*

i) *Si $V \cap E_0 \neq \{0\}$, V contient une fonction $\omega \in E_0$ telle que la famille $\omega(x-k)$, $k \in \mathbb{Z}$, soit orthonormée.*

ii) *Si $V \neq \{0\}$ et si*

(9.1) *$E_0 \cap V$ est dense dans V*

(9.2) $\Pi(E_0 \cap V)$ *est de dimension finie*

alors V *admet une base orthonormée de la forme*

$$\omega_i(x-k), \qquad 1 \leq i \leq \dim \Pi(E_0 \cap V)\,;\ k \in \mathbb{Z},$$

avec $\omega_i \in E_0$.

De plus toutes les bases de la forme $\Omega_i(x-k)$, $1 \leq i \leq L$, $k \in \mathbb{Z}$, *ont le même nombre de fonctions de base* Ω_i : $L = \dim \Pi(E_0 \cap V)$.

iii) *Les conditions* (9.1) *et* (9.2) *sont en particulier vérifiées lorsque le noyau* $p(x,y)$ *du projecteur orthogonal* P *de* L^2 *sur* V *est une fonction continue qui vérifie pour deux constantes* $C, D > 0$

$$|p(x,y)| \leq C\, e^{-D|x-y|} \tag{10}$$

La démonstration du Théorème 2-bis est élémentaire et reprend les idées de [11] et [13]. On choisit $\omega_0 \neq 0$, $\omega_0 \in E_0 \cap V$ et on note W le sous-espace fermé de $L^2(\mathbb{R})$ engendré par les $\omega(x-k)$, $k \in \mathbb{Z}$. On va montrer que W admet une base orthonormée $\omega(x-k)$, $k \in \mathbb{Z}$, avec $\omega \in E_0$. Pour cela, on considère la fonction

$$F(z) = \sum_{k \in \mathbb{Z}} \langle \omega_0(x), \omega_0(x-k) \rangle z^k\,,$$

c'est le développement de Laurent d'une fonction holomorphe au voisinage du cercle-unité $|z| = 1$. De plus, ses zéros sur le cercle-unité sont de multiplicité paire puisque

$$F(e^{-i\xi}) = \sum_{k \in \mathbb{Z}} |\hat{\omega}_0(\xi + 2k\pi)|^2 \geq 0$$

d'après la formule sommatoire de Poisson. $F(z)$ se factorise donc en $M(z)^2 G(z)$ où G ne s'annule pas sur $|z| = 1$ et où $M(z)$ est un polynôme dont toutes les racines sont de module 1.

On définit alors $\hat{\gamma}(\xi) = \hat{\omega}_0(\xi)/M(e^{-i\xi})$. On va montrer que $\gamma \in W \cap E_0$ et que les $\gamma(x-k)$, $k \in \mathbb{Z}$, forment une base de Riesz de W. Pour cela, on utilise le lemme suivant

Lemme 1. *Si* $h \in W \cap E_0$ *et si* $\sum_{k \in \mathbb{Z}} |\hat{h}(\xi_0 + 2k\pi)|^2 = 0$ *alors la fonction* η *définie par*

$$\hat{\eta}(\xi) = \frac{\hat{h}(\xi)}{e^{-i\xi} - e^{-i\xi_0}}$$

est dans $W \cap E_0$.

En effet on montre facilement que si h est à décroissance rapide alors la série $\sum |\hat{h}(\xi + 2k\pi)|^2$ converge en tout point vers une fonction C^∞. Si cette fonction s'annule en ξ_0 , alors on peut factoriser $|e^{-i\xi} - e^{-i\xi_0}|^2$ de cette fonction et on trouve un quotient C^∞. Cela montre que $\eta \in L^2$ puisque

$$\int_{-\infty}^{+\infty} |\hat{\eta}(\xi)|^2 d\xi = \int_0^{2\pi} \frac{\sum |\hat{h}(\xi + 2k\pi)|^2}{|e^{-i\xi} - e^{-i\xi_0}|^2} \, d\xi \, .$$

Par ailleurs, la formule sommatoire de Poisson donne que

$$\sum_{k \in \mathbb{Z}} e^{ik\xi_0} h(x - k) = 0 \, .$$

La fonction

$$\tilde{\eta}(x) = \sum_{k \leq -1} e^{i(k+1)\xi_0} \, h(x - k)$$

est alors dans E_0 : la décroissance exponentielle est clairement vérifiée pour $x \to +\infty$; pour $x \to -\infty$, il suffit de remarquer qu'on a également

$$\tilde{\eta}(x) = - \sum_{k \geq 0} e^{i(k+1)\xi_0} \, h(x - k) \, .$$

En particulier $\tilde{\eta} \in L^2$; comme de plus on a clairement

$$\tilde{\eta}(x) = -e^{i\xi_0} h(x) + e^{i\xi_0} \tilde{\eta}(x - 1) \, ,$$

on voit que $\tilde{\eta} = \eta$ et donc $\eta \in E_0$. De plus

$$- \sum_0^N e^{i(k+1)\xi_0} h(x - k) = \eta_N(x) \longrightarrow \eta$$

dans $\mathcal{D}'$ et

$$\|\eta_N\|_2 = \frac{1}{2\pi} \|\hat{\eta}(1 - e^{-i(N+1)(\xi - \xi_0)})\|_2 \leq 2 \, \|\eta\|_2$$

de sorte que $\eta_N \to \eta$ dans L^2-faible et donc $\eta \in W$. Le lemme est donc démontré.

En itérant la construction du lemme $\deg M$ fois, on voit que $\gamma \in E_0 \cap W$. Par ailleurs les $\gamma(x-k)$ engendrent W (puisque ω_0 est combinaison linéaire finie des $\gamma(x-k)$) et

$$\sum_{k\in\mathbb{Z}} |\hat{\gamma}(\xi+2k\pi)|^2 = c_0 e^{i(\deg M)\xi} G(e^{-i\xi})$$

avec $|c_0| = 1$, de sorte que les $\gamma(x-k)$, $k \in \mathbb{Z}$, forment une base de Riesz de W. La fonction

$$U(z) = \sum_{k\in\mathbb{Z}} \langle \gamma(x), \gamma(x-k)\rangle z^k$$

est holomorphe au voisinage de $|z| = 1$ et est strictement positive sur $|z| = 1$ (puisque $\sum |\hat{\gamma}(\xi+2k\pi)|^2$ ne s'annule pas), de sorte que $U(z)^{-1/2}$ est définie et holomorphe au voisinage de $|z| = 1$; on a alors $U(z)^{-1/2} = \sum_{k\in\mathbb{Z}} \alpha_k z^k$ avec α_k à décroissance exponentielle. Il suffit alors de définir ω comme $\hat{\omega} = U(e^{-i\xi})^{-1/2}\hat{\gamma}$, ou encore $\omega = \sum \alpha_k \gamma(x-k)$, pour obtenir une base orthonormée $\omega(x-k)$ $(k \in \mathbb{Z})$ de W avec $\omega \in E_0$. Le point i) est donc démontré.

Le point ii) est alors évident et se démontre par récurrence sur $\dim \Pi(E_0 \cap V)$. Il suffit pour cela, si $V \neq \{0\}$, de choisir $\omega \in V \cap E_0$ telle que les $\omega(x-k)$ soient orthonormées, de définir W l'espace engendré par les $\omega(x-k)$ et $V_0 = V \cap W^{\perp}$. Le projecteur orthogonal Q sur W est donné par $Qf = \sum \langle f, \omega(x-k)\rangle \omega(x-k)$ et on vérifie immédiatement que $Qf \in E_0$ lorsque $f \in E_0$, de sorte que $V_0 \cap E_0$ est dense dans V_0 (puisque $V_0 = (I-Q)V$). De plus on a

$$\Pi(V \cap E_0) = \Pi(W \cap E_0) \oplus \Pi(V_0 \cap E_0)$$

(la somme directe étant orthogonale) de sorte que $\dim \Pi(V_0 \cap E_0) = \dim \Pi(V \cap E_0) - 1$. On peut donc appliquer l'hypothèse de récurrence à V_0 et exhiber (si $(V_0 \neq \{0\})$ une base orthonormée de V_0 de la forme

$$\omega_i(x-k), \qquad 1 \leq i \leq \dim \Pi(V_0 \cap E_0),\ k \in \mathbb{Z},$$

avec $\omega_i \in E_0 \cap V_0$: il suffit de rajouter les fonctions $\omega(x-k)$, $k \in \mathbb{Z}$, pour obtenir une base orthonormée de V.

Que toutes les bases orthonormées de V de la forme $\omega_i(x-k)$, $1 \leq i \leq N$, $k \in \mathbb{Z}$, (avec $\omega_i \in E_0$) aient le même nombre de fonctions

de base ω_i est immédiat car les $\Pi\omega_i$, $1 \leq i \leq N$, forment alors une base orthonormée de $\Pi(E_0 \cap V)$. Le point ii) du théorème est donc démontré.

Pour démontrer le point iii), on introduit l'opérateur ΠP défini par: si $f \in L^2([0,1])$

$$\Pi P f(x) = \int_0^1 \sum_{k\in\mathbb{Z}} p(x-k,y) f(y)\, dy\,.$$

Si $f \in \Pi(E_0 \cap V)$, il est immédiat que $(\Pi P)f = f$:

$$\Pi P(f) = (\Pi P)(\sum F(x-k)) = \int_{-\infty}^{+\infty} p(x,y) \sum F(y-k)\, dy$$
$$= \sum F(x-k) = f\,.$$

Or la fonction $\sum_{k\in\mathbb{Z}} p(x-k,y)$ est continue. Si $(f_n)_{1\leq n\leq N}$ est orthonormée dans $\Pi(E_0 \cap V)$, on trouve comme pour la Proposition 1,

$$\sum |f_n(x_0)|^2 \leq \sum_{k\in\mathbb{Z}} p(x_0-k,x_0)$$

et

$$\dim \pi(E_0 \cap V) \leq \int_0^1 \sum_{k\in\mathbb{Z}} p(x_0-k,x_0)\, dx_0\,. \tag{11}$$

Par ailleurs il est évident d'après (10) que $P(E_0) \subset E_0$ et donc que $E_0 \cap V$ est dense dans E_0. Le Théorème 2-bis est donc démontré.

Du Théorème 2-bis on déduit directement

Théorème 1-bis. *Soit A un entier ≥ 2. Alors*

i) *Si $(V_j)_{j\in\mathbb{Z}}$ est une analyse multi-résolution de facteur de dilatation A, de multiplicité d et de fonctions d'échelle φ_ℓ $(1 \leq \ell \leq d)$ avec $\varphi_\ell \in E_0$ alors le complémentaire orthogonal W_0 de V_0 dans V_1 admet une base orthonormée de la forme*

$$\psi_\varepsilon(x-k)\,, \qquad 1 \leq \varepsilon \leq (A-1)d;\ k \in \mathbb{Z}\,,$$

avec $\psi_\varepsilon \in E_0$. De plus, si les φ_ℓ sont de classe E_α pour un $\alpha \geq 0$ alors les ψ_ε sont également de classe E_α.

ii) *Inversement si*

$$(A^{j/2}\psi_\varepsilon(A^j x - k))_{1\leq\varepsilon\leq E;\ k\in\mathbb{Z}}$$

est une base orthonormée de $L^2(\mathbb{R})$ avec les ψ_ε continues de classe E_0, les

$$\psi_\varepsilon(x-k), \qquad 1\leq\varepsilon\leq E;\ k\in\mathbb{Z})$$

forment une base hilbertienne d'un espace W_0 d'une analyse multi-résolution (V_j) de facteur de dilatation A, de multiplicité $d = E/(A-1)$ et de fonctions d'échelle φ_ℓ $(1\leq\ell\leq d)$ de classe E_0. De plus si les ψ_ε sont de classe E_α il en va de même pour les φ_ℓ.

Le point i) est immédiat puisque, si P_0 désigne le projecteur orthogonal sur V_0, $P_0(E_0)\subset E_0$ de sorte que $E_0\cap W_0 = (I-P_0)(E_0\cap V_1)$ est dense dans W_0 ; par ailleurs $\dim\Pi(E_0\cap W_0) = (A-1)d$ car on a

$$\Pi(E_0\cap V_1) = \Pi(E_0\cap V_0)\oplus\Pi(E_0\cap W_0).$$

Enfin le fait que les ψ_ε soient de classe E_α est immédiat: on vérifie immédiatement que $E_0\cap V_1\subset E_\alpha$ en décomposant les fonctions de $E_0\cap V_1$ sur la base $(\sqrt{A}\,\varphi_\ell(Ax-k))_{\ell,k}$.

Pour le point ii), il s'agit d'estimer la taille et la régularité du noyau du projecteur

$$P_0 = \sum_{j\leq -1} Q_j = I - \sum_{j\geq 0} Q_j .$$

On commence par remarquer que si $0<D'<D$ alors

$$\sum_{k\in\mathbb{Z}} e^{-D|X-k|}e^{-D|Y-k|} \leq C(D,D')\,e^{-D'|X-Y|}. \tag{12}$$

On obtient donc, lorsque $|x-y|\leq 2$ et $|h|\leq 1$ (et $0\leq p\leq N$, $\alpha = N+\rho$)

$$\left|\frac{\partial^p p_0}{\partial x^p}(x,y)\right| \leq C\sum_{j\leq -1} A^{j(1+p)}\,e^{-D'|x-y|A^j}$$

$$\leq C\sum_{j\leq -1} A^{j(1+p)} \leq C'$$

$$\left|\frac{\partial^N p_0}{\partial x^N}(x,y) - \frac{\partial^N p_0}{\partial x^N}\right| \leq C \sum_{j\leq -1} A^{j(1+N+\rho)}|h|^\rho e^{-D'|x-y|A^j}$$
$$\leq C'|h|^\rho.$$

Lorsque $|x-y| \geq 2$, on obtient

$$\left|\frac{\partial^p p_0}{\partial x^p}(x,y)\right| \leq C \sum_{j\geq 0} A^{j(1+p)}\, e^{-D'A^j|x-y|}$$
$$\leq C\, e^{-D''|x-y|} \sum_{j\geq 0} A^{j(1+p)}\, e^{-(D'-D'')A^j}$$
$$\leq C'\, e^{-D''|x-y|},$$

$(D'' < D')$, et enfin ($|x-y| \geq 2$, $|h| \leq 1$ et donc $|x+h-y| \geq |x-y|/2$)

$$\left|\frac{\partial^N p_0}{\partial x^N}(x,y) - \frac{\partial^N p_0}{\partial x^N}(x+h,y)\right|$$
$$\leq C \sum_{j\geq 0, A^j|h|\leq 1} |h|^\rho A^{j(1+N+\rho)}\, e^{-A^j|x-y|D'}$$
$$+ 2C \sum_{A^j|h|>1} A^{j(1+N)}\, e^{-A^j|x-y|D'/2}$$
$$\leq 2C \sum_{j\geq 0} |h|^\rho A^{j(1+N+\rho)}\, e^{-A^j|x-y|D'/2}$$
$$\leq 2C|h|^\rho e^{-|x-y|D''/2} \sum_{j\geq 0} A^{j(1+N+\rho)}\, e^{-(D'-D'')A^j}$$
$$\leq C'\,|h|^\rho e^{-D''|x-y|/2}.$$

On est alors dans les conditions du Théorème 2-bis. De plus si $f \in E_0$, on obtient immédiatement que $P_0 f \in E_\alpha$ à cause des estimations sur $p_0(x,y)$. Le Théorème 1-bis est donc démontré.

Pour conclure cette section, nous allons traiter le cas où les espaces considérés contiennent un sous-espace dense de fonctions à support compact.

Théorème 2-ter. *Soit $V \subset L^2(\mathbb{R})$ un sous-espace fermé tel que $f(x) \in V$ si et seulement si $f(x-1) \in V$. Alors*

i) *Si* $V \cap L^2_{\text{comp}} \neq \{0\}$, V *contient une fonction* $\omega \in L^2_{\text{comp}}$ *telle que la famille* $\omega(x-k)$, $k \in \mathbb{Z}$, *soit une base de Riesz de l'espace* W *qu'elle engendre. De plus si* $V \cap L^2_{\text{comp}}$ *est dense dans* V, *alors* $V \cap W^\perp \cap L^2_{\text{comp}}$ *est dense dans* $V \cap W^\perp$.

ii) *Si* $V \neq \{0\}$. *si* $V \cap L^2_{\text{comp}}$ *est dense dans* V *et si* $\dim \Pi(L^2_{\text{comp}} \cap V) < +\infty$ *alors* V *admet une base de Riesz de la forme*

$$\omega_i(x-k), \qquad 1 \leq i \leq \dim \Pi(L^2_{\text{comp}} \cap V).$$

De plus on peut choisir les ω_i *tels que* $\langle \omega_i(x), \omega_j(x-k)\rangle = 0$ *si* $i \neq j$ (*quel que soit* $k \in \mathbb{Z}$).

La démonstration du Théorème 2-ter suit l'organisation générale de celle du Théorème 2-bis. L'existence de la fonction ω du point i) se fait à nouveau à l'aide du Lemme 1, puisque lorsque $h \in L^2_{\text{comp}}$ et $\sum |\hat{h}(\xi_0 + 2k\pi)|^2 = 0$ alors il est immédiat que $\eta \in L^2_{\text{comp}}$ (où $\hat{\eta} = \hat{h}/(e^{-i\xi} - e^{-i\xi_0})$). Une fois ω construite, on note Ω la fonction duale de ω

$$\hat{\Omega} = \frac{\hat{\omega}}{\sum_{k\in\mathbb{Z}} |\hat{\omega}(\xi + 2k\pi)|^2}$$

de sorte que le projecteur orthogonal Q sur W est donné par

$$Qf = \sum \langle f, \Omega(x-k)\rangle \omega(x-k);$$

on pose

$$M(\xi) = \sum |\hat{\omega}(\xi + 2k\pi)|^2 = \sum_k \langle \omega, \omega(x-k)\rangle e^{-ik\xi}$$

et $\hat{F} = M(\xi)\hat{f}$ si $f \in L^2_{\text{comp}}$ alors $F \in L^2_{\text{comp}}$ et $QF \in L^2_{\text{comp}}$ (car $QF = \sum \langle f, \omega(x-k)\rangle \omega(x-k)$) de sorte que, si $f \in L^2_{\text{comp}} \cap V$, alors $(I-Q)f = g$ vérifie $\hat{g} = \hat{G}/M(\xi)$ avec $G = (I-Q)F \in L^2_{\text{comp}} \cap V \cap W^\perp$; comme $M(\xi)$ ne s'annule pas, g est dans l'adhérence des combinaisons linéaires des $G(x-k)$ et donc, si $L^2_{\text{comp}} \cap V$ est dense dans V, $L^2_{\text{comp}} \cap V \cap W^\perp$ est bien dense dans $V \cap W^\perp$.

Téorème 1-ter. *Soit* A *un entier* ≥ 2. *Alors si* $(V_j)_{j\in\mathbb{Z}}$ *est une analyse multi-résolution de facteur de dilatation* A *et de multiplicité* d *et si* V_0 *admet une base de Riesz de la forme*

$$\varphi_\ell(x-k), \qquad 1 \leq \ell \leq d, \ k \in \mathbb{Z},$$

avec φ_ℓ à support compact, alors W_0 (le complémentaire orthogonal de V_0 dans V_1) admet une base de Riesz de la forme

$$\psi_\varepsilon(x-k), \qquad 1 \leq \varepsilon \leq (A-1)d,\ k \in \mathbb{Z},$$

avec ψ_ε à support compact.

La réciproque est également vraie (mais présente peu d'intérêt): si W_0 a une base de Riesz de la forme $(\psi_\varepsilon(x-k))_{\varepsilon,k}$ avec les ψ_ε à support compact alors V_0 a une base de Riesz $(\varphi_\ell(x-k))_{\ell,k}$ avec les φ_ℓ à support compact.

Nous terminerons en remarquant que le processus d'orthogonalisation préservant la compacité des supports qui a été utilisé pour démontrer le Théorème 2-ter a d'abord été introduit dans le contexte de la théorie des ondelettes par Micchelli [9], Chui [1] et leurs collaborateurs.

4. Ondelettes bi-orthogonales à localisation exponentielle.

Le cas bi-orthogonal est essentiellement similaire au cas orthogonal dans le cas des fonctions à localisation exponentielle.

Pour comprendre le cas bi-orthogonal, nous nous ramenons une fois encore à l'étude des projecteurs $\mathbb{Z}$-invariants

Proposition 2. a) *Soient V, V^* deux sous-espaces fermés de $L^2(\mathbb{R})$ tels que*

i) *$f \in V$ si et seulement si $f(x-1) \in V$,*
$f \in V^*$ si et seulement si $f(x-1) \in V^*$,
ii) *$E_0 \cap V$ est dense dans V, $E_0 \cap V^*$ est dense dans V^*,*
iii) $\dim \Pi(E_0 \cap V) < +\infty$,
iv) $L^2 = V \oplus (V^*)^\perp$ *(somme directe non orthogonale).*

Alors il existe des fonctions φ_ℓ, $1 \leq \ell \leq \dim \Pi(E_0 \cap V)$, et φ^*_ℓ, $1 \leq \ell \leq \dim \Pi(E_0 \cap V)$, telles que

j) *$\varphi_\ell \in E_0 \cap V$ et $\varphi^*_\ell \in E_0 \cap V^*$,*
jj) *$(\varphi_\ell(x-k))_{\ell,k}$ est une base de Riesz de V et $(\varphi^*_\ell(x-k))_{\ell,k}$ est une base de Riesz de V^*.*
jjj) $\langle \varphi_\ell(x-k), \varphi^*_{\ell'}(x-k') \rangle = \delta_{\ell,\ell'}\delta_{k,k'}$

de sorte que le projecteur P de L^2 sur V parallèlement à $(V^*)^\perp$ s'écrit

$$Pf = \sum_{\ell=1}^{\dim \Pi(E_0 \cap V)} \sum_{k \in \mathbb{Z}} \langle f, \varphi_\ell^*(x-k) \rangle \, \varphi_\ell(x-k) \,. \tag{13}$$

b) *Si $P : L^2 \to L^2$ est un opérateur continu tel que*

k) $P \circ P = P$ (*P est un projecteur*),

et tel que $Pf = \int p(x,y) f(y)\, dy$ où, pour des constantes $C, D > 0$ et $\beta \in]0,1]$

kk) $p(x,y) = p(x-1, y-1)$ (*$\mathbb{Z}$ invariance*),

kkk) $|p(x,y)| \leq C\, e^{-D|x-y|}$,

kv) *Pour* $|h| \leq 1$, $|p(x,y) - p(x+h,y)| \leq C\, |h|^\beta e^{-D|x-y|}$,

alors les espaces $V = \operatorname{Im} P$ et $V^* = (\operatorname{Ker} P)^\perp$ vérifient les hypothèses (i) *à* (iv).

La preuve de cette proposition est très simple. L'hypothèse iv) est équivalente à ce que le projecteur orthogonal $P_\perp$ de L^2 sur V soit un isomorphisme de V^* sur V: il est évidemment continu et l'hypothèse iv) implique qu'il est bijectif; il est donc bicontinu par le Théorème de Banach. Par ailleurs on sait que V_0 admet une base orthonormée de la forme

$$\varphi_\ell(x-k), \qquad 1 \leq \ell \leq \dim \Pi(E_0 \cap V), \; k \in \mathbb{Z}$$

avec $\varphi_\ell \in E_0$.

On fixe alors $\omega^* \in V^* \cap E_0$ tel que les $\omega^*(x-k)$ soient orthonormées, et on note W^* l'espace engendré par les $\omega^*(x-k)$, $k \in \mathbb{Z}$. On note $\Omega = P_\perp \omega^*$; la famille $(\Omega(x-k))$ est une famille de Riesz, puisque

$$\| \sum a_k \, \Omega(x-k) \|_2 \approx \| \sum a_k \omega^*(x-k) \|_2 = (\sum_k |a_k|^2)^{1/2}.$$

Cela implique que $\sum |\hat{\Omega}(\xi + 2k\pi)|^2$ ne s'annule pas; or on a

$$\sum |\hat{\Omega}(\xi + 2k\pi)|^2 = \sum_{\ell=1}^{\dim \pi(\varepsilon_0 \cap V)} |\sum_k \hat{\omega}^*(\xi + 2k\pi)\, \bar{\hat{\varphi}}_\ell(\xi + 2k\pi)|^2 \,;$$

on pose

$$\alpha_\ell(\xi) = \sum_k \hat{\omega}^*(\xi + 2k\pi)\, \bar{\hat{\varphi}}_\ell(\xi + 2k\pi) \qquad \text{et} \qquad \alpha(\xi) = \sum_\ell |\alpha_\ell(\xi)|^2 \,;$$

encore une fois, elles s'écrivent $\alpha_\ell(\xi) = F_\ell(e^{-i\xi})$ et $\alpha(\xi) = F(e^{-i\xi})$ où les fonctions $F_\ell(z)$ et $F(z)$ sont holomorphes au voisinage de $|z| = 1$ et $F(z)$ est strictement positive sur $|z| = 1$; on peut alors poser

$$\hat{\omega}(\xi) = \sum_\ell \frac{\bar{\alpha}_\ell(\xi)}{\alpha(\xi)} \hat{\varphi}_\ell(\xi)$$

et on a $\omega \in E_0 \cap V$ et $\langle \omega(x), \omega^*(x-k) \rangle = \delta_{0,k}$. La famille $(\omega(x-k))$ est alors une famille de Riesz

$$\left(\sum |\hat{\omega}(\xi + 2k\pi)|^2 = \sum_\ell \frac{|\bar{\alpha}_\ell(\xi)|^2}{\alpha(\xi)^2} = \frac{1}{\alpha(\xi)} \right)$$

et engendre un sous-espace W de V.

On pose alors $X = V \cap (W^*)^\perp$ et $X^* = V^* \cap W^\perp$. On a alors

α) $V = W \oplus X$ et $V^* = W^* \oplus X^*$; en effet on note Q le projecteur sur W parallèlement à $(W^*)^\perp$:

$$Qf = \sum_k \langle f, \omega^*(x-k) \rangle \omega(x-k);$$

alors si $f \in V$, $Qf \in W$ et $(I-Q)f \in X$; de même si $f \in V^*$, on a $Q^*f \in W^*$ et $(I-Q^*)f \in X^*$.

β) $L^2 = X \oplus (X^*)^\perp$: en effet si $R = (I-Q)P$, on a $R \circ R = R$ car $QP = PQ = Q$ ($PQ = Q$ car $W \subset V$; $QP = Q$ car $Q^* = P^*Q^*$ puisque $W^* \subset V^*$) et R est donc un projecteur; on a $R|_X = I$ et $RL^2 \subset X$ d'où $\operatorname{Im} R = X$; de même $\operatorname{Im} R^* = X^*$ et donc $\operatorname{Ker} R = (X^*)^\perp$.

γ) X et X^* sont évidemment invariants par translation entière puisque V, V^*, W et W^* le sont (et donc aussi $W^\perp$ et $(W^*)^\perp$).

δ) $E_0 \cap X$ est dense dans X car $X = (I-Q)V$ et $(I-Q)E_0 \subset E_0$; de même $E_0 \cap X^*$ est dense dans X^*.

ε) enfin $\dim \Pi(E_0 \cap X) = \dim \Pi(E_0 \cap V) - 1$ car si $\Omega_\ell(x-k)$, $1 \leq \ell \leq \dim \Pi(E_0 \cap X)$, $k \in \mathbb{Z}$, est une base orthonormée de X alors $\Pi\Omega_\ell$, $1 \leq \ell \leq \dim \Pi(E_0 \cap X)$, augmentée de $\Pi\omega$ est une base de $\Pi(E_0 \cap V)$.

En itérant cette construction, on arrive à $X = \{0\}$ et alors $X^* = \{0\}$ de sorte qu'on a exhibé des fonctions $\varphi_1, \ldots, \varphi_L, \varphi_1^*, \ldots, \varphi_L^*$ ($L = \dim \Pi(E_0 \cap V)$) qui vérifient j) à jjj). Le point a) est donc démontré.

Le point b) est immédiat. La seule chose à vérifier est que $\dim \Pi(E_0 \cap V)$ est finie. ($E_0 \cap V$ est dense dans V car $V = PL^2$ et $PE_0 \subset E_0$). Mais pour cela il suffit de définir

$$\Pi P f = \int_0^1 \Big(\sum_{k \in \mathbb{Z}} p(x-k, y) \Big) f(y) dy$$

et de remarquer que ΠP est compacte sur $L^2([0,1])$ (estimations de taille et de régularité du noyau) et que $\Pi P(\Pi(E_0 \cap V)) = \Pi(E_0 \cap V)$. En effet la fonction $\sum_{k \in \mathbb{Z}} p(x-k, y)$ vérifie les estimations suivantes

$$\text{(14.1)} \qquad \text{Il existe } C > 0 \text{ tel que } \forall x, y, \quad |\sum_{k \in \mathbb{Z}} p(x-k, y)| \leq C$$

$$\text{(14.2)} \qquad \begin{aligned} &\text{Il existe } C > 0 \text{ et } \beta > 0 \text{ tel que } \forall x, y, \ \forall h \in [-1, 1] \\ &\qquad |\sum_{k \in \mathbb{Z}} p(x-k, y) - \sum_{k \in \mathbb{Z}} p(x+h-k, y)| \leq C\, |h|^{\beta}. \end{aligned}$$

Si g_n est une suite bornée de fonctions de $L^2([0,1])$, les fonctions $\Pi P(g_n)$ sont uniformément bornées et uniformément Höldériennes sur $[0,1]$; le Théorème d'Ascoli permet de conclure que $(\Pi P(g_n))$ admet une sous-suite uniformément convergente sur $[0,1]$, et donc convergente dans $L^2([0,1])$. ΠP est bien compact et donc $\dim \Pi P(E_0 \cap V) < +\infty$.

De la Proposition 2, on déduit alors facilement le théorème

Théorème 3. a) *Soit V_j, V_j^* deux analyses multi-résolutions de $L^2(\mathbb{R})$ de même facteur de dilatation A telles que $E_0 \cap V_0$ soit dense dans V_0, $E_0 \cap V_0^*$ soit dense dans V_0^* et $L^2 = V_0 \oplus (V_0^*)^{\perp}$. Alors elles ont la même multiplicité d et elles admettent des fonctions φ_ℓ, $1 \leq \ell \leq d$, et φ_ℓ^*, $1 \leq \ell \leq d$, telles que*

- $\varphi_\ell \in E_0 \cap V_0$, $\varphi_\ell^* \in E_0 \cap V_0^*$,
- *Les $\varphi_\ell(x-k)$, $1 \leq \ell \leq d$, $k \in \mathbb{Z}$, forment une base de Riesz de V_0, et les $\varphi_\ell^*(x-k)$, $1 \leq \ell \leq d$, $k \in \mathbb{Z}$ une base de Riesz de V_0^* ,*
- $\langle \varphi_\ell(x-k), \varphi_{\ell'}^*(x-k') \rangle = \delta_{\ell,\ell'} \delta_{k,k'}$.

(*fonctions d'échelle duales*).

On définit $W_0 = V_1 \cap (V_0^*)^\perp$ *et* $W_0^* = V_1^* \cap V_0^\perp$. *Alors*

$$L^2 = W_0 \oplus (W_0^*)^\perp$$

et il existe des bases de Riesz duales de W_0 *et* W_0^* *de la forme*

$$\psi_\varepsilon(x-k), \qquad 1 \leq \varepsilon \leq (A-1)d;\ k \in \mathbb{Z},$$

et

$$\psi_\varepsilon^*(x-k), \qquad 1 \leq \varepsilon \leq (A-1)d;\ k \in \mathbb{Z},$$

avec ψ_ε, $\psi_\varepsilon^* \in E_0$.

Si de plus les φ_ℓ *et les* φ_ℓ^* *sont de classe* E_α *pour un* $\alpha > 0$ *alors les familles*

$$(A^{j/2}\psi_\varepsilon(A^j x - k))_{1 \leq \varepsilon \leq (A-1)d;\ j,k \in \mathbb{Z}}$$

et

$$(A^{j/2}\psi_\varepsilon^*(A^j x - k))_{1 \leq \varepsilon \leq (A-1)d;\ j,k \in \mathbb{Z}}$$

forment des bases inconditionnelles bi-orthogonales de $L^2(\mathbb{R})$.

b) *Inversement si les fonctions* ψ_e *et* ψ_ε^* *sont de classe* E_α *et si les familles*

$$(A^{j/2}\psi_\varepsilon(A^j x - k))_{1 \leq \varepsilon \leq E;\ j,k \in \mathbb{Z}}$$

et

$$(A^{j/2}\psi_\varepsilon^*(A^j x - k)_{1 \leq \varepsilon \leq E;\ j,k \in \mathbb{Z}}$$

forment des bases inconditionnelles bi-orthogonales de $L^2(\mathbb{R})$ *alors elles proviennent d'analyses multi-résolutions bi-orthogonales de facteur de dilatation* A, *de multiplicité* $d = E/(A-1)$ *et à fonctions d'échelle duales* $\varphi_1, \ldots, \varphi_d$, $\varphi_1^*, \ldots, \varphi_d^*$ *de classe* E_α .

Le Théorème 3 a une démonstration exactement analogue à celle des divers théorèmes 1. Dans le sens des fonctions d'échelle aux ondelettes, on se borne à vérifier que $L^2 = W_0 \oplus (W_0^*)^\perp$ (car les projecteurs obliques P_0 sur V_0 parallèlement à $(V_0^*)^\perp$ et P_1 sur V_1 parallèlement à $(V_1^*)^\perp$ vérifient $P_1 P_0 = P_0$ ($V_0 \subset V_1$) et $P_0 P_1 = P_0$ ($V_0^* \subset V_1^*$), de sorte que $Q_0 = P_1 - P_0$ est un projecteur; or $W_0 = \operatorname{Im} Q_0$ et $(W_0^*)^\perp = \operatorname{Ker} Q_0$), que W_0 et W_0^* sont $\mathbb{Z}$-invariants (puisque c'est le cas de V_0, V_0^*, V_1 et V_1^*), que $E_0 \cap W_0$ est dense dans W_0 et $E_0 \cap W_0^*$ dans W_0^* (car $P_0 E_0 \subset E_0$ et $P_1 E_0 \subset E_0$) et enfin que $\dim \Pi(E_0 \cap W_0) \leq \dim \Pi(E_0 \cap V_0) < +\infty$; la Proposition 2 permet alors de conclure à

l'existence des ψ_ε, ψ_ε^*. Pour la propriété de base inconditionnelle, on remarque que si P_j est le projecteur sur V_j parallèlement à $(V_j^*)^\perp$ alors les P_j sont équicontinus (car $P_j = D_j P_0 D_j^{-1}$ où $D_j f(x) = f(2^j x)$) et donc, par densité de $\cup V_j$, que pour $f \in L^2$ on a $P_j f \to f$ dans L^2 quand $j \to +\infty$; cela force alors P_j à vérifier $P_j 1 = 1$ (condition de Strang et Fix), ou encore les φ_ℓ, φ_ℓ^* à vérifier

$$1 = \sum_{\ell=1}^{d} \left(\int \bar{\varphi}_\ell^* \, dx \right) \sum_{k \in \mathbb{Z}} \varphi_\ell(x-k) = \sum_{\ell=1}^{d} \left(\int \bar{\varphi}_\ell \, dx \right) \sum_{k \in \mathbb{Z}} \varphi_\ell^*(x-k)$$

(cela sera démontré dans le Lemme 2 ci-après) et donc les ψ_ε à vérifier $\int \psi_\varepsilon \, dx = \int \psi_\varepsilon \cdot 1 \, dx = 0$ puisque $\langle \psi_\varepsilon, \varphi_\ell^*(x-k) \rangle = 0$, et de même $\int \psi_\varepsilon^* \, dx = 0$. Si maintenant on a de plus $\psi_\varepsilon \in E_\alpha$ alors on a (*cf.* Lemme 3)

$$\Big\| \sum_\varepsilon \sum_j \sum_k A^{j/2} \, \alpha_{j,k,\varepsilon} \, \psi_\varepsilon(A^j x - k) \Big\|_2 \leq C \Big(\sum_{\varepsilon,j,k} |\alpha_{j,k,\varepsilon}|^2 \Big)^{1/2}$$

et de même pour les ψ_ε^* ; enfin on a

$$\begin{aligned} &\sum_{\varepsilon,j,k} |\alpha_{j,k,\varepsilon}|^2 \\ &= \langle \sum_{\varepsilon,j,k} A^{j/2} \, \alpha_{j,k,\varepsilon} \, \psi_\varepsilon(A^j x - k), \sum_{\varepsilon,j,k} A^{j/2} \, \alpha_{j,k,\varepsilon} \, \psi_\varepsilon^*(A^j x - k) \rangle \\ &\leq \Big\| \sum_\varepsilon \sum_j \sum_k A^{j/2} \, \alpha_{j,k,\varepsilon} \, \psi_\varepsilon(A^j x - k) \Big\|_2 \, C \, (\sum_{\varepsilon,j,k} |\alpha_{j,k,\varepsilon}|^2)^{1/2} \end{aligned}$$

et donc

$$\Big(\sum_{\varepsilon,j,k} |\alpha_{j,k,\varepsilon}|^2 \Big)^{1/2} \leq C \Big\| \sum_\varepsilon \sum_j \sum_k A^{j/2} \, \alpha_{j,k,\varepsilon} \, \psi_\varepsilon(A^j x - k) \Big\|_2 .$$

La famille $(A^{j/2} \psi_\varepsilon(A^j x - k))_{\varepsilon,j,k}$ est donc de Riesz dans L^2 ; de plus elle est totale (car $\cap V_j^* = \{0\}$; or si $f \in \cap W_j^\perp$, alors $Q_j^* f = 0$ et donc $P_k^* f = P_\ell^* f$ pour tous k, ℓ ; si $\ell \to +\infty$ on trouve $P_k^* f = f$; si $k \to -\infty$ on trouve $f \in \cap V_j^*$ et donc $f = 0$); c'est donc une base inconditionnelle de $L^2(\mathbb{R})$.

Dans le sens des ondelettes aux fonctions d'échelle, on note W_j l'espace engendré par les $\psi_{\varepsilon,j,k}$ ($1 \leq \varepsilon \leq E$, $k \in \mathbb{Z}$), W_j^* celui engendré par les $\psi^*_{\varepsilon,j,k}$ et Q_j le projecteur

$$Q_j f = \sum_{\varepsilon} \sum_{k} \langle f, \psi^*_{\varepsilon,j,k} \rangle \, \psi_{\varepsilon,j,k}$$

(où $\psi_{\varepsilon,j,k} = A^{j/2}\psi_\varepsilon(A^j x - k)$ et $\psi^*_{\varepsilon,j,k} = A^{j/2}\psi^*_\varepsilon(A^j x - k)$). Alors $Q_j Q_k = 0$ si $j \neq k$ et $Q_j Q_j = Q_j$. On désigne par P_0 l'opérateur $P_0 = \sum_{j \leq -1} Q_j$; alors P_0 est continu (car $(\psi_{\varepsilon,j,k})$ et $(\psi^*_{\varepsilon,j,k})$ sont des bases inconditionnelles de L^2) ; de plus P_0 est un projecteur, d'après les propriétés des Q_j citées plus haut et $P_0[f(x-1)] = [P_0 f](x-1)$ car $P_0 = I - \sum_{j \geq 0} Q_j$ et les Q_j sont $\mathbb{Z}$-invariants pour $j \geq 0$. Le noyau $p(x,y)$ de P_0 vérifie les estimations kkk) et kv) de la Proposition 2, par un calcul similaire à celui effectué pour démontrer le Théorème-1bis. On peut alors conclure et le Théorème 3 est démontré.

Il reste cependant à vérifier les lemmes 2 et 3 (qui sont classiques).

Lemme 2. *Si (V_j), (V_j^*) sont deux analyses multi-résolution bi-orthogonales alors $P_0 1 = 1$.*

En effet on a

$$P_j f = \mu(2^j x) f(x) + \sum_{\ell} \sum_{k} A^j \int \bar{\varphi}^*_\ell(A^j y - k)\big(f(y) - f(x)\big)\, dy \;\; \varphi_\ell(A^j x - k)$$

où

$$\mu = P_0 1 = \sum_{\ell} \sum_{k} \Big(\int \bar{\varphi}^*_\ell \, dy \Big) \varphi_\ell(x - k) .$$

Le second terme tend vers 0 dans L^2, car les opérateurs

$$T_j f = P_j f - \mu(2^j x) f$$

sont équicontinus sur L^2 et lorsque $f \in C_0^\infty$ il est immédiat que $T_j f \to 0$: $\|T_j f\|_\infty \leq C A^{-j}$ d'une part, et d'autre part lorsque

$$|x| \geq 2 \max_{y \in \operatorname{supp} f} |y| = x_0$$

on a

$$|f(x)| \leq C A^j e^{-DA^j|x|},$$

et donc

$$\int_{|x|\geq x_0} |T_j f|^2 dx \leq \int_{|x|\geq A^j x_0} C\, e^{-2D|x|}\, dx \to 0$$

car $x_0 > 0$ (si $f \neq 0$). Comme $P_j f \to f$, on obtient $\mu(2^j x) f \to f$. On teste cela sur $f = 1_{[0,1]}$

$$0 = \lim_{j\to+\infty} \int_0^1 |1-\mu(2^j x)|^2 dx = \lim_{j\to+\infty} \frac{1}{2^j} \int_0^{2^j} |1-\mu(x)|^2 dx$$
$$= \int_0^1 |1-\mu(x)|^2 dx$$

puisque $\mu(x-1) = \mu(x)$; cela implique que $1 = \mu(x)$.

Lemme 3. *Si $\theta \in E_\alpha$ et $\int \theta\, dx = 0$ alors on a*

$$\Big\| \sum_j \sum_k \lambda_{j,k}\, A^{j/2}\, \theta(A^j x - k) \Big\|_2 \leq C \Big(\sum_j \sum_k |\lambda_{j,k}|^2 \Big)^{1/2}.$$

En effet, on calcule la matrice de Gram des $\theta_{j,k} = A^{j/2}\, \theta(A^j x - k)$: si $j \geq \ell$ on écrit

$$\langle \theta_{j,k}, \theta_{\ell,p} \rangle = \int A^{(j+\ell)/2}\, \theta(A^j x - k)\, (\theta(A^\ell x - p) - \theta(A^\ell \frac{k}{A^j} - p))\, dx$$

et donc

$$|\langle \theta_{j,k}, \theta_{\ell,p} \rangle| \leq C\, A^{(j+\ell)/2} \int e^{-D|A^j x - k|}\, |\theta(A^\ell x - p) - \theta(A^\ell \frac{k}{A^j} - p)|\, dx.$$

On obtient

$$|\langle \theta_{j,k}, \theta_{\ell,p} \rangle|$$
$$\leq C\, A^{(\ell-j)/2} \int e^{-D|x|}\, |\theta(A^{\ell-j} x + A^{\ell-j} k - p) - \theta(A^{\ell-j} k - p)|\, dx.$$

• Si $|A^{\ell-j} x| \leq 1$ ou si $|A^{\ell-j} x + A^{\ell-j} k - p| \geq |A^{\ell-j} k - p|/2$ on obtient

$$|\theta(A^{\ell-j} x + A^{\ell-j} k - p) - \theta(A^{\ell-j} k - p)| \leq (A^{\ell-j})^\alpha |x|^\alpha e^{-D|A^{\ell-j} k - p|/2};$$

• Si $|A^{\ell-j}x| \geq 1$ et $|A^{\ell-j}x| \geq |A^{\ell-j}k-p|/2$, on obtient

$$e^{-D|x|} \leq (A^{\ell-j})^{\alpha}|x|^{\alpha}e^{-D|x|/2}e^{-D|A^{\ell-j}k-p|/4}$$

et au total

$$(14) \quad |\langle \theta_{j,k}, \theta_{\ell,p}\rangle| \leq C\, A^{(\ell-j)(1/2+\alpha)}e^{-D|A^{\ell-j}k-p|/4} \qquad \text{pour } \; j \geq \ell\,.$$

On vérifie alors que

$$\begin{aligned}\sum_{j,k} A^{(\ell-j)/2}|\langle \theta_{j,k}, \theta_{\ell,p}\rangle| &\leq C\Big(\sum_{j\geq\ell}\sum_{k} A^{(\ell-j)(\alpha+1)}e^{-D|A^{\ell-j}k-p|/4} \\ &\qquad + \sum_{j<\ell}\sum_{k} A^{(j-\ell)\alpha}e^{-D|A^{j-\ell}p-k|/4}\Big) \\ &\leq C'\Big(\sum_{j\geq\ell} A^{(\ell-j)(\alpha+1)}A^{j-\ell} + \sum_{j<\ell} A^{(j-\ell)\alpha}\Big) \\ &\leq C''\end{aligned}$$

de sorte que

$$\begin{aligned}\Big|\sum_{j,k}\sum_{\ell,p} a_{j,k}\, b_{j,p}\langle \theta_{j,k}, \theta_{\ell,p}\rangle\Big| &\leq \Big(\sum_{j,k}\sum_{\ell,p} |a_{j,k}|^2\, |\langle \theta_{j,k}, \theta_{\ell,p}\rangle| A^{(j-\ell)/2}\Big)^{1/2} \\ &\qquad \cdot \Big(\sum_{j,k}\sum_{\ell,p} |b_{\ell,p}|^2\, |\langle \theta_{j,k}, \theta_{\ell,p}\rangle|^2 A^{(\ell-j)/2}\Big)^{1/2} \\ &\leq C'\Big(\sum_{j,k} |a_{j,k}|^2\Big)^{1/2}\Big(\sum_{\ell,p} |b_{\ell,p}|^2\Big)^{1/2}\end{aligned}$$

et le Lemme 3 est démontré.

5. Ondelettes bi-orthogonales à support compact.

Nous allons traiter de même le cas des bases bi-orthogonales d'ondelettes à support compact.

Proposition 2-bis.

a) *Soit* $P : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ *un opérateur continu tel que*

i) $P \circ P = P$ (P *est un projecteur*) *et tel que* $Pf = \int p(x,y) f(y)\, dy$ *où,*

ii) $p(x,y) = p(x-1, y-1)$ ($\mathbb{Z}$-*invariance*),

iii) $p(x,y) = 0$ *si* $|x-y| \geq M$ (*où* M *est une constante* ≥ 0).

Alors si $\mathrm{Ind}_{\mathbb{Z}} \operatorname{Im} P < +\infty$ *il existe des fonctions* φ_ℓ *et* φ_ℓ^*, $1 \leq \ell \leq \mathrm{Ind}_{\mathbb{Z}} \operatorname{Im} P$, *telles que*

j) $\varphi_\ell \in \operatorname{Im} P$, $\varphi_\ell^* \in (\operatorname{Ker} P)^\perp$ *et* φ_ℓ, φ_ℓ^* *sont à support compact,*

jj) $(\varphi_\ell(x-k))_{\ell,k}$ *est une base de Riesz de* $\operatorname{Im} P$ *et* $(\varphi_\ell^*(x-k))_{\ell,k}$ *est une base de Riesz de* $(\operatorname{Ker} P)^\perp$,

jjj) $\langle \varphi_\ell(x-k), \varphi_{\ell'}^*(x-k') \rangle = \delta_{\ell,\ell'}\, \delta_{k,k'}$,

de sorte que le projecteur P *se décompose en*

$$Pf = \sum_{\ell=1}^{\mathrm{Ind}_{\mathbb{Z}} \operatorname{Im} P} \sum_{k \in \mathbb{Z}} \langle f, \varphi_\ell^*(x-k) \rangle \varphi_\ell(x-k)\,. \tag{15}$$

b) *Si* P *vérifie* i), ii), iii) *et si de plus, pour deux constantes* $C > 0$ *et* $\beta \in]0,1]$, *on a*

$$|p(x,y) - p(x+h,y)| \leq C\, |h|^\beta \tag{16}$$

alors $\mathrm{Ind}_{\mathbb{Z}} \operatorname{Im} P < +\infty$.

Si $P \neq 0$, il existe $\omega \in \operatorname{Im} P \cap L^2_{\mathrm{comp}}$ avec $\omega \neq 0$. On appelle W le sous-espace de L^2 engendré par les $\omega(x-k)$, $k \in \mathbb{Z}$. La démonstration du Théorème 2 ter nous a montré que W admettait une base de Riesz de la forme $\omega_0(x-k)$, $k \in \mathbb{Z}$, avec $\omega_0 \in L^2_{\mathrm{comp}}$, et donc que $\mathrm{Ind}_{\mathbb{Z}} W = 1$. Pour $f \in W \cap L^2_{\mathrm{comp}}$, on note

$$P_f(z) = \sum_{k \in \mathbb{Z}} \langle f, f(x-k) \rangle z^k$$

(de sorte que $P_f(e^{-i\xi})$ est le polynôme trigonométrique que nous avons déjà utilisé à plusieurs reprises) et on note φ un élément de $W \cap L^2_{\mathrm{comp}}$ tel que P_φ soit non nul ($\varphi \neq 0$) et de degré minimum. Alors $P_\varphi(z)$ ne s'annule pas sur le cercle-unité car nous avons vu que si $P_\varphi(e^{-i\xi_0}) = 0$ et si $\hat{\psi} = \hat{\varphi}/(e^{-i\xi} - e^{-\xi_0})$ alors $\psi \in W \cap L^2_{\mathrm{comp}}$ et

$$P_\varphi(e^{-i\xi}) = |e^{-i\xi} - e^{-i\xi_0}|^2 P_\psi(e^{-i\xi})\,,$$

ce qui contredit la minimalité de P_φ.

Pour $\gamma \in C_0^\infty$ on pose

$$Q_\gamma(z) = \sum_{k\in\mathbb{Z}} \langle \gamma, \varphi(x-k)\rangle z^k .$$

Alors l'ensemble des $Q_\gamma(e^{-i\xi})$ est un idéal de l'anneau des polynômes trigonométriques: si Q est un polynôme trigonométrique $\sum a_k e^{-ik\xi}$ et si $\eta = \sum a_k \gamma(x-k)$ alors $Q_\eta = QQ_\gamma$. Soit P_0 le générateur de cet idéal. Alors $P_0(z)$ ne s'annule pas sur le cercle-unité, puisque $Q_\varphi = P_\varphi$ ne s'y annule pas. Si P_0 s'annulait en un point z_0, alors nécessairement $\sum_{k\in\mathbb{Z}} \varphi(x-k) z_0^k$ serait nul dans $\mathcal{D}'$. La fonction

$$\theta = \sum_{k\geq 0} z_0^k \varphi(x-k) = -\sum_{k<0} z_0^k \varphi(x-k)$$

serait alors à support compact; de plus $\theta \in W$ car la première série converge dans L^2 si $|z_0| < 1$, et la seconde si $|z_0| > 1$. Enfin on a $\theta - z_0\theta(x-1) = \varphi$ et donc

$$P_\varphi(e^{-i\xi}) = P_\theta(e^{-i\xi})\, |1 - z_0 e^{-i\xi}|^2 ,$$

ce qui contredit la minimalité de P_φ. On obtient donc que P_0 ne s'annule pas, et donc $P_0 = 1$.

On fixe alors γ_0 tel que $P_{\gamma_0} = 1$ et $\varphi^* = P^*(\gamma_0)$. Alors $\varphi^* \in (\operatorname{Ker} P')^\perp \cap L^2_{\text{comp}}$ et

$$\langle \varphi^*, \varphi(x-k)\rangle = \langle \gamma_0\,, \varphi(x-k)\rangle = \delta_{0,k} .$$

On pose alors

$$\tilde{P} f = Pf - \sum_{k\in\mathbb{Z}} \langle f, \varphi^*(x-k)\rangle\, \varphi(x-k) .$$

$\tilde{P}$ est un projecteur et $\operatorname{Im}\tilde{P} = \operatorname{Im} P \cap \{\varphi^*(x-k)\}^\perp$, de sorte que $\operatorname{Ind}_{\mathbb{Z}} \operatorname{Im}\tilde{P} = \operatorname{Ind}_{\mathbb{Z}} \operatorname{Im} P - 1$ et que $\operatorname{Im} P = \{\varphi(x-k)\} \oplus \operatorname{Im}\tilde{P}$. Le noyau de $\tilde{P}$ est encore $\mathbb{Z}$-invariant et proprement supporté. En itérant la construction, on obtient les bases φ_ℓ, φ_ℓ^*.

Le point b) de la Proposition 2-bis est un cas particulier de la Proposition 2 et la Proposition 2-bis est donc démontrée.

On a alors le théorème

Théorème 3-bis.

a) *Soit V_j, V_j^* deux analyses multi-résolution de $L^2(\mathbb{R})$ de même facteur de dilatation A et de même multiplicité d. On suppose que de plus elles admettent des fonctions φ_ℓ, φ_ℓ^*, $1 \leq \ell \leq d$, telles que*

- *φ_ℓ et φ_ℓ^* sont à support compact,*
- *les $\varphi_\ell(x-k)$, $1 \leq \ell \leq d$, $k \in \mathbb{Z}$, forment une base de Riesz de V_0 et les $\varphi_\ell^*(x-k)$, $1 \leq \ell \leq d$, $k \in \mathbb{Z}$, forment une base de Riesz de V_0^*,*
- *$\langle \varphi_\ell(x-k), \varphi_{\ell'}^*(x-k') \rangle = \delta_{\ell,\ell'}\, \delta_{k,k'}$.*

On définit $W_0 = V_1 \cap (V_0^*)^\perp$ et $W_0^* = V_1^* \cap V_0^\perp$. Alors il existe des bases de Riesz duales de W_0 et W_0^* de la forme

$$\psi_\varepsilon(x-k), \qquad 1 \leq \varepsilon \leq (A-1)d;\ k \in \mathbb{Z},$$

et

$$\psi_\varepsilon^*(x-k), \qquad 1 \leq \varepsilon \leq (A-1)d;\ k \in \mathbb{Z},$$

avec ψ_ε, ψ_ε^* à support compact.

Si de plus les φ_ℓ et les φ_ℓ^* sont de classe C^α pour un $\alpha > 0$ alors les familles

$$(A^{j/2}\psi_\varepsilon(A^j x - k))_{1 \leq \varepsilon \leq (A-1)d;\ j,k \in \mathbb{Z}}$$

et

$$(A^{j/2}\psi_\varepsilon^*(A^j x - k))_{1 \leq \varepsilon \leq (A-1)d;\ j,k \in \mathbb{Z}}$$

forment des bases inconditionnelles bi-orthogonales de $L^2(\mathbb{R})$.

b) *Inversement si les fonctions ψ_ε et ψ_ε^* sont à support compact et de classe C^α pour un $\alpha > 0$ et si les familles $(A^{j/2}\psi_\varepsilon(A^j x - k))$ et $(A^{j/2}\psi_\varepsilon^*(A^j x - k))$ $(1 \leq \varepsilon \leq E;\ j,k \in \mathbb{Z})$ forment des bases inconditionnelles bi-orthogonales de $L^2(\mathbb{R})$ alors elles proviennent d'analyses multi-résolutions bi-orthogonales de facteur de dilatation A, de multiplicité $d = E/(A-1)$ et à fonctions d'échelle duales $\varphi_1, \ldots, \varphi_d$, $\varphi_1^*, \ldots, \varphi_d^*$ à support compact et de classe C^α.*

Ce théorème se montre de manière analogue aux théorèmes 1 et 3.

6. Annexe: Le Théorème de Gröchenig, corollaires et variantes.

Le Théorème de Gröchenig concerne les analyses multi-résolutions multi-dimensionnelles. Plus précisément, on considère une suite (V_j) de sous-espaces fermés de $L^2(\mathbb{R}^n)$ tels que

(5.1) $V_j \subset V_{j+1}$, $\bigcap_{j\in\mathbb{Z}} V_j = \{0\}$, $\bigcup_{j\in\mathbb{Z}} V_j$ est dense dans $L^2(\mathbb{R}^n)$,

(5.2) $f \in V_j$ si et seulement si $f(2x) \in V_{j+1}$,

(5.3) V_0 a une base orthonormée de la forme $\varphi(x-k)$, $k \in \mathbb{Z}^n$, où φ est à décroissance rapide $(\forall \alpha \in \mathbb{N}^n,\ x^\alpha \varphi \in L^2)$.

On note W_0 le complémentaire orthogonal de V_0 dans V_1 et on veut montrer que W_0 admet une base orthonormée de la forme $\psi_\varepsilon(x-k)$, $1 \le \varepsilon \le 2^n - 1$, $k \in \mathbb{Z}^n$, où les ψ_ε sont à décroissance rapide. Pour cela, on note $r_1, \ldots, r_{2^n}$ des représentants de $\mathbb{Z}^n/2\,\mathbb{Z}^n$, de sorte qu'on dispose d'une base orthonormée de V_1 constituée des $\omega_j(x-k)$, $1 \le j \le 2^n$, $k \in \mathbb{Z}^n$, où $\omega_j = 2^{n/2}\varphi(2\,x - r_j)$. Comme $\varphi \in V_1$, φ se décompose sur les $\omega_j(x-k)$ et on a

$$\hat{\varphi} = \sum_{j=1}^{2^n} \omega_j(\xi)\,\hat{\varphi}_j(\xi)$$

où les ω_j sont C^∞ et $2\pi\mathbb{Z}^n$-périodique. Or les familles $(f(x-k))$, $(g(x-k))$, $k \in \mathbb{Z}^n$, sont orthonormées pour $f, g \in V_1$, $(\hat{f} = \sum f_j(\xi)\hat{\varphi}_j(\xi)$, $\hat{g} = \sum g_j(\xi)\hat{\varphi}_j(\xi))$ si et seulement si on a: $\sum_j |f_j(\xi)|^2 = 1$, $\sum_j |g_j(\xi)|^2 = 1$, $\sum_j f_j(\xi)\bar{g}_j(\xi) = 0$ (puisque

$$\int f(x-k)\bar{g}(x)\,dx = \frac{1}{(2\pi)^n}\int_{[0,2\pi]^n} \sum_j f_j(\xi)\bar{g}_j(\xi)\,e^{-ik\xi}\,d\xi)\,.$$

Il s'agit donc de compléter de manière unitaire une matrice carrée de taille 2^n dont on connaît la première colonne ; les coefficients de la matrice doivent être des fonctions C^∞ sur $\mathbb{R}^n$, $2\pi\mathbb{Z}^n$-périodiques, ou encore des fonctions C^∞ sur $\mathbb{T}^n = S^1 \times \cdots \times S^1 = \mathbb{R}^n/2\pi\mathbb{Z}^n$. Le Théorème de Gröchenig est alors le suivant

Théorème 4. *Si $m : \mathbb{T}^n \to S^{2q-1} = \{z \in \mathbb{C}^q : \|z\|^2 = 1\}$ est C^∞, et si $n < 2q-1$, alors il existe une application: $M : \mathbb{T}^n \to U(q)$ de classe C^∞ telle que le premier vecteur colonne de M soit m.*

En effet, puisque m est C^∞ et que $n < 2q-1$, $m(\mathbb{T}^n)$ est de mesure nulle sur S^{2q-1} et donc n'est pas surjective. On peut toujours supposer que $(0,\dots,0,1)$ n'est pas atteint. Alors le déterminant de la matrice

$$M_\alpha = \begin{pmatrix} m_1 & \alpha & & 0 \\ \vdots & 0 & & \alpha \\ m_q & -\bar{m}_1 & \dots & -\bar{m}_{q-1} \end{pmatrix}$$

vaut

$$-\alpha^{q-1} m_q + (-1)^{q-1}\alpha^{q-2}(|m_1|^2 + \dots + |m_{q-1}|^2 \\ = (-1)^{q-1}\alpha^{q-2}(1 - |m_q|^2 + (-1)^q \alpha\, m_q)$$

et est donc nul pour $(-1)^{q-1}\alpha \in]0,\varepsilon[$, ε assez petit; en effet si ξ_ε est tel que

$$1 - |m_q(\xi_\varepsilon)|^2 - \varepsilon\, m_q(\xi_\varepsilon) = 0,$$

alors $|m_q(\xi_\varepsilon)| \to 1$ pour $\varepsilon \to 0$ et $m_q(\xi_\varepsilon) \in \mathbb{R}_+$ d'où $m_q(\xi_\varepsilon) \to 1$, ce qui est impossible. Il suffit ensuite d'orthonormaliser les vecteurs colonnes de M_α pour obtenir M. Le Théorème 4 est donc démontré.

Si la fonction φ est à localisation exponentielle, les fonctions m_j ont leurs coefficients de Fourier à décroissance exponentielle et il en va de même pour M_α. Cette propriété reste stable sous l'orthonormalisation de Gram-Schmidt des colonnes de M_α (la vitesse de décroissance exponentielle pouvant être modifiée), et on trouve des ondelettes ψ_ε à localisation exponentielle.

Si V_0 admet une base de Riesz de la forme $g(x-k)$, $k \in \mathbb{Z}^n$, avec g à support compact, on exprime les vecteurs de V_1 dans la base $2^{n/2} g(2(x-k) - r_j)$. En particulier, si $(\gamma(x-k))$ est une base de Riesz de V_0 où γ s'exprime à l'aide d'un nombre fini des $2^{n/2} g_j(x-k)$ (on peut prendre par exemple

$$\hat{\gamma} = \prod_{j=1}^{2^n} A\left(\frac{\xi}{2} + r_j \pi\right) \hat{g}(\xi)$$

avec

$$A(\xi) = \sum \langle g(x), g(x-k)\rangle e^{-ik\xi} = \sum |\hat{g}(\xi + 2k\pi)|^2 ;$$

A ne s'annule jamais, il en va donc de même du produit $\prod_{j=1}^{2^n} A(\xi/2 + r_j\pi)$ qui est $2\pi\mathbb{Z}^n$-périodique, de sorte que les $\gamma(x-k)$ forment bien une base de Riesz de V_0 ; enfin si

$$\prod_{j=2}^{2^n} A\left(\frac{\xi}{2} + r_j\pi\right) \hat{g}(\xi) = \hat{\Gamma}(\xi)$$

et si $r_1 = 0$, on a

$$\gamma = \sum_{j=1}^{2^n} \sum_{k\in\mathbb{Z}^n} \langle \Gamma, g_{j,k}\rangle\, g_{j,k}$$

avec

$$g_{j,k} = 2^{n/2} g(2(x-k) - r_j)\,;$$

Γ est à support compact et la somme ne contient qu'un nombre fini de termes non nuls), on doit compléter dans $GL(2^n, \mathbb{C})$ une matrice dont le premier vecteur colonne est à coefficients polynômes trigonométriques. La matrice M_α est alors elle-même à coefficients polynômes trigonométriques. On a donc complété $\gamma(x-k)$ à l'aide de fonctions $\gamma_\varepsilon(x-k)$, $1 \le \varepsilon \le 2^n - 1$, $k \in \mathbb{Z}^n$, avec γ_ε à support compact pour former une base de Riesz de V_1. Le procédé d'orthonormalisation décrit dans [1] et [9] permet alors d'obtenir une base de Riesz de W_0 de la forme

$$\Gamma_\varepsilon(x-k), \qquad 1 \le \varepsilon \le 2^n - 1,\ k \in \mathbb{Z}^n,$$

avec Γ_ε à support compact.

Le Théorème de Gröchenig s'applique également immédiatement au cas des analyses multi-résolutions de $L^2(\mathbb{R})$ de facteur de dilatation A et de multiplicité d. On est alors amené à compléter une application m de $\mathbb{T}^1$ dans $(S^{2Ad-1})^d$ telle que les vecteurs $m_1, \ldots, m_d$ soient orthogonaux deux à deux dans $\mathbb{C}^{Ad}$ en une application M de $\mathbb{T}^1$ dans $U(Ad)$; c'est exactement le cas de figure du Théorème de Gröchenig. On complète d'abord l'application m_1 en une application M_1. Mais alors les applications $m_2, \ldots, m_d$ (prenant des valeurs orthogonales à m_1) s'expriment à l'aide des vecteurs colonnes de M_1 orthogonaux à m_1 ; on est alors ramené à une application à valeurs dans $(S^{2Ad-3})^{d-1}$, et ainsi de suite.

Le Théorème de Gröchenig permet alors de passer des fonctions d'échelle aux ondelettes dans le cas des fonctions d'échelle à décroissance rapide, ou à localisation exponentielle (Théorème 1-bis, Théorème 3) et

dans le cas où V_0 admet une base de Riesz à support compact (Théorème 1-ter). Mais il ne permet pas de traiter le cas des fonctions d'échelle à support compact (Théorème 1) ni de passer des ondelettes aux fonctions d'échelle.

Conclusion.

Nous avons développé une méthode alternative au Théorème de Gröchenig pour l'étude des bases de V_0 et W_0 dans le cadre des analyses multi-résolutions uni-dimensionnelles. Cette méthode, basée sur l'étude des projecteurs associés, a permis de traiter le cas des fonctions d'échelle à support compact et le passage des ondelettes aux fonctions d'échelle. Mais elle ne paraît pas transposable aux dimensions supérieures, que ce soit pour l'idée de la propriété de minimisation utilisée dans la démonstration du Théorème 2 ou que ce soit pour le lemme "d'expulsion des zéros" (Lemme 1).

REMARQUE. Y. Meyer m'a signalé un argument plus simple pour prouver la compacité des projecteurs périodisés ΠP (de noyau $\sum_{k\in\mathbb{Z}} p(x-k,y)$) intervenant tout au long de cet article. En fait, le projecteur ΠP est compact parce qu'il est un opérateur de Hilbert-Schmidt:

$$\int_0^1 \int_0^1 |\sum_{k\in\mathbb{Z}} p(x-k,y)|^2 dx\, dy < +\infty\,.$$

Cela permet d'éliminer presque toute hypothèse de régularité sur les ondelettes ψ_ε.

Par exemple si $(A^{j/2}\psi_\varepsilon(A^j x-k))$ est une base orthonormée de $L^2(\mathbb{R})$ $(1\le\varepsilon\le E,\ j,k\in\mathbb{Z})$ avec ψ_ε à support compact et $\psi_\varepsilon\in L^{2+\beta}$ pour un $\beta>0$, alors l'opérateur ΠP est de Hilbert-Schmidt: si les ψ_ε sont à support compact contenu dans $[-M,M]$, le noyau

$$p(x,y)=\sum_{j\le -1}\sum_{k\in\mathbb{Z}}\sum_{\varepsilon=1}^{E} A^j\bar{\psi}_\varepsilon(A^j y-k)\psi_\varepsilon(A^j x-k)$$

est nul si $|x-y|\ge 2M$ et on a donc

$$\int_0^1 \int_0^1 |\sum_{k\in\mathbb{Z}} p(x-k,y)|^2 dx\, dy$$

$$\leq \int_0^1 \int_0^1 (2M+1) \sum_{k\in\mathbb{Z}} |p(x-k,y)|^2 dx\, dy$$
$$\leq (2M+1) \int_{x=-\infty}^{+\infty} \int_{y=0}^{1} |p(x,y)|^2 dy\,.$$

Mais

$$\left\| A^j \sum_{k\in\mathbb{Z}} \psi_\varepsilon(A^j x - k)\, \bar{\psi}_\varepsilon(A^j y - k) \right\|_{L^2(\mathbb{R}\times[0,1])}$$
$$\leq \sum_{k\in\mathbb{Z}} A^{j/2} \|\psi_\varepsilon(A^j y - k)\|_{L^2([0,1])}$$
$$\leq A^{j/2} \sum_{k\in\mathbb{Z}} \|\psi_\varepsilon(A^j y - k)\|_{L^{2+\beta}([0,1])}$$
$$\leq (2M+1)\, A^{j\,\beta/(2(2+\beta))}\, \|\psi_\varepsilon\|_{2+\beta}$$

et $\sum_{j\leq -1} A^{j\beta/(2(2+\beta))} < +\infty$. On procède de même dans le cas des bases bi-orthogonales.

Bibliographie.

[1] Chui, C. K., Stockler, J., Ward, J. D., Compactly supported box spline wavelets. Preprint, 1991.

[2] Cohen, A., Daubechies, I., Feauveau, J. C., Biorthogonal bases of compactly supported wavelets. *Comm. Pure Appl. Math.* **45** (1992), 485-560.

[3] Daubechies, I., Orthonormal bases of wavelets with compact support. *Comm. Pure Appl. Math.* **41** (1988), 909-996.

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

[5] Feauveau, J. C., Analyses multirésolution par ondelettes non orthogonales et bases de filtres numériques. Thèse Université Paris Sud, 1990.

[6] Goodman, T. N. T., Lee, S. L., Tang, W. S., Wavelets in wandering subspaces. A paraître aux *Trans. Amer. Math. Soc.*

[7] Gröchenig, K. H., Analyse multiéchelle et bases d'ondelettes. *C. R. Acad. Sci. Paris* (1987), 13-17.

[8] Hervé, L., Méthodes d'opérateurs quasi-compacts en analyse multirésolution, applications à la construction de bases d'ondelettes et à l'interpolation. Thèse Université Rennes I, 1992.

[9] Jia, R. Q., Micchelli, C. A., Using the refinement equation for the construction of pre-wavelets, II: Powers of two, in *Curves and surfaces,* P. J. Laurent, A. le Méhauté and L. L. Schumaker (eds.), Academic Press, 1991.

[10] Lemarié, P. G., Analyses multi-résolution et ondelettes à support compact, in *Les ondelettes en* 1989, P. G. Lemarié (ed.), Lectures notes in Math. **1438** (1990).

[11] Lemarié, P. G., Fonctions à support compact dans les analyses multi-résolution. *Revista Mat. Iberoamericana* **7** (1991), 157-182.

[12] Lemarié-Rieusset, P. G., Existence de fonctions-pères pour les ondelettes à support compact. *C. R. Acad. Sci. Paris* **314** (1992), 17-19.

[13] Lemarié-Rieusset, P. G., Sur l'existence des analyses multi-résolution en théorie des ondelettes. *Revista Mat. Iberoamericana* **8** (1992), 457-474.

[14] Lemarié, P. G., Meyer, Y., Ondelettes et bases hilbertiennes. *Revista Mat. Iberoamericana* **2** (1986), 1-18.

[15] Mallat, S., Multiresolution approximation and wavelet orthonormal bases of L^2. *Trans. Amer. Math. Soc.* **315** (1989), 69-88.

[16] Meyer, Y., Principe d'incertitude, bases hilbertiennes et algèbres d'opérateurs. *Sém. Bourbaki*, 1985-1986, n° 662.

[17] Meyer, Y., *Ondelettes et opérateurs.* Hermann, 1990.

Recibido: 10 de abril de 1.992

Pierre Gilles Lemarié-Rieusset
Département de Matématiques
Université de Paris-Sud
91405 Orsay, FRANCE

Revista Matemática Iberoamericana
Vol. 9, N.° 2, 1993

Further pseudodifferential operators generating Feller semigroups and Dirichlet forms

Niels Jacob

Abstract. We prove for a large class of symmetric pseudo differential operators that they generate a Feller semigroup and therefore a Dirichlet form. Our construction uses the Yosida-Hille-Ray Theorem and a priori estimates in anisotropic Sobolev spaces. Using these a priori estimates it is possible to obtain further information about the stochastic process associated with the Dirichlet form under consideration.

Introduction.

Generators of Feller semigroups are characterized by the Yosida-Hille-Ray Theorem, see [7] or [13]. This characterization involves the notion of the positive maximum principle. It was Ph. Courrège [6] who gave a characterization of operators satisfying this maximum principle. He proved that these operators are certain integro-differential-operators and later these results had been developed further in order to continue studies started by von Waldenfels [39]-[40]. More recent results on generators of Feller semigroups can be found in [8] or [37]. However, even in the paper [6], Courrège gave also a characterization of

the operators satisfying the positive maximum principle as pseudodifferential operators. These pseudodifferential operators have a symbol $a : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ with the property that for any $x \in \mathbb{R}^n$ the function $\xi \mapsto a(x, \xi)$ is a continuous negative definite function, see Definition 1.1. At that time the theory of pseudodifferential operators was rather young, see [18] and [28], and it seems to us that this characterization has never been used to construct Feller semigroups or Markov processes. Only recently there are some investigations using the theory of pseudodifferential operators to construct and study Markov processes. For example T. Komatsu in [29]-[32] uses the theory of elliptic standard pseudo differential operators to study perturbations of symmetric stable processes; and in [27] A. N. Kochubei uses the theory of hypersingular integrals due to S. G. Samko, see [36], to obtain Markov processes by constructing fundamental solutions of certain parabolic pseudo differential operators.

On the other hand continuous negative definite functions do enter in a quite different way in the theory of Markov processes. They do also characterize translation invariant Dirichlet forms, see [2], [9] and as a standard reference for Dirichlet forms [14]. In [23] we pointed out that it is possible to combine Hilbert space methods, which are very convenient when working with Dirichlet forms, with the result of Courrège in order to construct a Feller semigroup starting with a pseudodifferential operator having a symbol with the properties mentioned above. In [24] a special example of an elliptic pseudodifferential operator was discussed in detail.

The purpose of this paper is to give further examples of pseudo differential operators generating Feller semigroups and to study properties of corresponding objects like the associated Dirichlet form and Markov process. Now the operators under consideration are no longer elliptic pseudodifferential operators and some classical calculus of pseudodifferential operators is not applicable. While our strategy in constructing the semigroup is just the same as in [23], it was necessary to strengthen some results of [23]. In particular the commutator estimate in Section 6 improving an earlier result, see [22], plays an essential part in proving the fundamental estimates.

1. Notations and auxiliary results.

Most of our notations are standard. The spaces $C_0^\infty(\mathbb{R}^n)$, $C^m(\mathbb{R}^n)$, $0 \leq m \leq \infty$, $L^p(\mathbb{R}^n)$, $1 \leq p \leq \infty$, the Schwartz space $S(\mathbb{R}^n)$ and the Sobolev space $H^s(\mathbb{R}^n)$, $s \in \mathbb{R}$, are defined as usual. In the case $m = 0$ we write $C(\mathbb{R}^n)$ instead of $C^m(\mathbb{R}^n)$. By $C_\infty(\mathbb{R}^n)$ the space of all continuous functions $u : \mathbb{R}^n \to \mathbb{R}$ vanishing at infinity is denoted. Further we set $H^\infty(\mathbb{R}^n) = \bigcap_{s \geq 0} H^s(\mathbb{R}^n)$. The supremum norm is denoted by $\|\cdot\|_\infty$, the norm in $H^s(\mathbb{R}^n)$ is denoted by $\|\cdot\|_s$, in particular the norm in $L^2(\mathbb{R}^n)$ is $\|\cdot\|_0$ and $(\cdot,\cdot)_0$ is the scalar product in $L^2(\mathbb{R}^n)$. The norm in $L^p(\mathbb{R}^n)$, $p \neq 2, \infty$, is denoted by $\|\cdot\|_{L^p}$. For $s \in \mathbb{R}$ we define the function $\Lambda^s : \mathbb{R}^n \to \mathbb{R}$ by $\xi \mapsto \Lambda^s(\xi) = \left(1 + |\xi|^2\right)^{s/2}$ and we have $\|u\|_s = \|\Lambda^s(D)\, u\|_0$, where $\Lambda^s(D)$ is given by

$$\Lambda^s(D)\, u(x) = (2\pi)^{-n/2} \int_{\mathbb{R}^n} e^{ix\xi} \Lambda^s(\xi)\, \hat{u}(\xi)\, d\xi \,. \tag{1.1}$$

Here and in the following we denote by $\hat{u}$ the Fourier transform of u. By the Sobolev embedding theorem, see p. 121 of [33], we find for $s > m + n/2$, $m \in \mathbb{N}$, that

$$\|u\|_{C^m} \leq c\, \|u\|_s \tag{1.2}$$

holds for all $u \in H^s(\mathbb{R}^n)$, where

$$\|u\|_{C^m} = \max_{|\alpha| \leq m} \sup_{x \in \mathbb{R}^n} |\partial^\alpha u(x)| \,.$$

Moreover it follows that for $|\alpha| \leq m$ we have $\partial^\alpha u \in C_\infty(\mathbb{R}^n)$. In order that the last statement makes sense we have to identify $H^s(\mathbb{R}^n)$ with a subspace of $C^m(\mathbb{R}^n)$, which is possible by Sobolev's embedding theorem for $s > m + n/2$. Thus in this case we will always regard $H^s(\mathbb{R}^n)$ as a subspace of $C^m(\mathbb{R}^n)$. Since $C_0^\infty(\mathbb{R}^n) \subset H^s(\mathbb{R}^n)$ for all $s \geq 0$ and since $C_0^\infty(\mathbb{R}^n)$ is dense in $C_\infty(\mathbb{R}^n)$ with respect to the norm $\|\cdot\|_\infty$, it follows that for $s > n/2$ the space $H^s(\mathbb{R}^n)$ is dense in $C_\infty(\mathbb{R}^n)$ with respect to the supremum norm.

Next let us introduce the notion of a negative definite function which will be central in our paper.

Definition 1.1. *A function $a : \mathbb{R}^n \to \mathbb{C}$ is said to be negative definite if for all $m \in \mathbb{N}$ and $(x^1, \ldots, x^m)$, $x^j \in \mathbb{R}^n$, $1 \leq j \leq m$, the matrix*

$$\left(a(x^i) + \overline{a(x^j)} - a(x^i - x^j)\right)_{i,j=1,\ldots,m}$$

is positive Hermitian, i.e. if for all m-tuples $(c_1, \ldots, c_m) \in \mathbb{C}^m$

$$\sum_{i,j=1}^{m} \left(a\left(x^i\right) + \overline{a\left(x^j\right)} - a\left(x^i - x^j\right) \right) c_i \overline{c_j} \geq 0 \,. \tag{1.3}$$

A standard reference for negative definite functions is the book [1]. In particular the following lemma is proved there.

Lemma 1.1. *Let* $a : \mathbb{R}^n \to \mathbb{R}$ *be a continuous negative definite function. Then* a *is nonnegative and there exists a constant* c_a *such that*

$$a(\xi) \leq c_a \left(1 + |\xi|^2\right) \tag{1.4}$$

for all $\xi \in \mathbb{R}^n$. *Further* $a^{1/2}$ *is also a negative definite function.*

We also recall Lemma 2.1 of our paper [22].

Lemma 1.2. *Let* $a : \mathbb{R}^n \to \mathbb{R}$ *be a continuous negative definite function. Then we have for all* $\xi, \eta \in \mathbb{R}^n$

$$|a(\xi) - a(\eta)| \leq 4\, a^{1/2}(\xi)\, a^{1/2}(\xi - \eta) + a(\xi - \eta) \,. \tag{1.5}$$

As pointed out in [19, p. 327-328], (see also Section 10 of this paper) using examples of continuous negative definite functions given in [1], continuous negative definite functions need not be differentiable nor do they in general belong to classical symbol classes such as the class $S^m_{\rho,\delta}(\mathbb{R}^n)$, see [38] for the definition.

Finally let us remark that throughout this paper c denotes a nonnegative constant which may change from line to line.

2. Some function spaces.

In this section we will introduce a family of anisotropic Sobolev spaces related to a continuous negative definite function.

Definition 2.1. *Let* $a^2 : \mathbb{R}^n \to \mathbb{R}$ *be a continuous negative definite function and* $q \geq 0$ *a real number. We define the space* $H^{a^2,q}(\mathbb{R}^n)$ *by*

$$H^{a^2,q}(\mathbb{R}^n) = \left\{ u \in L^2(\mathbb{R}^n) : \int_{\mathbb{R}^n} \left(1 + a^2(\xi)\right)^{2q} |\hat{u}(\xi)|^2 d\xi < \infty \right\} \tag{2.1}$$

On $H^{a^2,q}(\mathbb{R}^n)$ we have the norm

$$\|u\|_{q,a^2}^2 = \int_{\mathbb{R}^n} \left(1 + a^2(\xi)\right)^{2q} |\hat{u}(\xi)|^2 d\xi \,. \tag{2.2}$$

With the norm (2.2) the space $H^{a^2,q}(\mathbb{R}^n)$ is a Hilbert space and $C_0^\infty(\mathbb{R}^n)$ is a dense subspace. Moreover, by Lemma 1.1 it follows that $H^{a^2,q}(\mathbb{R}^n)$ contains the space $H^{2q}(\mathbb{R}^n)$ and is contained in $L^2(\mathbb{R}^n)$. Later we will often assume that for some t, $0 < t \leq 2$, the estimate

$$c_t \left(1 + |\xi|^2\right)^{t/4} \leq \left(1 + a^2(\xi)\right)^{1/2} \tag{2.3}$$

holds. Clearly this implies that $H^{a^2,q}(\mathbb{R}^n)$ is continuously embedded into the space $H^{tq}(\mathbb{R}^n)$, *i.e.* we have

$$\|u\|_{tq} \leq c_t' \|u\|_{q,a^2} \,. \tag{2.4}$$

Thus when (2.3) holds and q is sufficiently large we can identify $H^{a^2,q}(\mathbb{R}^n)$ with a dense subspace of $C_\infty(\mathbb{R}^n)$.

The next lemma is proved as Proposition 1.5.A in [21].

Lemma 2.1. *Suppose that* $\lim_{|\xi|\to\infty} \left(1 + a^2(\xi)\right) = \infty$ *holds. Further let* $q \geq 0$ *be given and* $r > q$. *Then for any* $\eta > 0$ *there exists a constant* $c(\eta) = c\left(\eta; r, q, a^2\right)$ *such that*

$$\|u\|_{q,a^2} \leq \eta \|u\|_{r,a^2} + c(\eta) \|u\|_0 \tag{2.5}$$

for all $u \in H^{a^2,r}(\mathbb{R}^n)$.

We will need the following characterization of the dual space of $H^{a^2,q}(\mathbb{R}^n)$.

Proposition 2.1. *Let* a^2 *and* q *be as in Definition* 2.1. *Then the dual space of* $H^{a^2,q}(\mathbb{R}^n)$ *is the completion of* $L^2(\mathbb{R}^n)$ *with respect to the norm*

$$\|u\|_{-q,a^2} = \sup_{0 \neq v \in H^{a^2,q}(\mathbb{R}^n)} \frac{|(u,v)_0|}{\|v\|_{q,a^2}} \,. \tag{2.6}$$

Moreover, for $u \in L^2(\mathbb{R}^n)$ *we have*

$$\|u\|_{-q,a^2}^2 = \int_{\mathbb{R}^n} \left(1 + a^2(\xi)\right)^{-2q} |\hat{u}(\xi)|^2 d\xi \,. \tag{2.7}$$

Since $L^2(\mathbb{R}^n)$ is dense in $[H^{a^2,q}(\mathbb{R}^n)]^*$ with respect to the norm $\|\cdot\|_{-q,a^2}$ we have $[H^{a^2,q}(\mathbb{R}^n)]^* = H^{a^2,-q}(\mathbb{R}^n)$, where the space $H^{a^2,-s}(\mathbb{R}^n)$ is defined by (2.1) taking $-q$ instead of q and u to be a Schwartz distribution.

In the case of the usual Sobolev spaces this result can be found in [34, p. 31]. The proof of our proposition follows essentially the lines of the considerations in [11, p. 201-203], and is left to the reader. We also refer to Proposition 1.2 in [21].

3. The operator $L(x,D)$.

As pointed out in the introduction we want to construct a Feller semigroup and therefore a Dirichlet form by starting with a pseudodifferential operator $L(x,D)$. This operator will be introduced now. For $1 \leq j \leq n$ let $a_j^2 : \mathbb{R} \to \mathbb{R}$ be a continuous negative definite functions. Further, for $1 \leq j \leq n$ we assume that functions $b_j : \mathbb{R}^n \to \mathbb{R}$ are given. The operator $L(x,D)$ is defined by

$$L(x,D) = \sum_{j=1}^{n} b_j(x)\, a_j^2(D_j) \ , \tag{3.1}$$

where $a_j^2(D_j)$, $1 \leq j \leq n$, is the operator

$$a_j^2(D_j)\, u(x) = (2\pi)^{-n/2} \int_{\mathbb{R}^n} e^{ix\xi} a_j^2(\xi_j)\, \hat{u}(\xi)\, d\xi \ . \tag{3.2}$$

Clearly by Lemma 1.1 this operator is well defined on $C_0^\infty(\mathbb{R}^n)$. Later we will introduce a larger domain for $L(x,D)$. Let $a^2 : \mathbb{R}^n \to \mathbb{R}$ be the function

$$a^2(\xi) = \sum_{j=1}^{n} a_j^2(\xi_j) \ . \tag{3.3}$$

Since each of the functions a_j^2 is a continuous negative definite function it is clear that a^2 and $1 + a^2$ are continuous negative definite functions (see Section 10). Obviously we have with suitable constants for any $s \geq 0$

$$c_{n,s} \leq \frac{1 + \sum_{j=1}^{n} a_j^{2s}(\xi_j)}{(1 + a^2(\xi))^s} \leq \tilde{c}_{n,s} \ . \tag{3.4}$$

Now take $r, t \in (0, 2]$, $r \geq t > 0$ and let m_0 be the smallest even integer such that

$$t(m_0 + 1) \geq 3 + \left[\frac{n}{2}\right] \tag{3.5}$$

holds. Further suppose

$$0 < 1 - \frac{r}{2t} - \frac{r(r-t)m_0}{t^2} .$$

Let $\delta \in (0, 1 - r/(2t) - r(r-t)m_0/t^2)$ be fixed and define s by

$$s = (1 - \delta)\frac{t}{r} - \frac{1}{2} . \tag{3.6}$$

It follows that

$$s - \frac{(r-t)m_0}{t} > 0 . \tag{3.7}$$

Taking t and r as above we assume

$$c_t \left(1 + |\xi|^2\right)^{t/2} \leq 1 + a^2(\xi) \leq c_r \left(1 + |\xi|^2\right)^{r/2} . \tag{3.8}$$

Now we can state our assumptions on the coefficients b_j.

B.1. It is assumed that b_j is bounded and continuous.

B.2. We suppose that $b_j = d_j + c_j$, where c_j is a real number and $d_j : \mathbb{R}^n \to \mathbb{R}$ is a function satisfying

$$|\hat{d}_j(\xi)| \leq c_q \left(1 + |\xi|^2\right)^{-q} . \tag{3.9}$$

where $q = n + r(s + 1/2) + tm_0 + 1$.

B.3. For all $x \in \mathbb{R}^n$ we require

$$b_j(x) \geq \delta_1 > 0 \tag{3.10}$$

to hold.

B.4. For some $x_0 \in \mathbb{R}^n$ we assume

$$\max_{1 \le j \le n} \sup_{x \in \mathbb{R}^n} |b_j(x) - b_j(x_0)| \le \frac{\delta_1}{2n} . \tag{3.11}$$

In order to prove commutator estimates for the operators $a_j(D_j)$ we have in addition to assume that

$$c_j \left(1 + |\xi_j|^2\right)^{t_j/2} \le 1 + a_j^2(\xi_j) \le \tilde{c}_j \left(1 + |\xi_j|^2\right)^{r_j/2} , \tag{3.12}$$

where $0 < t \le t_j \le r_j \le r \le 2$.

In Section 10 we will give examples of operators satisfying all the conditions stated above. Note that our assumptions are far from being the most general or the sharpest. In particular when considering special operators as it was done in [24] or [16] much of our assumptions could be relaxed. Beside the operator $L(x, D)$ we also will often consider the operator $L^\lambda(x, D) = L(x, D) + \lambda$, $\lambda \in \mathbb{R}$.

The next lemma will be used frequently

Lemma 3.1. *For all $u \in C_0^\infty(\mathbb{R}^n)$ we have*

$$\|a_j(D_j)u\|_0^2 \le \sum_{j=1}^n \|a_j(D_j)u\|_0^2 \le \|u\|_{1/2,a^2}^2 . \tag{3.13}$$

PROOF. For $u \in C_0^\infty(\mathbb{R}^n)$ we have

$$\begin{aligned} \|a_j(D_j)u\|_0^2 &= \int_{\mathbb{R}^n} a_j^2(\xi_j) |\hat{u}(\xi)|^2 d\xi \le \int_{\mathbb{R}^n} \sum_{j=1}^n a_j^2(\xi_j) |\hat{u}(\xi)|^2 d\xi \\ &\le \int_{\mathbb{R}^n} (1 + \sum_{j=1}^n a_j^2(\xi_j)) |\hat{u}(\xi)|^2 d\xi = \|u\|_{1/2,a^2}^2 . \end{aligned}$$

4. On the bilinear form associated with $L(x, D)$.

The operator $L(x, D)$ can be regarded as a pseudo differential operator with symbol

$$L(x, \xi) = \sum_{j=1}^n b_j(x) a_j^2(\xi_j) .$$

This operator is clearly defined on $C_0^\infty(\mathbb{R}^n)$ by

$$\begin{aligned} L(x,D)u(x) &= (2\pi)^{-n/2}\int_{\mathbb{R}^n} e^{ix\xi}L(x,\xi)\hat{u}(\xi)\,d\xi \\ &= (2\pi)^{-n/2}\sum_{j=1}^{n} b_j(x)\int_{\mathbb{R}^n} e^{ix\xi}\,a_j^2(\xi_j)\hat{u}(\xi)\,d\xi\,. \end{aligned} \tag{4.1}$$

Since $a_j^2(D_j)u \in C_\infty(\mathbb{R}^n)$ for any $u \in C_0^\infty(\mathbb{R}^n)$, see [21, Proposition 1.4] and $b_j \in L^\infty(\mathbb{R}^n)$, it follows that we can define on $C_0^\infty(\mathbb{R}^n)$ the bilinear form

$$B(u,v) = (L(x,D)u,v)_0 = \sum_{j=1}^{n}\left(b_j(\cdot)\,a_j^2(D_j)u,v\right)_0\,. \tag{4.2}$$

The bilinear form associated with $L^\lambda(x,D)$ is denoted by B_λ. Our next aim is to prove that B has a continuous extension onto $H^{a^2,1/2}(\mathbb{R}^n)$. For this we need

Proposition 4.1. *Suppose that* (3.12) *holds. Then for any* $\eta > 0$ *there exists a constant* $c(\eta) \geq 0$ *such that*

$$\|[a_j(D_j),b_j(\cdot)]u\|_{1/2,a^2} \leq \eta\,\|u\|_{1/2,a^2} + c(\eta)\,\|u\|_0 \tag{4.3}$$

for all $u \in H^{a^2,1/2}(\mathbb{R}^n)$.

As usual we denote by $[a_j(D_j),b_j(\cdot)]$ the commutator of $a_j(D_j)$ and $b_j(\cdot)$, *i.e.* the operator

$$u \mapsto a_j(D_j)(b_j u)(\cdot) - b_j(\cdot)\,a_j(D_j)u(\cdot)\,.$$

The proof of Proposition 4.1 is analogous to that of Corollary 3.2 in [22]. It will be given in Section 6 where more general commutator estimates are discussed. Using the commutator $[a_j(D_j),b_j]$ we can write B in a more appropiate way, namely

$$\begin{aligned} B(u,v) &= \sum_{j=1}^{n}(b_j(\cdot)a_j(D_j)u, a_j(D_j)v)_0 \\ &\quad + \sum_{j=1}^{n}(a_j(D_j)u,[a_j(D_j),b_j(\cdot)]v)_0\,. \end{aligned} \tag{4.4}$$

Now we claim

Theorem 4.1. *For all $u, v \in C_0^\infty(\mathbb{R}^n)$ we have*

$$|B(u,v)| \leq c \|u\|_{1/2,a^2} \|v\|_{1/2,a^2} . \tag{4.5}$$

PROOF. Let $u, v \in C_0^\infty(\mathbb{R}^n)$. Then it follows that

$$\begin{aligned}
|B(u,v)| &\leq \sum_{j=1}^n |(b_j(\cdot)a_j(D_j)u, a_j(D_j)v)_0| \\
&\quad + \sum_{j=1}^n |(a_j(D_j)u, [a_j(D_j), b_j(\cdot)]v)_0| \\
&\leq c \sum_{j=1}^n \|a_j(D_j)u\|_0 \, \|a_j(D_j)v\|_0 \\
&\quad + \sum_{j=1}^n \|a_j(D_j)u\|_0 \, \|[a_j(D_j), b_j(\cdot)]v\|_0 \\
&\leq c \|u\|_{1/2,a^2} \, \|v\|_{1/2,a^2} ,
\end{aligned}$$

where we used Lemma 3.1 and Proposition 4.1 for the last step.

Obviously (4.5) holds also for B_λ, $\lambda \in \mathbb{R}$. Thus B_λ has a continuous extension onto $H^{a^2,1/2}(\mathbb{R}^n)$ which is again denoted by B_λ. Furthermore we have

Theorem 4.2. *For all $u \in H^{a^2,1/2}(\mathbb{R}^n)$ we have with a suitable constant d_0*

$$B(u,u) \geq \frac{\delta_1}{2} \|u\|^2_{1/2,a^2} - d_0 \|u\|_0^2 . \tag{4.6}$$

PROOF. It is sufficient to prove (4.6) for all $u \in C_0^\infty(\mathbb{R}^n)$. For these u we find

$$\begin{aligned}
B(u,u) &= \sum_{j=1}^n (b_j(\cdot)a_j^2(D_j)u, u)_0 \\
&\geq \sum_{j=1}^n (b_j(\cdot)a_j(D_j)u, a_j(D_j)u)_0
\end{aligned}$$

$$
\begin{aligned}
&- |\sum_{j=1}^{n}(a_j(D_j)u, [a_j(D_j), b_j(\cdot)]u)_0| \\
&= B_1 - |B_2| \, .
\end{aligned}
$$

Now we get using B.3

$$
(4.7) \qquad B_1 = \int_{\mathbb{R}^n} \sum_{j=1}^{n} b_j(x)|a_j^2(D_j)u(x)|^2 dx \geq \delta_1 \|u\|_{1/2,a^2} - \delta_1 \|u\|_0^2 \, .
$$

Using Proposition 4.1 we can estimate B_2 as follows

$$
\begin{aligned}
|B_2| &= |\sum_{j=1}^{n}(a_j(D_j)u, [a_j(D_j), b_j(\cdot)]u)_0| \\
&\leq \sum_{j=1}^{n} \|u\|_{1/2,a^2} \left(\eta \|u\|_{1/2,a^2} + c(\eta) \|u\|_0\right) \\
&\leq \varepsilon \|u\|_{1/2,a^2}^2 + c(\varepsilon) \|u\|_0^2 \, , \qquad (4.8)
\end{aligned}
$$

where $\eta > 0$ and therefore $\varepsilon > 0$ are sufficiently small constants. Thus by (4.7) and (4.8) we have

$$
B(u,u) \geq (\delta_1 - \varepsilon) \|u\|_{1/2,a^2}^2 - (c(\varepsilon) + \delta_1) \|u\|_0^2 \, ,
$$

which implies (4.6).

It follows that $L^\lambda(x,D)$, $\lambda \in \mathbb{R}$, has a closed extension L^λ, called the Friedrichs extension, with domain $D(L^\lambda)$ defined as the set of the functions $u \in H^{a^2,1/2}(\mathbb{R}^n)$ such that

(4.9) there exists $f \in L^2(\mathbb{R}^n)$ such that for all $v \in H^{a^2,1/2}(\mathbb{R}^n)$:

$$
B_\lambda(u,v) = (f,v)_0 \, .
$$

Note that L^λ is the only closed extension of $L^\lambda(x,D)$ with the property that $D(L^\lambda) \subset H^{a^2,1/2}(\mathbb{R}^n)$, see [26] or [41]. Moreover $-L^\lambda$ is the generator of an analytic semigroup of contractions provided λ is sufficiently large. Our next goal is to characterize the domain $D(L^\lambda)$.

5. A characterization of $D(L^\lambda)$.

First of all let us prove

Proposition 5.1. *The operator $L^\lambda(x,D)$, $\lambda \in \mathbb{R}$, has a continuous extension onto $H^{a^2,1}(\mathbb{R}^n)$, i.e. $L^\lambda(x,D) : H^{a^2,1}(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$ is a continu- ous operator.*

PROOF. For $u \in C_0^\infty(\mathbb{R}^n)$ we find using B.1

$$\|L^\lambda(x,D)u\|_0 \leq \|\sum_{j=1}^n b_j(\cdot)a_j^2(D_j)u\|_0 + |\lambda|\,\|u\|_0 \leq c'\,\|u\|_{1,a^2}\ .$$

The next estimate will give an important regularity result for solutions of the equation $L^\lambda(x,D)u = f$.

Theorem 5.1. *Under the assumptions* B.1-B.4 *and* (3.12) *there exists a constant c_λ such that*

$$\|u\|_{1,a^2} \leq c_\lambda\left(\|L^\lambda(x,D)u\|_0 + \|u\|_0\right) \tag{5.1}$$

holds for all $u \in L^2(\mathbb{R}^n)$ with $L^\lambda(x,D)u \in L^2(\mathbb{R}^n)$.

PROOF. Using Proposition 2.1 we find

$$\begin{aligned}
&\|u\|_{1,a^2}\,\|\sum_{j=1}^n b_j(\cdot)a_j^2(D_j)u + \lambda u\|_0 \\
&\quad = \|u\|_{1,a^2}\,\|(1+\sum_{l=1}^n a_l^2(D_l))\,(\sum_{j=1}^n b_j(\cdot)a_j^2(D_j)u + \lambda u)\|_{-1,a^2} \\
&\quad \geq (u,(1+\sum_{l=1}^n a_l^2(D_l))(\sum_{j=1}^n b_j(\cdot)a_j^2(D_j)u + \lambda u))_0 \\
&\quad = ((1+\sum_{l=1}^n a_l^2(D_l))u, \sum_{j=1}^n b_j(\cdot)a_j^2(D_j)u + \lambda u)_0 \\
&\quad = \sum_{j,l=1}^n (a_l^2(D_l)u, b_j(\cdot)a_j^2(D_j)u)_0 + \lambda\,\|u\|_0^2
\end{aligned}$$

$$+ \lambda \Big(\sum_{l=1}^{n} a_l^2(D_l)u, u\Big)_0 + \sum_{j=1}^{n} (u, b_j(\cdot)a_j^2(D_j)u)_0 \,.$$

For $\lambda \geq 0$ we find

$$\lambda \Big(\sum_{l=1}^{n} a_l^2(D_l)u, u\Big)_0 = \lambda \sum_{l=1}^{n} \|a_l(D_l)u\|_0^2 \geq 0 \,. \tag{5.2}$$

Further we have with x_0 as in B.4

$$\begin{aligned}\sum_{j,l=1}^{n} (a_l^2(D_l)u, b_j(\cdot)a_j^2(D_j)u)_0 &= \sum_{j,l=1}^{n} (a_l^2(D_l)u, b_j(x_0)a_j^2(D_j)u)_0 \\ &\quad + \sum_{j,l=1}^{n} ((b_j(\cdot) - b_j(x_0))a_j^2(D_j)u, a_l^2(D_l)u)_0 \\ &= A_1 + A_2 \,.\end{aligned}$$

Now, by B.3 and Lemma 2.1 we get for $\eta_1 > 0$

$$\begin{aligned} A_1 &= \sum_{j,l=1}^{n} (b_j(x_0)a_j^2(D_j)u, a_l^2(D_l)u)_0 \\ &\geq \delta_1 \sum_{j,l=1}^{n} \int_{\mathbb{R}^n} a_j^2(\xi_j)a_l^2(\xi_l)\,|\hat{u}(\xi)|^2 d\xi \\ &= \delta_1 \int_{\mathbb{R}^n} \Big(1 + \sum_{j=1}^{n} a_j^2(\xi_j)\Big)^2 |\hat{u}(\xi)|^2 d\xi - \delta_1 \|u\|_0^2 \\ &\quad - 2\,\delta_1 \int_{\mathbb{R}^n} \sum_{j=1}^{n} a_j^2(\xi_j)\,|\hat{u}(\xi)|^2 d\xi \\ &\geq (\delta_1 - \eta_1)\,\|u\|_{1,a^2}^2 - c(\delta_1, \eta_1)\,\|u\|_0^2 \,. \end{aligned} \tag{5.3}$$

Now we estimate A_2 by taking into account B.4:

$$\begin{aligned} |A_2| &\leq \sum_{j,l=1}^{n} |(b_j(\cdot) - b_j(x_0))a_j^2(D_j)u, a_l^2(D_l)u)_0| \\ &\leq \max_{1\leq j\leq n} \sup_{x\in\mathbb{R}^n} |b_j(x) - b_j(x_0)| \sum_{j,l=1}^{n} \|a_j^2(D_j)u\|_0\,\|a_l^2(D_l)u\|_0 \end{aligned}$$

$$\leq \max_{1\leq j\leq n} \sup_{x\in\mathbb{R}^n} |b_j(x) - b_j(x_0)| \Big(\sum_{j=1}^n \|a_j^2(D_j)u\|_0 \Big) \Big(\sum_{l=1}^n \|a_l^2(D_l)u\|_0 \Big)$$
$$\leq n \max_{1\leq j\leq n} \sup_{x\in\mathbb{R}^n} |b_j(x) - b_j(x_0)| \, \|u\|_{1,a^2}^2$$
$$\leq \frac{\delta_1}{2} \|u\|_{1,a^2}^2 .$$

Thus

$$|A_2| \leq \frac{\delta_1}{2} \|u\|_{1,a^2}^2 . \tag{5.4}$$

Moreover using Proposition 4.1 and Lemma 2.1 we find for any $\eta_2 > 0$

$$|\sum_{j=1}^n (u, b_j(\cdot)a_j^2(D_j)u)_0|$$
$$\leq c \sum_{j=1}^n |(a_j(D_j)u, a_j(D_j)u)_0| + \sum_{j=1}^n |([b_j(\cdot), a_j(D_j)]u, a_j(D_j)u)_0|$$
$$\leq c \|u\|_{1/2,a^2}^2 \leq \eta_2 \|u\|_{1,a^2}^2 + c(\eta_2) \|u\|_0^2 .$$

Thus we find for $\lambda \geq 0$

$$\|u\|_{1,a^2} \| \sum_{j=1}^n b_j(\cdot)a_j^2(D_j)u + \lambda u\|_0$$
$$\geq (\delta_1 - \delta_1/2 - \eta_1 - \eta_2)\|u\|_{1,a^2}^2$$
$$+ \lambda \sum_{l=1}^n \|a_l(D_l)u\|_0^2 + (\lambda - c(\delta_1, \eta_1, \eta_2)) \|u\|_0^2 .$$

For $\eta_1 = \eta_2 = \delta_1/8$ and $\lambda \geq \lambda_0 = c(\delta_1)$ we find using (3.12)

$$\|L^\lambda(x, D)u\|_0 \geq \frac{\delta_1}{4} \|u\|_{1,a^2} . \tag{5.5}$$

Now let $\lambda \in \mathbb{R}$ be arbitrary. Then it follows that

$$\|u\|_{1,a^2} \leq c \|L^{\lambda_0}(x, D)u\|_0$$
$$= c \, \|L^\lambda(x, D)u + (\lambda_0 - \lambda)u\|_0$$
$$\leq c \, (\|L^\lambda(x, D)u\|_0 + |\lambda_0 - \lambda| \, \|u\|_0)$$

$$\leq c' \left(\|L^\lambda(x,D)u\|_0 + \|u\|_0 \right),$$

which proves the theorem.

Note that Theorem 5.1 also follows from Theorem 7.1 below, even the structure of the proof is the same. However our proof of Theorem 5.1 does not use Theorem 6.1, that is why we have given it separately.

Corollary 5.1. *Let us consider $L^\lambda(x,D)$ as an operator on $L^2(\mathbb{R}^n)$ with domain $H^{a^2,1}(\mathbb{R}^n) \subset L^2(\mathbb{R}^n)$. Then $L^\lambda(x,D)$ is a closed operator.*

PROOF. We prove that $H^{a^2,1}(\mathbb{R}^n)$ equipped with the graph norm $\|u\|_0 + \|L^\lambda(x,D)u\|_0$ is a Hilbert space. By Proposition 5.1 we have

$$\|u\|_0 + \|L^\lambda(x,D)u\|_0 \leq c \|u\|_{1,a^2}$$

for all $u \in H^{a^2,1}(\mathbb{R}^n)$. Conversely, by Theorem 5.1 we find

$$c' \|u\|_{1,a^2} \leq \|u\|_0 + \|L^\lambda(x,D)u\|_0$$

for all $u \in H^{a^2,1}(\mathbb{R}^n)$, since for these u we have by Proposition 5.1 that $L^\lambda(x,D)u \in L^2(\mathbb{R}^n)$. Thus the graph norm is equivalent to the norm $\|\cdot\|_{1,a^2}$ which implies that $H^{a^2,1}(\mathbb{R}^n)$ is a Hilbert space with respect to the graph norm.

Following [26, p. 325-326], we get

Theorem 5.2. *Let $L^\lambda(x,D)$ be as above and let L^λ be its Friedrichs extension. Then we have $D\left(L^\lambda\right) = H^{a^2,1}(\mathbb{R}^n)$ and $L^\lambda = L^\lambda(x,D)$ as operators defined on $H^{a^2,1}(\mathbb{R}^n)$.*

Note that Theorem 5.2 is a regularity result for solutions of the representation problem

$$B_\lambda(u,\varphi) = (f,\varphi)_0 \qquad \text{for all } \varphi \in C_0^\infty(\mathbb{R}^n), \tag{5.6}$$

where $f \in L^2(\mathbb{R}^n)$ is a given function. For $\lambda \geq d_0$ the non-symmetric version of the Lax-Milgram theorem implies that (5.6) always has a solution in $H^{a^2,1/2}(\mathbb{R}^n)$, while the definition of $D\left(L^\lambda\right)$ together with Theorem 5.2 gives that this solution belongs already to $H^{a^2,1}(\mathbb{R}^n)$.

6. Some commutator estimates.

In order to get further regularity results for solutions of the equation $L^\lambda(x,D)u = f$ we have to prove some commutator estimates. First we note the trivial identity

$$x^m - y^m = (x-y)\sum_{l=0}^{m-1} x^l y^{m-1-l} \tag{6.1}$$

which holds for all $x, y \in \mathbb{R}$ and $m \in \mathbb{N}$.

Lemma 6.1. *Let $a^2 : \mathbb{R}^n \to \mathbb{R}$ be a continuous negative definite function and $m \in \mathbb{N}$. Then we have for all $\xi, \eta \in \mathbb{R}^n$*

$$\begin{aligned} &|(1+a^2(\xi))^m - (1+a^2(\eta))^m| \\ &\leq 4\,(a(\xi-\eta)+a^2(\xi-\eta))\sum_{l=0}^{m-1}(1+a^2(\xi))^{l+1/2}(1+a^2(\eta))^{m-1-l}\,. \end{aligned} \tag{6.2}$$

PROOF. By (6.1) we have

$$\begin{aligned} &|(1+a^2(\xi))^m - (1+a^2(\eta))^m| \\ &= |(1+a^2(\xi)) - (1+a^2(\eta))|\sum_{l=0}^{m-1}(1+a^2(\xi))^l(1+a^2(\eta))^{m-1-l} \end{aligned}$$

Since a^2 is a continuous negative definite function we find using Lemma 1.2

$$\begin{aligned} &|(1+a^2(\xi))^m - (1+a^2(\eta))^m| \\ &\leq (4\,a(\xi)a(\xi-\eta)+a^2(\xi-\eta))\sum_{l=0}^{m-1}(1+a^2(\xi))^l(1+a^2(\eta))^{m-1-l} \\ &\leq (4\,a(\xi-\eta)+a^2(\xi-\eta))\sum_{l=0}^{m-1}(1+a^2(\xi))^{l+1/2}(1+a^2(\eta))^{m-1-l}\,, \end{aligned}$$

which proves the lemma.

The proof of Theorem 6.1 requires the following two lemmas

Lemma 6.2. ([35], Lemma 2.2.1) *For any $q \in \mathbb{R}$ and all $\xi, \eta \in \mathbb{R}^n$ the inequality*

$$(1 + |\xi|^2)^q (1 + |\eta|^2)^{-q} \leq 2^{|q|} (1 + |\xi - \eta|^2)^{|q|} \tag{6.3}$$

holds.

Lemma 6.3. ([35], Lemma 2.2.4) *Let $k \in L^1(\mathbb{R}^n)$. Then we have for all $u, v \in L^2(\mathbb{R}^n)$*

$$\left| \int_{\mathbb{R}^n} \Big(\int_{\mathbb{R}^n} k(\xi - \eta) u(\eta) v(\xi) d\eta \Big) d\xi \right| \leq \|k\|_{L^1} \|u\|_0 \|v\|_0 . \tag{6.4}$$

Now we prove

Theorem 6.1. *Let b_j be as in* B.2, *in particular suppose* (3.9) *with $q = n + r(s + 1/2) + t m_0 + 1$. Then for all $u \in H^{a^2, m_0}(\mathbb{R}^n)$ we have*

$$\|[(1 + a^2(D))^{m_0}, b_j(\cdot)]u\|_{s_0, a^2} \leq c \|u\|_{m_0 - \delta, a^2} , \tag{6.5}$$

where $s_0 = s - \dfrac{r - t}{t} m_0$ with m_0, s and δ as in Section 3.

PROOF. First we note that for $u \in C_0^\infty(\mathbb{R}^n)$ we have

$$\begin{aligned} [(1 + a^2(D))^{m_0}, b_j(\cdot)]u &= [(1 + a^2(D))^{m_0}, d_j(\cdot) + c_j]u \\ &= [(1 + a^2(D))^{m_0}, d_j(\cdot)]u , \end{aligned}$$

thus we only have to prove (6.3) with d_j instead of b_j. By a straightforward calculation we find for $u \in C_0^\infty(\mathbb{R}^n)$ that

$$\begin{aligned} &([(1 + a^2(D))^{m_0}, d_j(\cdot)]u)\hat{}\,(\xi) \\ &\qquad = \int_{\mathbb{R}^n} \hat{d}_j(\xi - \eta)((1 + a^2(\xi))^{m_0} - (1 + a^2(\eta))^{m_0}) \hat{u}(\eta) d\eta . \end{aligned}$$

Furthermore, for $v \in L^2(\mathbb{R}^n)$ we have

$$\begin{aligned} &|([(1 + a^2(D))^{m_0}, d_j(\cdot)]u, v)_0| \\ &= \left| \int_{\mathbb{R}^n} \Big(\int_{\mathbb{R}^n} \overline{\hat{d}_j(\xi - \eta)}((1 + a^2(\xi))^{m_0} - (1 + a^2(\eta))^{m_0}) \overline{\hat{u}(\eta)} \hat{v}(\xi) d\eta \Big) d\xi \right| \end{aligned}$$

$$\underset{(6.2)}{\leq} c \sum_{l=0}^{m_0-1} \int_{\mathbb{R}^n} \Big(\int_{\mathbb{R}^n} |\hat{d}_j(\xi-\eta)|\,(a(\xi-\eta)+a^2(\xi-\eta)) \cdot (1+a^2(\xi))^{l+1/2}(1+a^2(\eta))^{m_0-1-l}|\hat{u}(\eta)|\,|\hat{v}(\xi)|d\eta\Big)\,d\xi$$

$$\leq c \sum_{l=0}^{m_0-1} \int_{\mathbb{R}^n} \Big(\int_{\mathbb{R}^n} |\hat{d}_j(\xi-\eta)|(1+|\xi-\eta|^2) \cdot \frac{(1+a^2(\xi))^l}{(1+a^2(\eta))^l} \frac{(1+a^2(\xi)^{1/2}}{(1+a^2(\eta))^{-m_0+1}} \cdot (1+a^2(\xi))^s(1+a^2(\eta))^{-m_0+\delta}(1+a^2(\xi))^{-s}|\hat{v}(\xi)| \cdot (1+a^2(\eta))^{m_0-\delta}|\hat{u}(\eta)|d\eta\Big)\,d\xi$$

$$\underset{(3.8)}{\leq} c \sum_{l=0}^{m_0-1} c_l \int_{\mathbb{R}^n} \Big(\int_{\mathbb{R}^n} |\hat{d}_j(\xi-\eta)|\,(1+|\xi-\eta|^2) \cdot \frac{(1+|\xi|^2)^{rl/2}\,(1+|\xi|^2)^{r(s+1/2)/2}}{(1+|\eta|^2)^{tl/2}\,(1+|\eta|^2)^{t(1-\delta)/2}} \cdot (1+a^2(\xi))^{-s}|\hat{v}(\xi)|(1+a^2(\eta))^{m_0-\delta}|\hat{u}(\eta)|d\eta\Big)\,d\xi$$

$$\underset{(3.6)}{=} c \sum_{l=0}^{m_0-1} c_l \int_{\mathbb{R}^n} \Big(\int_{\mathbb{R}^n} |\hat{d}_j(\xi-\eta)|\,(1+|\xi-\eta|^2)\,(1+|\xi|^2)^{l(r-t)/2} \cdot \left(\frac{1+|\xi|^2}{1+|\eta|^2}\right)^{(r(s+1/2)+tl)/2} (1+a^2(\xi))^{-s}|\hat{v}(\xi)| \cdot (1+a^2(\eta))^{m_0-\delta}|\hat{u}(\eta)|d\eta\Big)\,d\xi$$

$$\underset{(3.8),(6.3)}{\leq} c \sum_{l=0}^{m_0-1} \tilde{c}_l \int_{\mathbb{R}^n} \Big(\int_{\mathbb{R}^n} |\hat{d}_j(\xi-\eta)|(1+|\xi-\eta|^2) \cdot (1+|\xi-\eta|^2)^{(r(s+1/2)+tl)/2} \cdot (1+a^2(\xi))^{-s+(r-t)(l/t)} \cdot |\hat{v}(\xi)|(1+a^2(\eta))^{m_0-\delta}|\hat{u}(\eta)|d\eta\Big)\,d\xi$$

$$\underset{(3.9)}{\leq} c \int_{\mathbb{R}^n} \Big(\int_{\mathbb{R}^n} (1+|\xi-\eta|^2)^{-q}(1+|\xi-\eta|^2)^{(r(s+1/2)+tm_0+2)/2} \cdot (1+a^2(\xi))^{-s+(r-t)(m_0/t)}|\hat{v}(\xi)|$$

$$\cdot (1+a^2(\eta))^{m_0-\delta}|\hat{u}(\eta)|d\eta\Big)\,d\xi$$

$$\underset{(6.4)}{\leq} c\,\|v\|_{-s_0,a^2}\,\|u\|_{m_0-\delta,a^2}\,\cdot$$

Thus we find

$$\frac{|([(1+a^2(D))^{m_0}, d_j(\cdot)]u, v)_0|}{\|v\|_{-s_0,a^2}} \leq c\,\|u\|_{m_0-\delta,a^2}\,,$$

which implies by Proposition 2.1

$$\|[(1+a^2(D))^{m_0}, d_j(\cdot)]u\|_{s-2(r-t)(m_0/t)} \leq c\,\|u\|_{m_0-\delta,a^2}\,,$$

thus the theorem is proved.

Corollary 6.1. *Let b_j be as in Theorem* 6.1. *Then for any $\eta > 0$ there exists $c(\eta) \geq 0$ such that*

$$\|[(1+a^2(D))^{m_0}, b_j(\cdot)]u\|_0 \leq \eta\,\|u\|_{m_0,a^2} + c(\eta)\,\|u\|_0 \tag{6.6}$$

holds for all $u \in H^{a^2,m_0}(\mathbb{R}^n)$.

PROOF. For $u \in H^{a^2,m_0}(\mathbb{R}^n)$ we have using (6.5) and Lemma 2.1

$$\begin{aligned}\|[(1+a^2(D))^{m_0}, b_j(\cdot)]u\|_0 &\leq \|[(1+a^2(D))^{m_0}, b_j(\cdot)]u\|_{s-(r-t)(m_0/t),a^2}\\ &\leq c\,\|u\|_{m_0-\delta,a^2}\\ &\leq \eta\,\|u\|_{m_0,a^2} + c(\eta)\,\|u\|_0\,.\end{aligned}$$

Finally we have to give the

PROOF OF PROPOSITION 4.1. Let $\alpha_l \in (0, 1 - r_l/2t_l)$ and set $s_l = (1-\delta)t_l/r_l - 1/2$. Then we may proceed as in the proof of Theorem 6.1 (or as in the proof of Theorem 3.1 in [22]) to get for $u \in C_0^\infty(\mathbb{R}^n)$ and $v \in L^2(\mathbb{R}^n)$

$$\begin{aligned}&|([a_l(D_l), b_j(\cdot)]u, v)_0| = |([a_l(D_l), d_j(\cdot)]u, v)_0|\\ &\leq c\int_{\mathbb{R}^n}\Big(\int_{\mathbb{R}^n}|\hat{d}_j(\xi-\eta)|(1+|\xi-\eta|^2)\frac{(1+a_l^2(\xi_l))^{s_l+1/2}}{(1+a_l^2(\eta_l))^{1-\alpha_l}}\\ &\qquad\cdot(1+a_l^2(\xi_l))^{-s_l}|\hat{v}(\xi)|(1+a_l^2(\eta_l))^{1-\alpha_l}|\hat{u}(\eta)|d\eta\Big)\,d\xi\\ &\underset{(3.12)}{\leq} c\int_{\mathbb{R}^n}\Big(\int_{\mathbb{R}^n}|\hat{d}_j(\xi-\eta)|(1+|\xi-\eta|^2)\left(\frac{1+|\xi|^2}{1+|\eta|^2}\right)^{r_l(s_l+1/2)/2}\\ &\qquad\cdot(1+a_l^2(\xi_l))^{-s_l}|\hat{v}(\xi)|(1+a_l^2(\eta_l))^{1-\alpha_l}|\hat{u}(\eta)|d\eta\Big)\,d\xi\,,\end{aligned}$$

which implies

$$\|[a_l(D_l), b_j(\cdot)]u\|_{s_l,a^2} \leq c\,\|u\|_{1-\alpha_l,a^2}\ ,$$

from which the proposition follows as Corollary 6.1 follows from Theorem 6.1.

7. A regularity result.

We will need a stronger regularity result for solutions of the equation

$$L^\lambda(x,D)u = f \tag{7.1}$$

For this we give

Theorem 7.1. *Let $L^\lambda(x,D)$ be as before, i.e. assume* B.1-B.4. *Suppose further that $L^\lambda(x,D)u \in H^{a^2,m_0}(\mathbb{R}^n)$ for some $u \in L^2(\mathbb{R}^n)$. Then $u \in H^{a^2,m_0+1}(\mathbb{R}^n)$ and*

$$\|u\|_{m_0+1,a^2} \leq c(\|L^\lambda(x,D)u\|_{m_0,a^2} + \|u\|_0) \tag{7.2}$$

holds.

PROOF. Let u be as stated in the theorem. Then we have using Proposition 2.1

$$\begin{aligned}
&\|u\|_{m_0+1,a^2}\ \|L^\lambda(x,D)u\|_{m_0,a^2}\\
&= \|u\|_{m_0+1,a^2}\ \|(1+a^2(D))^{1+2m_0}L^\lambda(x,D)u\|_{-m_0-1,a^2}\\
&\geq (u,(1+a^2(D))^{1+2m_0}L^\lambda(x,D)u)_0\\
&= ((1+a^2(D))^{1+m_0}u,(1+a^2(D))^{m_0}L^\lambda(x,D)u)_0\\
&= ((1+a^2(D))^{1+m_0}u,(1+a^2(D))^{m_0}(\sum_{j=1}^n b_j(\cdot)a_j^2(D_j)u+\lambda u))_0\\
&= \lambda\,((1+a^2(D))^{1+m_0}u,(1+a^2(D))^{m_0}u)_0\\
&\quad + ((1+a^2(D))^{1+m_0}u,(1+a^2(D))^{m_0}\sum_{j=1}^n b_j(\cdot)a_j^2(D_j)u)_0\\
&= \lambda\,\|u\|^2_{m_0+1/2,a^2} + D_1 + D_2\ ,
\end{aligned}$$

where

$$D_1 = ((1 + a^2(D))^{1+m_0}u, (1 + a^2(D))^{m_0} \sum_{j=1}^{n} b_j(x_0)a_j^2(D_j)u)_0$$

and

$$D_2 = ((1 + a^2(D))^{1+m_0}u, (1 + a^2(D))^{m_0} \sum_{j=1}^{n} (b_j(\cdot) - b_j(x_0))a_j^2(D_j)u)_0 \ .$$

Here x_0 is again the fixed point given by (3.11). First we estimate D_1 using (3.10):

$$\begin{aligned} D_1 &= \int_{\mathbb{R}^n} (1 + a^2(\xi))^{1+m_0}(1 + a^2(\xi))^{m_0} \sum_{j=1}^{n} b_j(x_0)a_j^2(\xi_j)|\hat{u}(\xi)|^2 d\xi \\ &\geq \delta_1 \int_{\mathbb{R}^n} (1 + a^2(\xi))^{1+2m_0} \sum_{j=1}^{n} a_j^2(\xi_j)|\hat{u}(\xi)|^2 d\xi \\ &\geq \delta_1 \|u\|^2_{m_0+1,a^2} - \delta_1 \|u\|^2_{m_0+1/2,a^2} \ . \end{aligned} \tag{7.3}$$

Now let us turn to D_2:

$$\begin{aligned} D_2 &= ((1 + a^2(D))^{1+m_0}u, (1 + a^2(D))^{m_0} \sum_{j=1}^{n} (b_j(\cdot) - b_j(x_0))a_j^2(D_j)u)_0 \\ &= \sum_{j=1}^{n} ((1 + a^2(D))^{1+m_0}u, (b_j(\cdot) - b_j(x_0))(1 + a^2(D))^{m_0}a_j^2(D_j)u)_0 \\ &\quad + \sum_{j=1}^{n} ((1 + a^2(D))^{1+m_0}u, [(1 + a^2(D))^{m_0}, b_j(\cdot)]a_j^2(D_j)u)_0 \\ &= D_{12} + D_{22} \ . \end{aligned}$$

By (3.11) we get

$$\begin{aligned} |D_{12}| &\leq \max_{1\leq j\leq n} \sup_{x\in\mathbb{R}^n} |b_j(x) - b_j(x_0)| \\ &\qquad \cdot \sum_{j=1}^{n} \|(1 + a^2(D))^{1+m_0}u\|_0 \ \|(1 + a^2(D))^{m_0}a_j^2(D_j)u\|_0 \\ &\leq \max_{1\leq j\leq n} \sup_{x\in\mathbb{R}^n} |b_j(x) - b_j(x_0)| \end{aligned}$$

$$\cdot \|(1+a^2(D))^{1+m_0}u\|_0 \sum_{j=1}^{n} \|(1+a^2(D))^{m_0}a_j^2(D_j)u\|_0$$

$$\leq n \max_{1\leq j\leq n} \sup_{x\in\mathbb{R}^n} |b_j(x)-b_j(x_0)|\, \|u\|^2_{1+m_0,a^2}$$

(7.4)

$$\leq \frac{\delta_1}{2}\, \|u\|^2_{1+m_0,a^2}\ .$$

Furthermore we find

$$|D_{22}| = |\sum_{j=1}^{n}((1+a^2(D))^{1+m_0}u, [(1+a^2(D))^{m_0}, b_j(\cdot)]a_j^2(D_j)u)_0|$$

$$\leq \sum_{j=1}^{n} \|u\|_{1+m_0,a^2}\, \|[(1+a^2(D))^{m_0}, b_j(\cdot)]a_j^2(D_j)u\|_0$$

(7.5)

$$\leq \varepsilon\, \|u\|^2_{1+m_0,a^2} + c(\varepsilon)\, \|u\|^2_0\ ,$$

where $\varepsilon > 0$ is an arbitrary number. Combining (7.3)-(7.5) we find

$$\|u\|_{1+m_0,a^2}\|L^\lambda(x,D)u\|_{m_0,a^2} \geq (\delta_1 - \frac{\delta_1}{2} - \varepsilon)\, \|u\|^2_{1+m_0,a^2}$$
$$+ (\lambda-\delta_1)\, \|u\|^2_{m_0+1/2,a^2} - c(\varepsilon)\, \|u\|^2_0\ .$$

Since $\|u\|_0 \leq \|u\|_{m_0+1/2,a^2}$ we get for $\lambda \geq \delta_1$

$$\|u\|_{1+m_0,a^2}\|L^\lambda(x,D)u\|_{m_0,a^2}$$
$$\geq (\frac{\delta_1}{2} - \varepsilon)\, \|u\|^2_{1+m_0,a^2} + (\lambda - \delta_1 - c(\varepsilon))\, \|u\|^2_0\ .$$

Thus taking $\varepsilon = \delta_1/4$ and $\lambda \geq \delta_1 + c\,(\delta_1/4)$ we find

$$\|L^\lambda(x,D)u\|_{m_0,a^2} \geq \frac{\delta_1}{4}\, \|u\|_{m_0+1,a^2}\ .$$

Now let $\lambda \in \mathbb{R}$ be arbitrary and set $\lambda_0 = \delta_1 + c\,(\delta_1/4)$. Then it follows using Lemma 2.1 for any $\eta > 0$ that

$$\|u\|_{m_0+1,a^2} \leq \frac{4}{\delta_1}\, \|L^{\lambda_0}(x,D)u\|_{m_0,a^2}$$

$$\leq \frac{4}{\delta_1}\left(\|L^\lambda(x,D)u\|_{m_0,a^2} + |\lambda-\lambda_0|\|u\|_{m_0,a^2}\right)$$
$$\leq \frac{4}{\delta_1}\left(\|L^\lambda(x,D)u\|_{m_0,a^2} + |\lambda-\lambda_0|\,\eta\,\|u\|_{m_0+1,a^2} + c(\eta)\,|\lambda-\lambda_0|\,\|u\|_0\right).$$

For $\eta = \delta_1\,|\lambda-\lambda_0|/8$ we finally get

$$\|u\|_{m_0+1,a^2} \leq \frac{8}{\delta_1}\|L^\lambda(x,D)u\|_{m_0,a^2} + \tilde{c}\,\|u\|_0$$
$$\leq c\left(\|L^\lambda(x,D)u\|_{m_0,a^2} + \|u\|_0\right),$$

which proves the theorem.

From Theorem 7.1 it follows that any solution $u \in L^2(\mathbb{R}^n)$ of the equation

$$L^\lambda(x,D)u = f, \qquad f \in C_0^\infty(\mathbb{R}^n), \tag{7.6}$$

belongs to $H^{a^2,m_0+1}(\mathbb{R}^n) \subset C_\infty(\mathbb{R}^n)$. Furthermore, by Theorem 4.2 and Theorem 5.2 we know that for $\lambda \geq d_0$ there exists for any $f \in L^2(\mathbb{R}^n)$ a unique solution $u \in H^{a^2,1}(\mathbb{R}^n)$ of (7.6).

We close this section with

Theorem 7.2. *For $\lambda \in \mathbb{R}$ the operator $L^\lambda(x,D)$ maps $H^{a^2,m_0+1}(\mathbb{R}^n)$ continuously into the space $H^{a^2,m_0}(\mathbb{R}^n)$.*

PROOF. Let $u \in C_0^\infty(\mathbb{R}^n)$. Then we have using Corollary 6.1 and Lemma 2.1

$$\|L^\lambda(x,D)u\|_{m_0,a^2}$$
$$\leq \|(1+a^2(D))^{m_0}\sum_{j=1}^n b_j(\cdot)a_j^2(D_j)u\|_0 + |\lambda|\,\|u\|_{m_0,a^2}$$
$$\leq \sum_{j=1}^n \|b_j(\cdot)(1+a^2(D))^{m_0}a_j^2(D_j)u\|_0 + |\lambda|\,\|u\|_{m_0,a^2}$$
$$+ \sum_{j=1}^n \|[(1+a^2(D))^{m_0}, b_j(\cdot)]a_j^2(D_j)u\|_0$$

$$\leq c\,\|u\|_{m_0+1,a^2} + |\lambda|\,\|u\|_{m_0,a^2} + c'\,\|u\|_{m_0+1,a^2} + \tilde{c}\,\|u\|_0$$
$$\leq \overline{c}\,\|u\|_{m_0+1,a^2}\ .$$

REMARK 7.1. The proofs of Theorem 7.1 and Theorem 7.2 together with assumptions B.1-B.4 show that both theorems hold for any $k \in \mathbb{N}$, $k \leq m_0$, instead of m_0.

8. On the operator $[L^\lambda]^{m_0}$.

Let $L^\lambda(x,D)$ be as in the previous section. In order to apply results of [17] and [25], see also [16], we need a characterization of $D([L^\lambda]^{m_0})$, where L^λ is the Friedrichs extension of $L^\lambda(x,D)$, see Section 4. Since L^μ is self-adjoint we can define the operator $[L^\mu]^k$ using the functional calculus or by iteration. It is well known, see [12, Corollary XII.2.8., p. 1200], that these two definitions coincide and that $[L^\mu]^k$ is a closed operator on its domain $D([L^\mu]^k)$. Furthermore we have (see [12, Definition XII, 1.1, p. 1186])

$$D([L^\mu]^k) = \left\{u \in D([L^\mu]^{k-1}):\quad [L^\mu]^{k-1}u \in D(L^\mu)\right\} \tag{8.1}$$

and

$$D([L^\mu]^k) = \left\{u \in D([L^\mu]):\quad L^\mu u \in D([L^\mu]^{k-1})\right\}\ . \tag{8.2}$$

Now we claim

Theorem 8.1. *Let L^μ be the Friedrichs extension of the operator $L^\mu(x,D)$, where $L(x,D)$ satisfies the assumptions* B.1-B.4. *Then we have for any $k \leq m_0$*

$$D([L^\mu]^k) = H^{a^2,k}(\mathbb{R}^n)\ . \tag{8.3}$$

PROOF. We prove (8.3) by induction. For $k = 1$ (8.3) was proved in Theorem 5.2. Next we prove that $D([L^\mu]^k) \subset H^{a^2,k}(\mathbb{R}^n)$ provided we know that $D([L^\mu]^{k-1}) = H^{a^2,k-1}(\mathbb{R}^n)$. Let $u \in D([L^\mu]^k)$. Then we have $[L^\mu]^k u = [L^\mu]^{k-1} L^\mu u$ and $L^\mu u \in D([L^\mu]^{k-1}) = H^{a^2,k-1}(\mathbb{R}^n)$. But by Theorem 7.1 and Remark 7.1 it follows now that $u \in H^{a^2,k}(\mathbb{R}^n)$. Finally let us prove that $H^{a^2,k}(\mathbb{R}^n) \subset D([L^\mu]^k)$ assuming that

$D([L^\mu]^{l-1}) = H^{a^2,l-1}(\mathbb{R}^n)$ for $l \leq k$. Let $u \in H^{a^2,k}(\mathbb{R}^n)$, then by Theorem 7.2 and Remark 7.1 we find that $[L^\mu]^{k-1}u \in H^{a^2,1}(\mathbb{R}^n)$ but $H^{a^2,1}(\mathbb{R}^n) = D(L^\mu)$, which by (8.1) proves the theorem.

By the Sobolev embedding theorem we get

Corollary 8.1. *Let L^μ as above. Then we have by* (3.5)

$$D([L^\mu]^{m_0}) \subset C_\infty(\mathbb{R}^n)\,. \tag{8.4}$$

9. On the Feller semigroup generated by $-L^\lambda(x,D)$.

By definition a Feller semigroup on $\mathbb{R}^n$ is a family of linear operators $(T_t)_{t\geq 0}, T_t : C_\infty(\mathbb{R}^n) \to C_\infty(\mathbb{R}^n)$, satisfying the following conditions

F.1. For all $t, s \geq 0$ we have $T_{s+t} = T_s T_t$ and $T_0 = I$.

F.2. For all $u \in C_\infty(\mathbb{R}^n)$ it follows that $\lim_{t\to 0} \|T_t u - u\|_\infty = 0$.

F.3. Let $u \in C_\infty(\mathbb{R}^n)$ and $0 \leq u \leq 1$ in $\mathbb{R}^n$. Then it is required that $0 \leq T_t u \leq 1$ holds for all $t \geq 0$.

The generator of a Feller semigroup is the operator

$$Au = \lim_{t\to 0} \frac{T_t u - u}{t}\,, \tag{9.1}$$

which is defined on $D(A) \subset C_\infty(\mathbb{R}^n)$, where $D(A)$ consists of all $u \in C_\infty(\mathbb{R}^n)$ such that (9.1) exists. The following theorem, often called the Yosida-Hille-Ray Theorem, will be of greater importance to us.

Theorem 9.1. ([7, p. 3-44], or [13, p. 165]) *Let $D(A)$ be a linear subspace of $C_\infty(\mathbb{R}^n)$ and let $A : D(A) \to C_\infty(\mathbb{R}^n)$ be a linear operator. Suppose further that $D(A)$ is dense in $C_\infty(\mathbb{R}^n)$, that A satisfies the positive maximum principle on $D(A)$, i.e. if $u \in D(A)$ and $x_0 \in \mathbb{R}^n$ such that $\sup_{x\in\mathbb{R}^n} u(x) = u(x_0) \geq 0$ then it follows that $Au(x_0) \leq 0$, and suppose that for some $\lambda \geq 0$ the operator $\lambda I - A$ maps $D(A)$ onto a dense subspace of $C_\infty(\mathbb{R}^n)$. Then A has a closed extension which is the generator of a Feller semigroup.*

It was Ph.Courrège who gave a characterization of operators satisfying the positive maximum principle on $C_0^\infty(\mathbb{R}^n)$.

Theorem 9.2. ([6, p. 2-36]) *Let $a : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be a continuous function such that for each $x \in \mathbb{R}^n$ the function $\xi \mapsto a(x,\xi)$ is negative definite. Then the operator $-a(x,D)$ defined on $C_0^\infty(\mathbb{R}^n)$ by*

$$-a(x,D)u(x) = (2\pi)^{-n/2} \int_{\mathbb{R}^n} e^{ix\xi}(-a(x,\xi))\,\hat{u}(\xi)\,d\xi \tag{9.2}$$

satisfies the positive maximum principle on $C_0^\infty(\mathbb{R}^n)$.

Theorem 9.3. *Suppose that $L(x,D)$ satisfies* B.1-B.4. *Then for each $x \in \mathbb{R}^n$ the function*

$$\xi \mapsto \sum_{j=1}^{n} b_j(x) a_j^2(\xi_j) \tag{9.3}$$

is negative definite and $-L(x,D)$ satisfies the positive maximum principle as an operator defined on $H^{a^2,m_0+1}(\mathbb{R}^n)$.

PROOF. The fact that the function defined by (9.3) is negative definite follows from Definition 1.1 and (3.10). By (3.5) we know that $H^{a^2,m_0+1}(\mathbb{R}^n) \subset C_\infty(\mathbb{R}^n)$. Furthermore, Theorem 7.2 combined with the Sobolev embedding theorem gives

$$\|L(x,D)u\|_\infty \leq c\,\|u\|_{m_0+1,a^2}\,. \tag{9.4}$$

Now let $u \in H^{a^2,m_0+1}(\mathbb{R}^n) \subset C_\infty(\mathbb{R}^n)$ and $x_0 \in \mathbb{R}^n$ such that $u(x_0) = \sup_{x\in\mathbb{R}^n} u(x) \geq 0$. Take $\chi \in C_0^\infty(\mathbb{R}^n)$ such that $\chi(x_0) = 1$ and $\chi|_{\mathbb{R}^n\setminus\{x_0\}} < 1$. Then for any $\eta > 0$ the function $u + \eta\chi$ belongs to $H^{a^2,m_0+1}(\mathbb{R}^n)$, $\sup_{x\in\mathbb{R}^n}(u+\eta\chi)(x) = u(x_0) + \eta > 0$ and

$$(u+\eta\chi)|_{\mathbb{R}^n\setminus\{x_0\}} < u(x_0) + \eta\,. \tag{9.5}$$

Let $(\varphi_\nu^\eta)_{\nu\in\mathbb{N}}$, $\varphi_\nu^\eta \in C_0^\infty(\mathbb{R}^n)$, be a sequence converging to $u+\eta\chi$ in $H^{a^2,m_0+1}(\mathbb{R}^n)$ and therefore also in $C_\infty(\mathbb{R}^n)$. Denote by $x_\nu \in \mathbb{R}^n$ a point defined by $\varphi_\nu^\eta(x_\nu) = \sup_{x\in\mathbb{R}^n}\varphi_\nu^\eta(x)$. Since $\varphi_\nu^\eta \to u+\eta\chi$ in $C_\infty(\mathbb{R}^n)$ it follows that $\varphi_\nu^\eta(x_\nu) \to u(x_0)+\eta$. We claim that a subsequence of $(x_\nu)_{\nu\in\mathbb{N}}$ converges to x_0. If no subsequence of $(x_\nu)_{\nu\in\mathbb{N}}$ converges to x_0, then there exists an open neigbourhood $U_\delta(x_0)$ such that at most a finite number of members of that sequence lie in $U_\delta(x_0)$. By (9.5) we can find some ε, $0 < \varepsilon < \eta$, such that

$$(u+\eta\chi)|_{\mathbb{R}^n\setminus U_\delta(x_0)} < u(x_0) + \eta - \varepsilon\,.$$

But this is a contradiction to the fact that $\varphi_\nu^\eta(x_\nu) \to u(x_0) + \eta$. In the following we may suppose that the whole sequence $(x_\nu)_{\nu\in\mathbb{N}}$ converges to x_0, since otherwise we have to take a subsequence. Since we have $\varphi_\nu^\eta(x_\nu) \to u(x_0) + \eta \geq \eta > 0$, we can also suppose that $\varphi_\nu^\eta(x_\nu) \geq 0$ for all $\nu \in \mathbb{N}$. By Theorem 9.2 the operator $-L(x,D)$ satisfies the positive maximum principle on $C_0^\infty(\mathbb{R}^n)$. Thus we have $-L(x,D)\varphi_\nu^\eta(x_\nu) \leq 0$. But this implies $-L(x,D)(u+\eta\chi)(x_0) \leq 0$, where we used (9.4) and the convergence properties of $(\varphi_\nu^\eta)_{\nu\in\mathbb{N}}$. Thus we have for any $\eta > 0$ that

$$-L(x,D)u(x_0) \leq \eta\,(L(x,D)\chi)(x_0)\,,$$

and for $\eta \to 0$ the theorem follows.

Now from Theorem 4.2, Theorem 5.2, Theorem 7.1 and Theorems 9.1-9.3 we get

Theorem 9.4. *Let $L(x,D)$ satisfy the assumptions of Theorem 9.3. Then for all $\lambda \geq 0$ the operator*

$$-L(x,D) : H^{a^2,m_0+1}(\mathbb{R}^n) \to H^{a^2,m_0}(\mathbb{R}^n) \subset C_\infty(\mathbb{R}^n)$$

has a closed extension which is the generator of a Feller semigroup on $\mathbb{R}^n$.

An immediate consequence of Theorem 9.4 is

Corollary 9.1. *Suppose that $L(x,D)$ satisfies the assumptions of Theorem 9.3 and is symmetric. Then B_λ is a regular Dirichlet form with domain $H^{a^2,1/2}(\mathbb{R}^n)$.*

10. Examples.

In this section we want to give examples of operators $L(x,D)$ we can apply to Theorem 9.4 and the results of the theorems leading to Theorem 9.4. For this we have to recall some basic properties of continuous negative definite functions. Our standard reference is the book [1]. First we want to note the following representation formula, see [9, p. 5-9] or [1, p. 184].

Theorem 10.1. (Lévy-Khinchin Formula) *Every real-valued continuous negative definite function $a : \mathbb{R}^n \to \mathbb{R}$ has the following representation*

$$a(\xi) = c + Q(\xi) + \int_{\mathbb{R}^n} (1 - \cos(\xi, \eta)) \frac{1 + |\eta|^2}{|\eta|^2} d\sigma(\eta)\,, \tag{10.1}$$

where $c \geq 0$, Q is a non-negative quadratic form on $\mathbb{R}^n$ and σ is a positive measure on $\mathbb{R}^n$ which does not charge the origin and has finite total mass. Conversely, given c, Q, and σ with the properties mentioned above, then the function a defined by (10.1) *is a continuous negative definite function.*

Sometimes it is convenient to consider continuous negative definite functions of the form

$$a(\xi) = \int_{\mathbb{R}^n} (1 - \cos(\xi, \eta)) k(\eta)\, d\eta\,. \tag{10.2}$$

We will call k the kernel associated with a. For $0 < s < 1$ an example is $\lambda^{2s}(\xi) = |\xi|^{2s}$, where the associated kernel is given by $K_{2s}(\eta) = c(n, s)|\eta|^{-n-2s}$.

Clearly the set of all continuous negative definite functions forms a convex cone. Further, if $a_j : \mathbb{R}^{n_j} \to \mathbb{R}$, $j = 1, 2$, are two continuous negative definite functions then the function $a : \mathbb{R}^{n_1+n_2} \to \mathbb{R}$ defined by $(\xi, \eta) \mapsto a_1(\xi) + a_2(\eta)$ is again a continuous negative definite function. This fact is verified by a direct calculation using (1.3). Since for $0 < s \leq 1$ the function λ^{2s} is a continuous negative definite one, it follows that for s_j, $0 < s_j \leq 1$, and $b_j \geq 0$ by $\xi \mapsto \sum_{j=1}^n b_j \lambda^{2s_j}(\xi_j)$ a continuous negative definite function is given. It is rather easy to construct continuous negative definite functions on $\mathbb{R}$. By Proposition 10.6 in [1] any continuous function $a : \mathbb{R} \to [0, \infty)$ which is even and when restricted to $[0, \infty)$ increasing and concave is negative definite.

From the previous considerations it follows that for any choice of t_j and r_j, $0 < t_j \leq r_j \leq 2$, there are a lot of continuous negative definite functions $a_j^2 : \mathbb{R} \to \mathbb{R}$, $1 \leq j \leq n$, satisfying

$$c_j\, (1 + |\xi_j|^2)^{t_j/4} \leq (1 + a_j^2(\xi_j))^{1/2} \leq \tilde{c}_j\, (1 + |\xi_j|^2)^{r_j/4}\,. \tag{10.3}$$

Since the square root of a_j^2 is again a continuous negative definite function, we can now start to construct an operator $L(x, D)$ satisfying the

assumptions of Theorem 9.3. It is clear that

$$a^2(\xi) = \sum_{j=1}^{n} a_j^2(\xi_j) \tag{10.4}$$

is a continuous negative definite function on $\mathbb{R}^n$ and that

$$c(1+|\xi|^2)^{t/2} \leq (1+a^2(\xi)) \leq \tilde{c}(1+|\xi|^2)^{r/2} \tag{10.5}$$

holds, where $t = \min_{1\leq j\leq n} t_j$ and $r = \max_{1\leq j\leq n} r_j$. Thus given t and n, we can find m_0 such that (3.5) holds. Then we will determine $r \geq t$ such that $0 < 1 - r/(2t) - r(r-t)m_0/t^2$ holds, *i.e.* $r \in [t, t(1+\overline{m})\wedge 2]$, where

$$\overline{m} = \frac{1}{4m_0}\left(((2m_0+1)^2+8m_0)^{1/2} - (2m_0+1)\right).$$

Now we take continuous negative definite functions $a_j^2 : \mathbb{R} \to \mathbb{R}$ satisfying (10.3), where $t \leq t_j \leq r_j \leq r$. Note that $r = t$ is the elliptic case, *i.e.* in that case we will handle an elliptic pseudodifferential operator. Then it is possible to let m_0 tend to infinity and the regularity results are just (hypo-)elliptic regularity results. In particular, we get for coefficients $b_j = d_j + c_j$, $d_j \in S(\mathbb{R}^n)$, satisfying B.1-B.4, that any solution $u \in L^2(\mathbb{R}^n)$ of the equation $L^\lambda(x,D)u = f$, $f \in H^\infty(\mathbb{R}^n)$, lies also in $H^\infty(\mathbb{R}^n)$. However, when we want to handle non-elliptic operators m_0 must be finite and therefore in general we do not get hypoellipticity results.

Now, taking $b_j : \mathbb{R}^n \to \mathbb{R}$ satisfying B.1-B.4 it is clear that the operator

$$\sum_{j=1}^{n} b_j(x) a_j^2(D_j) \tag{10.6}$$

satisfies all assumptions of Theorem 9.3, in particular all results proved in sections 4-8 do hold.

11. Some probabilistic consequences.

In this section we will show that the validity of estimate (4.6) is not only helpful in constructing a Feller semigroup or a Dirichlet form, but once it is known that B_λ is a symmetric Dirichlet form on $H^{a^2,1/2}(\mathbb{R}^n)$

and that the continuous negative definite function a^2 satisfies (3.8), then (4.6) has some probabilistic consequences for the stochastic process associated with B_λ. We will consider two of these consequences.

For the first one we recall a result due to M. Fukushima [14].

Theorem 11.1. ([14], Theorem 2) *Suppose a symmetric Dirichlet form E on $L^2(\mathbb{R}^n)$ is regular and satisfies*

$$\|u\|_{L^q}^2 \leq c(E(u,u) + c_0 \|u\|_0^2) \tag{11.1}$$

for some $q > 2$ and $c_0 \geq 0$. Then the associated standard Markov process M possesses the following properties. There exists a Borel set N of zero capacity such that $\mathbb{R}^n \setminus N$ is M-invariant and further the following assertions hold:

i) The resolvent kernel $R_\mu(x,\cdot)$ is absolutely continuous with respect to the Lebesgue measure for each $\mu > 0$ and $x \in \mathbb{R}^n \setminus N$.

ii) The transition function $p_t(x,\cdot)$ is absolutely continuous with respect to the Lebesgue measure for each $t > 0$ and $x \in \mathbb{R}^n \setminus N$.

iii) A set $A \subset \mathbb{R}^n \setminus N$ is of zero capacity if and only if A is polar, that is, almost all sample paths starting at $x \in \mathbb{R}^n \setminus N$ do not hit A at positive time.

Now we claim

Theorem 11.2. *Let M^λ be the standard Markov process associated with B_λ, where B_λ is generated by $L^\lambda(x,D)$ which is assumed to be symmetric and to fulfill the assumptions of Theorem* 9.3. *Then the assertions of Theorem* 11.1 *do hold for M^λ.*

PROOF. It remains to prove that

$$\|u\|_{L^q}^2 \leq c(B_\lambda(u,u) + c_0 \|u\|_0^2) \tag{11.2}$$

holds for some $q > 2$ and a constant $c_0 \geq 0$. But combining (3.8) with (4.6) we have with a suitable constant c

$$\|u\|_{t/2}^2 \leq c(B_\lambda(u,u) + d_0 \|u\|_0^2)$$

for all $u \in H^{a^2,1/2}(\mathbb{R}^n) = D(B_\lambda)$. Now applying the Sobolev inequality (see [38, p. 20]) we find

$$\|u\|_{L^q} \leq c' \|u\|_{t/2}$$

with $q = 2n/(n-t)$, hence $q > 2$. Thus with this value of q we get (11.2).

Note that one can use (11.1) to get L^∞-bounds for the resolvent of the semigroup generated by $-L^\lambda$, see [14] and also [20].

Our second application is concerned with the asymptotic behaviour of the semigroup $(T_t)_{t>0}$ generated by $-L^\lambda$ on $L^2(\mathbb{R}^n)$. In their work [5] E. Carlen, S. Kusuoka and D. Stroock proved the following result which we formulate here for our special situation again assuming that B_λ is symmetric.

Theorem 11.3. ([5], Theorem 2.16) *Let $\nu \in (2,\infty)$ and $q = 2\nu/(\nu-2) > 2$. Suppose further that with some constants c_1 and c_2*

$$\|u\|_{L^q}^2 \leq c_1 \left(B_\lambda(u,u) + c_2 \|u\|_0^2\right) \tag{11.3}$$

holds for all $u \in H^{a^2,1/2}(\mathbb{R}^n)$. Then there exist constants c_1' and c_2' such that the Nash-type inequality

$$\|u\|_0^{2+4/\nu} \leq c_1' \left(B_\lambda(u,u) + c_2' \|u\|_0^2\right) \|u\|_{L^1}^{4/\nu} \tag{11.4}$$

holds.

Further they showed the next theorem, which we again state only in a formulation convenient for our purposes.

Theorem 11.4 ([5], Theorem 2.2) *Suppose that (11.4) holds for all $u \in H^{a^2,1/2}(\mathbb{R}^n)$. Then there exists a constant $d > 0$ such that for the semigroup $(T_t^\lambda)_{t>0}$ generated by B_λ on $L^2(\mathbb{R}^n)$ the estimate*

$$\|T_t^\lambda\|_{L^1 \to L^\infty} \leq d \frac{e^{c_2' t}}{t^{\nu/2}} \tag{11.5}$$

holds for $t > 0$. Here $\|\cdot\|_{L^1 \to L^\infty}$ denotes the operator norm for continuous linear operators mapping $L^1(\mathbb{R}^n)$ into $L^\infty(\mathbb{R}^n)$.

Again using (3.8) and (4.6) we get by the Sobolev embedding theorem estimate (11.3). Thus (11.5) follows with $\nu = 2n/t > 2$. In [5] further results related to these of Theorem 11.3 and Theorem 11.4 are given.

We also want to mention that Theorem 8.1 enables us to apply some results of the theory of (r,p)-capacities developed by M.

Fukushima and H. Kaneko in [17] and by H. Kaneko in [25]. For details we refer to the paper [16].

Finally let us remark that it seems to us to be possible to use some of the Feller semigroups constructed in this paper to obtain examples for balayage spaces in the sense of [3].

Acknowledgement. Parts of this work had been done while the author, supported by DFG-contract Ja 522/1-1, was visiting the University of Osaka. The author wants to thank Prof. M. Fukushima for his hospitality and many discussions about the topics of this paper. Further the author wants to thank W. Hoh, Erlangen, for discussions and support while writing this paper.

NOTE ADDED IN THE PROOFS. Meanwhile some of our results had been improved. A report on these results is given in N. Jacob: Pseudodifferential operators with negative definite functions as symbol: Applications in probability theory and mathematical physics. In *Operator theory: Advances and Applications* **57** (1992), 149-161.

References.

[1] Berg, C. and Forst, G. *Potential theory on locally compact Abelian groups.* Springer Verlag, 1975.

[2] Beurling, A. and Deny, J. Dirichlet spaces. *Proc. Natl. Acad. Sci. U.S.A.* **45** (1959), 208-215.

[3] Bliedtner, J. and Hansen, W. *Potential theory - An analytic and probabilistic approach to balayage.* Springer Verlag, 1986.

[4] Bony, J. M., Courrège, Ph., and Priouret, P. Semi-groupes de Feller sur une variété à bord compacte et problème aux limites intégrodifférentiels du second ordre donnant lieu au principe du maximum. *Ann. Inst. Fourier,* **18** (1968), 369-521.

[5] Carlen, E.A., Kusuoka, S., and Stroock, D.W. Upper bounds for symmetric Markov transition functions. *Ann. Inst. Henri Poincaré, Probabilités et Statistique,* Sup. au n° 2, **23** (1987), 245-287.

[6] Courrège, Ph. Sur la forme intégro-différentielle des opérateurs de C_K^∞ dans C satisfaisant du principe du maximum. *Sém. Théorie du Potentiel* (1965/66). 38 p.

[7] Courrège, Ph. Sur la forme intégro-différentielle du générateur infinitésimal d'un semi-groupe de Feller sur une variété. *Sém. Théorie du Potentiel* (1965/66), 48 p.

[8] Demuth, M., Van Casteren, J. A. On spectral theory of self-adjoint Feller generators. *Rev. Math. Phys.* **1** (1989), 325-414.

[9] Deny, J. Sur les espaces de Dirichlet. *Sém. Théorie du Potentiel* (1957), 12 p.

[10] Deny, J. Méthodes Hilbertiennes et théorie du potentiel, in *Potential Theory, C.I.M.E., Roma* (1970), 123-201.

[11] Doppel, K. and Jacob, N. Zur Konstruktion periodischer Lösungen von Pseudodifferentialgleichungen mit Hilfe von Operatorenalgebren. *Ann. Acad. Sci. Fenn. Ser.* A.I. *Math.* **8** (1983), 193-217.

[12] Dunford, N., Schwartz, J. T. *Linear Operators,* II., John Wiley Interscience Publ., 1963.

[13] Ethier, S. N. and Kurtz, Th. G. *Markov processes - characterization and convergence.* John Wiley & Sons, 1985.

[14] Fukushima, M. On an L^p-estimate of resolvents of Markov processes. *Publ. R.I.M.S.* **13** (1977), 277-284.

[15] Fukushima, M. *Dirichlet forms and Markov processes.*, North Holland Pub. Co., 1980.

[16] Fukushima, M., Jacob, N., and Kaneko, H. On $(r, 2)$-capacities for a class of elliptic pseudodifferential operators. *Math. Ann.* **293** (1992), 343-348.

[17] Fukushima, M. and Kaneko, H. On (r,p)-capacities for general Markov semi-groups. *Infinite-dimensional analysis and stochastic processes.* Proc. USP-meeting at Bielefeld 1983, Pitman Research Notes in Math. **124**, (1985), 41-47.

[18] Hörmander, L. Pseudodifferential operators. *Comm. Pure Appl. Math.* **18** (1965), 501-517.

[19] Jacob, N. Dirichlet forms and pseudodifferential operators. *Expo. Math.* **6** (1988), 313-351.

[20] Jacob, N. L^∞-bounds for solutions of pseudodifferential boundary problems in Dirichlet spaces. *Expo. Math.* **6** (1988), 363-371.

[21] Jacob, N. A Gårding inequality for certain anisotropic pseudodifferential operators with non-smooth symbols. *Osaka J. Math.* **26** (1989), 857-879.

[22] Jacob, N. Commutator estimates for pseudodifferential operators with negative definite functions as symbols. *Forum Math.* **2** (1990), 155-162.

[23] Jacob, N. Feller semigroups, Dirichlet forms, and pseudodifferential operators. *Forum Math.* **4** (1992), 433-446.

[24] Jacob, N. A class of elliptic pseudodifferential operators generating symmetric Dirichlet forms. *Potential Anal.* **1** (1992), 221-232.

[25] Kaneko, H. On (r,p)-capacities for Markov processes. *Osaka J. Math.* **23** (1986), 325-336.

[26] Kato, T. *Perturbation theory for linear operators.*, Springer Verlag, 1966.

[27] Kochubei, A. N. Parabolic pseudodifferential equations, hypersingular integrals, and Markov processes. *Math. U.S.S.R. Izvestiya* **33** (1989), 233-259.

[28] Kohn, J. J. and Nirenberg, L. On the algebra of pseudodifferential operators. *Comm. Pure Appl. Math.* **18** (1965), 269-305.

[29] Komatsu, T. Markov processes associated with certain integrodifferential operators. *Osaka J. Math.* **10** (1973), 271-303.

[30] Komatsu, T. On the martingale problem for generators of stable processes with perturbations. *Osaka J. Math.* **21** (1984), 113-132.

[31] Komatsu, T. Pseudodifferential operators and Markov processes. *J. Math. Soc. Japan* **36** (1984), 387-418.

[32] Komatsu, T. Continuity estimates for solutions of parabolic equations associated with jump type Dirichlet forms. *Osaka J. Math.* **25** (1988), 697-728.

[33] Kumano-go, H. *Pseudodifferential operators.* M.I.T. Press, 1981.

[34] Lions, J. L. and Magenes, E. *Nonhomogeneous boundary value problems and applications I.*, Springer-Verlag, 1972.

[35] Oleinik, O. A. and Radkevic, E. V. *Second order equations with nonnegative characteristic form.* Amer. Math. Soc. Plenum Press, 1973.

[36] Samko, S. G. *Hypersingular integrals and their applications.* Izdat. Rostov Univ., Rostov-on Don, 1984 (in Russian).

[37] Taira, K. *Diffusion processes and partial differential equations.* Academic Press, 1988.

[38] Taylor, M. *Pseudodifferential operators*, Princeton University Press, 1981.

[39] von Waldenfels, W. *Eine Klasse stationärer Markowprozesse.* Berichte der Kernforschungsanlage Jülich, 1961.

[40] von Waldenfels, W. Fast positive Operatoren. *Z. Wahrscheinlichkeitstheorie verw. Geb.* **4** (1965), 159-174.

[41] Weidmann, J. *Lineare Operatoren in Hilberträumen.* Mathematische Leitfäden, B.G. Teubner, 1976.

Recibido: 5 de marzo de 1.992

Niels Jacob
Mathematisches Institut
Universität Erlangen-Nürnberg
D-8520 Erlangen, GERMANY

REVISTA MATEMÁTICA IBEROAMERICANA
VOL. 9, N.° 3, 1993

Aperiodicity of the Hamiltonian flow in the Thomas-Fermi potential

Charles L. Fefferman and Luis A. Seco

"...que para sacar una verdad en limpio
menester son muchas pruebas y repruebas."

"Don Quijote de la Mancha", **M. de Cervantes.**

In [FS1] we announced a precise asymptotic formula for the ground-state energy of a non-relativistic atom. The purpose of this paper is to establish an elementary inequality that plays a crucial role in our proof of that formula. The inequality concerns the Thomas-Fermi potential $V_{TF}(r) = -y(ar)/r$, $a > 0$, where $y(r)$ is defined as the solution of

$$\text{(1.1)} \qquad \begin{cases} y''(x) = x^{-1/2}\, y^{3/2}(x), \\ y(0) = 1, \\ y(\infty) = 0. \end{cases}$$

(Without loss of generality, in what follows we will take $a = 1$).

Define

$$F(\Omega) = F_y(\Omega) = \int \left(\frac{y(x)}{x} - \frac{\Omega^2}{x^2} \right)_+^{1/2} dx \ , \qquad \Omega \in (0, \Omega_c)$$

where

$$\Omega_c^2 = \sup_{r>0} u(r) = u(r_c)\,, \qquad u(r) = r\,y(r)\,.$$

The subscript for F will be used whenever we want to emphasize the dependence of F on y. Then, $F(\Omega)$ depends smoothly on Ω, [SW2], and our main result here is as follows:

Theorem 1.1.

$$F''(\Omega) \le c < 0\,, \qquad \textit{for all} \quad \Omega \in (0, \Omega_c)\,. \tag{1.2}$$

This is a quantitative form of the non-periodicity of almost all zero-energy orbits for the Hamiltonian

$$H = |\xi|^2 + V_{TF}(|x|)$$

on

$$\mathbb{R}^6 = \{(x,\xi):\ x \in \mathbb{R}^3,\ \xi \in \mathbb{R}^3\}\,.$$

In fact, an easy computation shows that a zero-energy orbit with angular momentum Ω is periodic if and only if the derivative $F'(\Omega)$ is a rational multiple of π (see [Ar]). Hence, Theorem 1.1 shows that closed zero-energy orbits arise for only countably many Ω.

Theorem 1.1 will be used in our later papers ([FS5] and [FS6]) to control the density and eigenvalue sum arising from the three dimensional Schrödinger operator

$$H_Z = -\Delta + Z^{4/3} V_{TF}\big(Z^{1/3}|x|\big)\,, \qquad \text{for large } Z\,.$$

Aperiodicity of zero-energy Hamiltonian paths is well-known to play a crucial role in the study of eigenvalues and eigenfunctions. In our setting, Theorem 1.1 enters because our formulas for the eigenvalue sum and density involve expressions of the form

$$S = \sum_{1 \le l < Z^{1/3}\Omega_c} \beta\left(\frac{Z^{1/3}}{\pi} F(Z^{-1/3} l)\right),$$

for elementary functions such as $\beta(t) = t - [t] - 1/2$. (Here $[t]$ is the greatest integer in t). Since β is bounded, we obtain trivially the estimate $S = O\big(Z^{1/3}\big)$. If $F(\Omega) = \pi\mu\,\Omega + \nu$ with μ rational, then the trivial

estimate for S is easily seen to be the best possible. On the other hand, if $d^2F/d\Omega^2 < c < 0$, then one can prove that the numbers

$$\phi_l = Z^{1/3} F(Z^{-1/3} l)$$

are equidistributed modulo π. (The argument is close to Hardy's estimates on the number of lattice points in a disc.) Since $\beta(t)$ is periodic and has average zero, it follows that $S = O(Z^\gamma)$ with $\gamma < 1/3$.

Thus, Theorem 1.1 allows us to improve on the trivial estimate for the sum S, which appears in the eigenvalue sum and density for H_Z. The complete proof of our results on atoms is contained in this paper together with [FS2], [FS3], [FS4], [FS5], [FS6] and [FS7].

The proof of Theorem 1.1 is necessarily rather delicate. For small perturbations of V_{TF} in a natural topology, the analog of Theorem 1.1 fails. Therefore, we have to make strong use of the differential equation defining $y(r)$. Our proof uses computer-assisted methods to solve that equation and to obtain bounds for F''. We remark, however, that without a computer it can also be seen that F'' vanishes at most finitely many times (Proposition 4.8 below; see also the recent independent proof in [HKSW]), which also implies that zero-energy periodic orbits have measure zero, which in turn also implies the same results stated above for sums S, and therefore our result for atomic energies. Theorem 1.1, however, is better because it implies better error terms for all those formulas. Moreover, if one wants to understand ground-state energies to a greater accuracy, then Theorem 1.1, with all its strength, is unavoidable.

In what follows, our proofs will *not* be computer-assisted unless stated otherwise.

It would be interesting to prove the aperiodicity of almost all zero-energy Hamiltonian paths in the Thomas-Fermi potential for a molecule.

The complete programs used in our proof are publicly available by anonymous `ftp` from the machine `math.utexas.edu` (Internet number 128.83.133.215) This machine also supports other standard methods of such as `gopher` and `wais`. The interested parties should contact their administrators about availability and usage of these programs on their machine. The machine `math.utexas.edu` has a user called `anonymous`

whose password is the e-mail address of the actual user. Our programs are stored in the directory `/pub/papers/feffseco`. We refer the reader to the file `README` there for instructions on how to download the programs. Each one of them has instructions on how to use them.

More information about how to interact with `math.utexas.edu` is available from the Mathematical Physics Preprint Archive. In particular, the user can obtain detailed instructions on how to install the public domain programs `gopher` and `wais`. Send e-mail to

`mp_arc@math.utexas.edu`

for details.

We also remark that the American Mathematical Society maintains the `e-math` account in the machine `e-math.ams.com` (Internet number 130.44.1.100). This account includes a menu, one of whose entries is gopher. At the moment, the mp_arc gopher connection is in the main menu. Going through different submenus, one can also reach the U.T. Math. gopher server. The user may find out other machines that provide public access to Internet services.

1. Preliminaries.

In this section we consider a smooth function y that looks like the Thomas-Fermi function. More precisely, let $u(x) = x\,y(x)$; then, we assume the following holds

a) $y > 0$, $y(0) = 1$ and $\lim_{x\to\infty} y(x) = 0$.

b) There exists a point r_c such that $u(x) < u(r_c)$ for $x \neq r_c$, $u'(x) > 0$ for $0 \leq x \leq r_c$, and $u'(x) < 0$ for $r_c \leq x$. Also, $u''(r_c) < 0$.

We will denote the two solutions of $u(r) = \Omega^2$ by $r_1(\Omega) < r_2(\Omega)$. We start by giving convenient formulas for the derivatives of F. We point out that similar formulas were given in [SW2]. One of the reasons we need formulas of the kind stated below is to obtain expressions such as (1.7) and (1.8) below. Also, we will see that in the case of an analytic y, not only is F analytic on $(0,\Omega_c)$, but it admits an analytic extension beyond Ω_c. However, 0 will be in general an essential singularity.

Lemma 1.2. *Let y be as above. The following formulas hold*

$$F(\Omega) = \int \left(u(x) - \Omega^2\right)_+^{1/2} \frac{dx}{x}\,,$$

$$F'(\Omega) = -\Omega \int \left(u(x) - \Omega^2\right)_+^{-1/2} \frac{dx}{x},$$

$$F''(\Omega) = -\lim_{\delta \to 0} \left(\int_{r_1(\Omega)+\delta}^{r_2(\Omega)-\delta} \left(u(x) - \Omega^2\right)^{-3/2} y(x)\, dx + c(\Omega)\, \delta^{-1/2} \right),$$

where $c(\Omega)$ is uniquely specified by requiring the finiteness of the limit.

Moreover, if b is any number less than $r_2(\Omega)$, then

$$\frac{d^2}{d\Omega^2} \int_{r_1(\Omega)}^{b} \left(u(x) - \Omega^2\right)_+^{1/2} \frac{dx}{x}$$

equals

$$-\lim_{\delta \to 0} \left(\int_{r_1(\Omega)+\delta}^{b} \left(u(x) - \Omega^2\right)^{-3/2} y(x)\, dx + c_1(\Omega)\, \delta^{-1/2} \right)$$

again, for a constant c_1 that makes the limit finite. The corresponding symmetric case also holds.

PROOF. The first two formulas are trivial. For the third, let

$$H(\delta, \Omega) = \Omega \int_{r_1(\Omega)+\delta}^{r_2(\Omega)-\delta} \left(u(r) - \Omega^2\right)^{-1/2} \frac{dr}{r}.$$

Note that the formula for F'' amounts to showing that

$$\frac{d}{d\Omega} \lim_{\delta \to 0} H(\delta, \Omega) = \lim_{\delta \to 0} \frac{d}{d\Omega} H(\delta, \Omega). \tag{1.3}$$

Indeed, the left hand side equals $-F''$, whereas the right hand side equals

$$\lim_{\delta \to 0} \Bigg\{ \Omega^2 \int_{r_1(\Omega)+\delta}^{r_2(\Omega)-\delta} \left(u(r) - \Omega^2\right)^{-3/2} \frac{dr}{r} + \int_{r_1(\Omega)+\delta}^{r_2(\Omega)-\delta} \left(u(r) - \Omega^2\right)^{-1/2} \frac{dr}{r}$$
$$+ \Omega \Big(\frac{\left(u(r_2 - \delta) - u(r_2)\right)^{-1/2}}{r_2 - \delta}\, r_2'(\Omega)$$
$$- \frac{\left(u(r_1 + \delta) - u(r_1)\right)^{-1/2}}{r_1 + \delta}\, r_1'(\Omega) \Big) \Bigg\}$$

$$= \lim_{\delta \to 0} \left\{ \int_{r_1(\Omega)+\delta}^{r_2(\Omega)-\delta} \left(u(r) - \Omega^2\right)^{-3/2} u(r) \frac{dr}{r} \right.$$
$$\left. - \Omega \sum_{i=1,2} \frac{|u'(r_i)|^{-1/2} |r_i'(\Omega)| \, \delta^{-1/2} (1 + O(\delta))}{r_i(\Omega)} \right\}$$

which agrees with the formula asserted for $-F''$, provided that this previous expression for $c(\Omega)$,

$$c(\Omega) = -\Omega \sum_{i=1,2} \frac{|u'(r_i)|^{-1/2} |r_i'(\Omega)|}{r_i(\Omega)} ,$$

actually makes the limit above finite.

Therefore, the lemma will follow if we show that both $H(\delta, \Omega)$ and $(\partial/\partial\Omega)H(\delta, \Omega)$ converge uniformly on compact subsets of $(0, \Omega_c)$ to C^1 functions. This will imply, first, that we can interchange limits in (1.3), and, second, that the expression for $c(\Omega)$ above is the right one.

In order to see this, consider the change of variables given by

$$t(r) = \begin{cases} \left(\Omega_c^2 - u(r)\right)^{1/2} , & \text{if } r \geq r_c , \\ -\left(\Omega_c^2 - u(r)\right)^{1/2} , & \text{if } r \leq r_c . \end{cases} \tag{1.4}$$

Note that t is smooth and strictly increasing in the range $(0, \infty)$. We can therefore consider its inverse, $r(t)$, and use it to rewrite

$$H(\delta, \Omega) = \Omega \int_{t_1(\delta,\Omega)}^{t_2(\delta,\Omega)} \left(D^2 - t^2\right)^{-1/2} w(t) \, dt ,$$

where

$$t_1 = t(r_1 + \delta) , \qquad t_2 = t(r_2 - \delta) , \qquad D^2 = \Omega_c^2 - \Omega^2 , \qquad w(t) = \frac{r'(t)}{r(t)} .$$

Note that w is smooth on $(0, \Omega_c)$, and that

$$\begin{aligned} t_1 &= -D\,(1 + \tau_1(\delta)) , \qquad t_2 = D\,(1 + \tau_2(\delta)) , \\ c\,\delta &\leq |\tau_i| \leq C\,\delta , \quad \text{for } i = 1, 2 , \end{aligned} \tag{1.5}$$

uniformly on compact subsets of $(0,\Omega_c)$, which implies that

$$H(\delta,\Omega)=\Omega\int_{D^{-1}t_1}^{D^{-1}t_2}\left(1-t^2\right)^{-1/2}w(tD)\,dt$$

converges uniformly to the C^1 function

$$H(0,\Omega)=\Omega\int_{-1}^{1}\left(1-t^2\right)^{-1/2}w(tD)\,dt=-F'(\Omega). \tag{1.6}$$

As for $(d/d\Omega)H(\delta,\Omega)$,

$$\frac{d}{d\Omega}H(\delta,\Omega)=\int_{D^{-1}t_1}^{D^{-1}t_2}(1-t^2)^{-1/2}\frac{\partial}{\partial\Omega}\Big(\Omega\,w(tD)\Big)\,dt+\Omega\sum_{i=1,2}G_i(\delta,\Omega)\,,$$

with

$$G_i(\delta,\Omega)=\pm\left(1-D^{-2}t_i^2\right)^{-1/2}w(t_i)\,\frac{\partial}{\partial\Omega}\left(D^{-1}t_i\right)\,,$$

the first term above converges with δ to the smooth function

$$\int_{-1}^{1}(1-t^2)^{-1/2}\frac{\partial}{\partial\Omega}\Big(\Omega\,w(tD)\Big)\,dt$$

uniformly on compact subsets of $(0,\Omega_c)$. Thus, the lemma will follow if we prove that G_i goes to zero with δ uniformly in Ω. By (1.5), this will in turn follow if we prove that

$$\frac{\partial}{\partial\Omega}\left(D^{-1}t_i\right)=O(\delta)\,.$$

By (1.5) again, it is enough to prove that

$$\frac{\partial}{\partial\Omega}\left(D^{-1}t_i\right)^2=O(\delta)\,.$$

But, for $i=1$,

$$\begin{aligned}\frac{\partial}{\partial\Omega}\left(D^{-1}t_1\right)^2&=\frac{\partial}{\partial\Omega}\left(\frac{u(r_1+\delta)-\Omega_c^2}{u(r_1)-\Omega_c^2}\right)\\&=\frac{\left(u(r_1)-\Omega_c^2\right)u'(r_1+\delta)r_1'(\Omega)-\left(u(r_1+\delta)-\Omega_c^2\right)u'(r_1)\,r_1'(\Omega)}{\left(u(r_1)-\Omega_c^2\right)^2}\end{aligned}$$

$$= \frac{r_1'(\Omega)}{\left(u(r_1) - \Omega_c^2\right)^2}$$

$$\cdot \Big(\big(u(r_1)u'(r_1+\delta) - u(r_1+\delta)u'(r_1)\big) - \Omega_c^2 \left(u'(r_1+\delta) - u'(r_1)\right) \Big) .$$

The first factor above is trivial. The other is clearly bounded by $C\delta$, and, doing the same for $i = 2$, the lemma follows.

The last remark in the statement of the lemma follows in exactly the same way, with the only modification that one of the G_i is in fact constant in δ, which of course does not affect the uniform approach to a C^1 function.

A closer look at (1.6) yields the following remark.

Corollary 1.3. *Define $w(t)$ as in the proof of the previous lemma. Then*

$$-F''(\Omega) = \int_{-1}^{1} (1-t^2)^{-1/2} \frac{\partial}{\partial \Omega} \big(\Omega\, w(tD)\big)\, dt .$$

In particular, if $y \in C^k(0,\infty)$, then $F_y \in C^{k-1}(0,\Omega_c)$, $k \geq 2$. Also, if y is analytic $F(\Omega)$ admits an analytic extension to a complex neighborhood of $(0,\Omega_c]$.

PROOF. If $y \in C^k$, the same is true for u. Therefore, $t \in C^{k-1}$, thus $r \in C^{k-1}(-\Omega_c,\Omega_c)$ and $r(t) \neq 0$, which implies $w \in C^{k-2}(-\Omega_c,\Omega_c)$, and, by (1.6), $F' \in C^{k-2}$.

In the case of an analytic y, since w is analytic in some neighborhood around 0, it admits a convergent power series expansion

$$w(t) = \sum_{n=0}^{\infty} w_n\, t^n , \qquad t < t_0 . \tag{1.7}$$

This implies

$$-F'(\Omega) = \Omega \sum_{n=0}^{\infty} w_{2n}\, D^{2n} \int_{-1}^{1} (1-t^2)^{-1/2} t^{2n}\, dt , \tag{1.8}$$

since the odd terms clearly yield an integral 0, and thus drop out of the sum. This, in particular shows that F can be defined as an analytic

function around Ω_c. Since, by (1.6), F is analytic also in $(0,\Omega_c)$, the corollary follows.

We will see later (Proposition 4.8) that the limit

$$\lim_{\Omega\to 0} F''(\Omega)\,\Omega^{\gamma}\,, \qquad \gamma = \frac{9-\sqrt{73}}{2} > 0\,,$$

exists, is finite and not zero. This shows, in particular, that F has an essential singularity at 0 and that F is not a linear function.

The proof of (1.2) will now go as follows: We make an initial division of $(0,\Omega_c)$ into two intervals $(0,\bar{\Omega})$ and $[\bar{\Omega},\Omega_c]$, that we will refer to as Zone I and Zone II, respectively. In Zone I, we will use the formula in Lemma 1.2 to prove (1.2) uniformly on very little subintervals of $(0,\bar{\Omega})$. We will deal with this in Section 4. Then, formula (1.8) will allow us to show (1.2) uniformly on Zone II, as explained in Section 5.

Our proof will rely on a very precise knowledge of the solution to the Thomas-Fermi equation. For this, we will use computer assisted techniques. The next section deals with a description of how the computer will be used to yield theorems.

2. Computer-Assisted Analysis.

Let $\mathcal{R}$ be the set of "representable numbers" in a computer, that is those numbers that the computer can represent exactly. Depending on the specific machine, they are usually real numbers with some finite binary expansion.

It is well known that computers can only perform arithmetic in an approximate way: the addition -for example- of two representable numbers is another representable number that will probably be close to the true sum, but is not exactly the true sum.

The idea to perform rigorous arithmetic is to instruct the computer on how to produce upper and lower bounds to the true results of arithmetic operations between representable numbers; in other words, we work with intervals with endpoints in $\mathcal{R}$, and we implement arithmetic operations on intervals in such a way that given two intervals, the

computer will produce a third that is guaranteed to contain the result of all arithmetic operations between points in the initial intervals. This is usually called "interval arithmetic".

We denote the set of all these intervals by $\mathcal{I}$. Also, given a real function $f(x)$, we denote

$$f(I) = \{f(x) : \ x \in I\}, \qquad I \in \mathcal{I}.$$

Binary functions of intervals are defined accordingly. In particular, a statement like $I_1 > I_2$ means that $x > y$ for all pairs (x,y), $x \in I_1$, $y \in I_2$. Also, given $I = [a,b]$ and $\varepsilon \geq 0$, we introduce the shorthand notation $I \pm \varepsilon$ to denote an interval containing $[a-\varepsilon, b+\varepsilon]$. We also point out, although it really is redundant, that in what follows, finite decimal expressions for numbers represent the rational numbers with exactly those decimal expansions.

The next step is to perform a similar kind of arithmetic, but where objects are functions in some Banach space, not numbers. A convenient Banach space to use in this theory is the space of piecewise analytic functions, with a lower bound on the size of the domains of analyticity.

Occasionally, it will be convenient to switch to genuine real variable theory, for which we will do our work on $C^0[-1,1]$. The reason for this is that inversion of functions in $\mathbb{R}^1$ is a little easier than the complex counterpart, mainly because the domain of definition problem is trivial in the real case. We remark though, that the use of C^0 is not essential, and the same analysis could be carried over to H^1 with a little more work.

More precisely, consider the Banach Algebras

$$H^1 = \left\{ f(z) : \ f(z) = \sum_{n=0}^{\infty} a_n z^n, \quad \sum_{n=0}^{\infty} |a_n| < \infty \right\}$$

and

$$C^0 = \{f(x) : \ f \text{ is continuous on } [-1,1]\}$$

with norms

$$\|f\|_1 = \sum_{n=0}^{\infty} |a_n|, \qquad \|f\|_\infty = \sup |f(x)|,$$

respectively. H^1 is a subspace of the set of analytic functions in the unit disk.

Then, our substitute for intervals are sets $\mathcal{U}^1(I_0, \ldots, I_N; C_h, C_g; k)$ of the form

$$\left\{ f(z) = \sum_{n=0}^{\infty} a_n z^n + z^k g(z) : \ a_n \in I_n, \quad 0 \le n \le N, \right. \tag{2.1}$$
$$\left. \sum_{n=N+1}^{\infty} |a_n| \le C_h, \quad \|g\|_1 \le C_g \right\}$$

where C_h and C_g are positive real numbers and I_n are intervals in the real line. The parameter k will generally be problem-dependent and fixed. For the computer implementation, C_h and C_g will run over the set of computer-representable numbers, and the intervals will be those with representable endpoints. We refer to C_h and C_g as high and general order error terms respectively, for obvious reasons. If intervals have nonempty interior and $C_h > 0$, or if $k = 0$ and $C_g > 0$, then these sets are in fact a neighborhood basis for the topology induced by $\|\cdot\|_1$. For this reason, we will refer to these $\mathcal{U}$ as "neighborhoods", even if in general they will not be. We will refer to them as neighborhoods of *type* k whenever we want to emphasize the integer k in definition (2.1). If $C_g = 0$, we refer to them as *type* ∞. In general, $\mathcal{U}(k)$ means that $\mathcal{U}$ is a neighborhood of type k. Also, we will refer to them as being of *order* N to indicate that they consist of $N+1$ intervals. In our implementation, N will not be fixed, but chosen adaptatively during the execution of the programs. The reason why this is a convenient space to work in is because elementary operations, such as addition, product, integration, differentiation (composed with a slightly contracting dilation), evaluation at a point and integration of initial value problems in ordinary differential equations can be conveniently bounded by elementary formulas in terms of this set of neighborhoods. By trivial scaling, we will be able to do analysis on

$$H^1(|z - z_0| \le r) = \left\{ f(z) : \ f(z) = \sum_{n=0}^{\infty} a_n \left(\frac{z - z_0}{r} \right)^n, \quad \sum_{n=0}^{\infty} |a_n| < \infty \right\},$$

a subspace of the set of analytic functions on the disk of center z_0 and radius r.

As for C^0, we will use sets (that we will also refer to as "neighborhoods") of the type:

$$\begin{aligned}&\mathcal{U}^0(I_0,\dots,I_N;C_h,C_g;k;S)\\&\quad=\Big\{f(z)=\sum_{n=0}^{N}a_n\,z^n+z^{N+1}h(z)+z^k g(z)\ :\\&\qquad a_n\in I_n,\quad 0\le n\le N,\\&\qquad \sup_{z\in S}|h(z)|\le C_h\,,\ \sup_{z\in S}|g(z)|\le C_g\Big\}\end{aligned}\tag{2.2}$$

where S is a subset of $[-1,1]$, and h and g are continuous functions on S. We will use the superscript 0 or 1 whenever we want to emphasize in which topology we are taking these "neighborhoods".

Note the natural inclusion

$$\mathcal{U}^1(I_0,\dots,I_N;C_h,C_g;k)\subset\mathcal{U}^0(I_0,\dots,I_N;C_h,C_g;k;S)\,,$$

for any $S\subset[-1,1]$.

These sets of neighborhoods $\mathcal{U}^0(k)$ will not allow us to perform as many operations as their smaller brothers the $\mathcal{U}^1(k)$, but we can still add, multiply, raise to fractional powers and integrate (among others) in terms of them; furthermore, the formulas for these neighborhood operations are exactly the same as those for the $\mathcal{U}^1(k)$.

We illustrate this neighborhood analysis describing how we can raise neighborhoods to real powers. At this point, we make the following remark concerning our use and description of algorithms:

Algorithms describe a procedure that, if successful, will allow us to construct (usually upper and lower bounds for) certain numbers. When we describe these algorithms, we will state under which conditions they *fail*; a failure means that the procedure is stopped, an error reported, and no theorem proved. Obviously, if during the description of an algorithm, we use another algorithm, a failure in the execution of the latter implies also a failure of the former algorithm.

Lemma 2.1. *Let* $0<r<1$. *Then*

$$\sup_{n\ge N}(nr^n)\le\begin{cases}Nr^N & \text{if } N\ge\dfrac{1}{|\log r|}\,,\\[2ex]\dfrac{1}{e\,|\log r|} & \text{otherwise}.\end{cases}$$

PROOF. The function $x\,r^x$ attains its maximum when $x = |\log r|^{-1}$.

Lemma 2.2. *Consider, in any commutative Banach Algebra, the operators*

$$T^\alpha(y) = (1+y)^\alpha$$

acting on $\|y\| \le r < 1$. *Then, we have*

$$\|T^\alpha(y)\| \le K_{2.2}(\alpha, \|y\|)\,,$$
$$\|T^\alpha\|_{\mathrm{Lip}} \le C_{2.2}(\alpha, r)\,,$$

where

$$K_{2.2}(\alpha, \|y\|) = \min\left\{(1-\|y\|)^{-|\alpha|}\,,\ 1 + |\alpha|\frac{\|y\|}{1-\|y\|}\right\}$$

if $-1 \le \alpha \le 2$,

$$K_{2.2}(\alpha, \|y\|) = (1-\|y\|)^{-|\alpha|}$$

otherwise, and

$$C_{2.2}(\alpha, r) = \min\left\{|\alpha| + r\frac{|\alpha|\,|\alpha-1|}{(1-r)^2}\,,\ |\alpha|(1-r)^{-|\alpha-1|}\right\}$$

if $-1 \le \alpha \le 2$,

$$C_{2.2}(\alpha, r) = |\alpha|(1-r)^{-|\alpha-1|}$$

otherwise.

PROOF. First, if $-1 \le \alpha \le 2$ and $\|y_1\|,\ \|y_2\| \le r$,

$$(1+y_1)^\alpha - (1+y_2)^\alpha = \sum_{n=1}^{\infty}\binom{\alpha}{n}(y_1^n - y_2^n)\,.$$

Now,

$$y_1^n - y_2^n = (y_1 - y_2)\sum_{k=0}^{n-1} y_1^k y_2^{n-1-k}$$

and

$$\left\|\sum_{k=0}^{n-1} y_1^k y_2^{n-1-k}\right\| \le n\,r^{n-1}\,.$$

Since $2 \geq \alpha \geq -1$, $\left|\binom{\alpha}{n}\right|$ is a decreasing sequence in n, for $n \geq 1$. Therefore,

$$\frac{\|T(y_1) - T(y_2)\|}{\|y_1 - y_2\|} \leq \sum_{n=1}^{\infty} \left|\binom{\alpha}{n}\right| n\, r^{n-1} \leq |\alpha| + r\frac{|\alpha|\,|\alpha - 1|}{(1-r)^2}\,.$$

On the other hand, for α in the same range,

$$\|T(y)\| \leq \sum_{n=0}^{\infty} \left|\binom{\alpha}{n}\right| \|y\|^n \leq 1 + |\alpha|\,\frac{\|y\|}{1 - \|y\|}\,.$$

Now, for general α, and $\|y_1\|$, $\|y_2\| \leq r$, note that

$$\|e^y\| \leq e^{\|y\|}$$

and

$$\|\log(1+y)\| \leq -\log(1 - \|y\|)\,.$$

Therefore,

$$\|T^\alpha(y)\| = \left\|e^{\alpha \log(1+y)}\right\| \leq (1 - \|y\|)^{-|\alpha|}$$

and

$$\begin{aligned}\|T^\alpha(y_1) - T^\alpha(y_2)\| &= |\alpha| \left\|(y_1 - y_2) \int_0^1 T^{\alpha-1}\big(ty_1 + (1-t)y_2\big)\, dt\right\| \\ &\leq |\alpha|\,\|y_1 - y_2\|\,(1-r)^{-|\alpha-1|}\,.\end{aligned}$$

Algorithm 2.3. *Given a neighborhood $\mathcal{U}(I_0, \ldots, I_N; C_h, C_g; k)$ satisfying*

1. $I_0 > 0$,
2. $|I_0| > \sum_{n>0} |I_n| + C_h + C_g$,
3. *If $\alpha > 2$, then $2N > \alpha - 1$,*

we construct another, $\tilde{\mathcal{U}}(k)$, such that, if $f \in \mathcal{U}$ then $f^\alpha \in \tilde{\mathcal{U}}(k)$.

The algorithm is independent of k, and of whether the neighborhoods are in H^1 or C^0.

If $C_g = 0$ for $\mathcal{U}$, then the same is true for $\tilde{\mathcal{U}}$.

DESCRIPTION: Assume first that $\alpha \geq -1$. Let $f \in \mathcal{U}$. Put $\tilde{f} = \big(f(0)\big)^{-1} f$, so $\tilde{f} = 1 + y(z) + z^k g(z)$, where $y(z) = z\,\tilde{y}(z)$,

$$1 + y(z) \in \mathcal{U}(I'_0, \ldots, I'_N; C'_h, 0; 0) \tag{2.3}$$

for $I'_i = f(0)^{-1} I_i$, $C'_h = f(0)^{-1} C_h$, and $\|g\| \leq C_g / f(0)$. Bounds for all this can be computed easily since we know that $f(0) \in I_0$.

Now,

$$\big(1 + y(z)\big)^\alpha = \sum_{n=0}^{N} \binom{\alpha}{n} y(z)^n + h(z)\,,$$

where $h(z) = z^{N+1}\tilde{h}(z)$. In the H^1 topology, $\|h\|_1 = \|\tilde{h}\|_1$, and

$$\|\tilde{h}\|_1 \leq \sum_{n>N} \left|\binom{\alpha}{n}\right| \|y\|_1^n \leq \left|\binom{\alpha}{N+1}\right| \frac{r^{N+1}}{1-r}$$

for

$$r = \sum_{i=1}^{N} |I'_i| + C'_h\,,$$

where we have used condition 3. in the statement of the algorithm. In the C^0 topology,

$$\left|\tilde{h}(z)\right| \leq \sum_{n>N} \left|\binom{\alpha}{n}\right| \left|\frac{y(z)}{z}\right|^n \leq \left|\binom{\alpha}{N+1}\right| \frac{r^{N+1}}{1-r}$$

for

$$r = \sum_{i=1}^{N} |I'_i| + C'_h\,.$$

As a result of this, the computation of $\|\tilde{h}\|$ is done in exactly the same way whether we are in the C^0 or H^1 topologies.

Concerning the computation of the factor $1/(1-r)$, it is done as follows: we first check that $r \in (0,1)$ the check for $r > 0$ being unnecessary, harmless but convenient; then, we compute an upper bound for $1/(1-r)$ with our interval arithmetic package, knowing that an overflow will be reported and the program terminated if we cannot find such upper bound with machine-numbers.

Also,

$$(1 + y(z))^\alpha - \big(\tilde{f}(z)\big)^\alpha = O(z^k)\,,$$

which implies that general errors are of type k. In the case that we are in H^1, since multiplication by z is an isomorphism, by Lemma 2.2, we see that general errors are bounded by

$$\left\|(1+y(z))^\alpha - \tilde{f}(z)^\alpha\right\| \leq \|g\|\ C_{2.2}(\alpha, \|y\| + \|g\|)\,.$$

If, however, we are in C^0, apply Lemma 2.2 to $(\mathbb{R}^1,+,\cdot)$, to get

$$\begin{aligned}\left|(1+y(z))^\alpha - \tilde{f}(z)^\alpha\right| &\leq \left|1+y(z)-\tilde{f}(z)\right|\ C_{2.2}\,(\alpha, \|g\|_\infty + \|y\|_\infty) \\ &\leq |z^k|\ \|g\|_\infty\ C_{2.2}\,(\alpha, \|g\|_\infty + \|y\|_\infty)\end{aligned}$$

since $|y(z)|$, $|\tilde{f}(z)-1| \leq \|g\|_\infty + \|y\|_\infty$ and $C_{2.2}(\alpha,t)$ is increasing in t.

Therefore, say that

$$\sum_{n=0}^{N} \binom{\alpha}{n} y(z)^n \in \mathcal{U}_1(\tilde{I}_0, \ldots, \tilde{I}_N; \tilde{C}_h, 0; \infty)$$

by (2.3). Then,

$$\tilde{f}^\alpha \in \mathcal{U}(\tilde{I}_0, \ldots, \tilde{I}_N; \tilde{\tilde{C}}_h, \tilde{\tilde{C}}_g; k)\,,$$

with

$$\tilde{\tilde{C}}_h = \tilde{C}_h + \left|\binom{\alpha}{N+1}\right| \frac{r^{N+1}}{1-r}$$

and

$$\tilde{\tilde{C}}_g = f(0)^{-1}\, C_g\ C_{2.2}\,(\alpha, \|y\| + C_g f(0)^{-1})$$

and

$$f^\alpha \in f(0)^\alpha \cdot \mathcal{U}(\tilde{I}_0, \ldots, \tilde{I}_N; \tilde{\tilde{C}}_h, \tilde{\tilde{C}}_g; k)\,.$$

In the case $\alpha < -1$, we can find an integer k such that $2^{-k}\alpha \geq -1$. Then, we can find a neighborhood containing $f^{2^{-k}\alpha}$. By ordinary multiplication we can thus construct a neighborhood containing

$$f^\alpha = \left(f^{2^{-k}\alpha}\right)^{2^k}\,.$$

Although computer-assisted analysis has become fairly standard, we refer the reader to [Mo] and [KM] for a description of the basic ideas. The technique for solving ODE's is adapted from [Se2] and [Se1], and

is tailored to handle our particular ODE. See [Lo] for a thorough discussion on ODE solving techniques, with very good general algorithms. Also, we refer the reader to [EKW], [EW], [FL], [LL], [Ll] and [Ra] for a sample of computer-assisted proofs of a wide variety of problems. Main ideas in our approach go back to those proofs.

Our interval arithmetic package is an adaptation of the one used in [Se1] and [Se2], which in turn is an adaptation of the one developed by D. Rana. See [Ra] and [Se1] for details on the software.

3. The Thomas-Fermi Equation.

In this section we will be concerned with the problem of getting good bounds for the solution of the Thomas-Fermi equation (1.1).

It is well known ([Hi]) that

$$-w_0 = \lim_{r\to 0} y'(r) < 0 \tag{3.1}$$

exists, and that y admits a power series expansion

$$y(r) = 144\, r^{-3}\Big(\sum_{n=0}^{\infty} b_n\, r^{-n\alpha}\Big) \tag{3.2}$$

convergent for r large enough, with $b_0 = 1$, $b_1 < 0$ and $\alpha = (\sqrt{73}-7)/2$.

Also, y is always positive, decreasing, and it is the only such solution of the ODE satisfying (3.1) and (3.2).

The Initial Value Problem away from the Singularities.

In this section we will be concerned with the solution to the Initial Value Problem

$$\begin{cases} u''(x) = x^{-1/2}\, u^{3/2}(x)\,, \\ u(x_0) = u_0\,, \\ u'(x_0) = u_1\,, \end{cases} \tag{3.3}$$

for $x_0, u_0 > 0$. The solution to this problem will be in terms of a function $f \in H^1$ satisfying

$$u(x) = u_0 + u_1\, r\, z + z^2 f(z)\,,$$

where $z = (x - x_0)/r$ and r is a small positive representable number (in particular, $r < x_0$).

Note that the solution of (3.3) can be viewed as the fixed point of

$$T(u) = u_0 + \int_{x_0}^{x} \left(u_1 + \int_{x_0}^{t} \frac{u^{3/2}(s)}{s^{1/2}}\, ds \right) dt$$

and that T induces in a trivial way an operator $\tilde{T}$ of which f is its fixed point.

Throughout this section, we will do our work on H^1, and $\|\cdot\|$ will always denote $\|\cdot\|_1$.

Algorithm 3.1. *We deduce conditions on u_0, u_1, x_0, r and α under which $\tilde{T}$ is a well-defined contraction in $B(0,\alpha) \subset H^1$, and we compute an upper bound for $\|\tilde{T}\|_{\mathrm{Lip}}$.*

DESCRIPTION: Let $g = \sum a_n z^n$. Consider the operators

$$\begin{aligned} T_1(f) &= r\,u_1\,z + z^2 f(z)\,, \\ T_2(g) &= (u_0 + g(z))^{3/2}\,, \\ T_3(g) &= (rz + x_0)^{-1/2} \cdot g\,, \\ T_4(g) &= r^2 \sum_{n\geq 0} \frac{a_n z^n}{(n+1)(n+2)}\,. \end{aligned} \tag{3.4}$$

It is clear that

$$T(u) = u_0 + r\,u_1\,z + z^2\,(T_4 \circ T_3 \circ T_2 \circ T_1)(f) \tag{3.5}$$

and thus, $\tilde{T} = T_4 \circ T_3 \circ T_2 \circ T_1$.

Now, T_1 is affine with an isometry as the linear part, and

$$\|T_4\|_{\mathrm{Lip}} \leq \frac{1}{2}\, r^2\,. \tag{3.6}$$

Using Lemma 2.2, and putting $\beta = r/x_0$, we can see that

$$\|T_3\|_{\mathrm{Lip}} \leq \left\| (rz + x_0)^{-1/2} \right\|_1 \leq x_0^{-1/2}\, K_{2.2}\Big(-\frac{1}{2}, \beta\Big)\,. \tag{3.7}$$

Here we assume $\beta < 1$, otherwise we say the algorithm fails. For T_2, we have

$$\|T_2\|_{\text{Lip}} \le u_0^{1/2}\, C_{2.2}\Big(\frac{3}{2}, \gamma\Big) \tag{3.8}$$

whenever

$$\gamma \ge u_0^{-1} \sup_{\|f\| \le \alpha} \|T_1(f)\| \,.$$

Since

$$u_0^{-1} \sup_{\|f\| \le \alpha} \|T_1(f)\| \le \frac{r|u_1| + \alpha}{u_0} \stackrel{\text{def}}{=} \gamma_0 \,,$$

we have

$$\|\tilde{T}\|_{\text{Lip}} \le \frac{1}{2}\, r^2 x_0^{-1/2}\, u_0^{1/2}\, K_{2.2}\Big(-\frac{1}{2}, \beta\Big)\, C_{2.2}\Big(\frac{3}{2}, \gamma_0\Big) \,. \tag{3.9}$$

Also, here we assume $\gamma_0 < 1$, or else the algorithm fails.

Next, we need to show that $\tilde{T}$ maps $B(0, \alpha)$ into itself. In order to do this, note that $\|T_4(g)\| \le (r^2/2)\, \|g\|$, which implies

$$\begin{aligned} \|\tilde{T}(0)\| &\le \frac{1}{2}\, r^2 \left\|(rz + x_0)^{-1/2}\right\| \left\|(u_0 + r\, u_1\, z)^{3/2}\right\| \\ &\le \frac{1}{2}\, r^2 x_0^{-1/2} u_0^{3/2}\, K_{2.2}\Big(-\frac{1}{2}, \beta\Big)\, K_{2.2}\left(\frac{3}{2}, \frac{r|u_1|}{u_0}\right) \,. \end{aligned}$$

Note that our assumption on γ_0 guarantees that the last term above is well-defined. Then, since

$$\left\|\tilde{T}(f)\right\| \le \left\|\tilde{T}(0)\right\| + \alpha \left\|\tilde{T}\right\|_{\text{Lip}}$$

we see that $\tilde{T}$ maps $B(0, \alpha)$ into itself provided

$$\frac{1}{2}\, r^2 x_0^{-1/2} u_0^{3/2}\, K_{2.2}\Big(-\frac{1}{2}, \beta\Big)\, K_{2.2}\left(\frac{3}{2}, \frac{r|u_1|}{u_0}\right) \le \alpha(1 - L)$$

whenever L is an upper bound for $\|\tilde{T}\|_{\text{Lip}}$. The algorithm also reports a failure if the upper bound L obtained using (3.9) is not strictly less than 1.

Note that if the previous conditions are satisfied, we also know that the solution u is strictly positive on $[x_0 - r, x_0 + r]$. Also, we know that it is defined as an analytic function on $|z - x_0| < r$.

Algorithm 3.2. *Given intervals x^*, u_0^* and u_1^*, and representable r, we construct a neighborhood $\mathcal{U}(I_0, \ldots, I_N; 0, C_g; 0)$ such that for any $x_0 \in x^*$, $u_0 \in u_0^*$, and $u_1 \in u_1^*$, and any solution u of (3.3) with any of these initial conditions, we have*

$$u(x) = u_0 + u_1 (x - x_0) + z^2 f(z), \qquad \text{with } z = \frac{x - x_0}{r},$$

for some $f \in \mathcal{U}$.

We can also make that the neighborhood to have the form

$$\mathcal{U}(I_0, \ldots, I_N; C_h, 0; \infty).$$

DESCRIPTION: First, we construct, in a heuristic way, a polynomial

$$p(z) = \sum_0^N p_i z^i$$

which approximately solves $\tilde{T}p = p$, and we set α such that $\|p\| \leq \alpha$. Next, we look for $\alpha_0 \geq \alpha$ such that the conditions on x_0, u_0, u_1, r and α_0 given by Algorithm 3.1 hold uniformly for all $x_0 \in x^*$, $u_0 \in u_0^*$ and $u_1 \in u_1^*$.

Next, since f is the fixed point of $\tilde{T}$, we have

$$\|p - f\| \leq \frac{\|p - \tilde{T}p\|}{1 - \|\tilde{T}\|_{\mathrm{Lip}}}.$$

Now, formulas (3.4) and (3.5) allow us to compute an upper bound for the numerator, Algorithm 3.1 allows us to compute a lower bound for the denominator, and we set C_g to be the resulting upper bound for the ratio. This immediately yields the required $\mathcal{U}$, by putting $I_i = [p_i, p_i]$ for $i = 0, \ldots, N$.

In order to obtain neighborhoods of type ∞, note that by power matching, for a given i, we can produce an interval I_i that contains any of the i'th Taylor coefficient for any of the solutions to the ODE

for all $x_0 \in x^*$, $u_0 \in u_0^*$ and $u_1 \in u_1^*$. Next, we pick any polynomial $p(z) = \sum_0^N p_i z^i$, with $p_i \in I_i$, and carry out the previous procedure, to obtain an upper bound C for $\|p - f\|$. It is clear then that $f \in \mathcal{U}(I_0, \ldots, I_N; C, 0; 0)$, since, if $f = \sum a_n z^n$, then

$$\sum_{n>N} |a_n| \leq \|f - p\| \leq C\,.$$

REMARK. Note that the previous algorithm enables us to construct a neighborhood of type 2 that contains u as a function of z.

Algorithm 3.3. *Given disjoint intervals x_0^* and x_1^*, and representable u_0 and u_1, we construct intervals y_0^* and y_1^* such that the solutions u to* (3.3) *with initial values u_0 and u_1 for $x \in x_0^*$ satisfy*

$$u(x') \in y_0^*\,, \qquad u'(x') \in y_1^*\,, \qquad \textit{for any } \; x' \in x_1^*\,.$$

DESCRIPTION: Choose a representable r such that $r \geq |x_0^* - x_1^*|$, (if we can't, we report a failure) and run the previous algorithm for this r. Then, y_0^* can be readily obtained by simply evaluating the neighborhood $\mathcal{U}$ produced by the algorithm at the interval x_1^*. In order to obtain y_1^*, we note that

$$u'(x') = u_1 + \int_x^{x'} u^{3/2}(s)\, s^{-1/2}\, ds$$

and this can be also easily computed. For a sharp bound, note that by the previous remark, we have $u(s)$ (as a function of $z = (s - x)/r) \in \mathcal{U}(I_0, \ldots, I_N; 0, C; 2)$, and thus, we also have

$$u(s)^{3/2}\, s^{-1/2} \text{ (as a function of } z) \in \mathcal{U}(I_0, \ldots, I_N; C_h, C_g; 2)$$

After integration, this reduces general error terms by a factor 3 compared to the ones that would follow from the weaker statement

$$u(s)^{3/2}\, s^{-1/2} \text{ (as a function of } z) \in \mathcal{U}(I_0, \ldots, I_N; C_h, C_g; 0)\,.$$

The following lemma has a trivial proof.

Lemma 3.4. *Say y_1 and y_2 are positive solutions of $y'' = x^{-1/2} y^{3/2}$ on the interval $[x_1, x_2]$, with $x_1 > 0$.*

1. *If* $y_1(x_1) \geq y_2(x_1)$ *and* $y_1'(x_1) \geq y_2'(x_1)$ *for all* $x \in [x_1, x_2]$, *then we have that* $y_1(x) \geq y_2(x)$ *and* $y_1'(x) \geq y_2'(x)$ *for all* $x \in [x_1, x_2]$.
2. *If* $y_1(x_2) \geq y_2(x_2)$ *and* $y_1'(x_2) \leq y_2'(x_2)$ *for all* $x \in [x_1, x_2]$, *then we have that* $y_1(x) \geq y_2(x)$ *and* $y_1'(x) \leq y_2'(x)$ *for all* $x \in [x_1, x_2]$.

Definition. *Let* $x_i^* = [x_i^{\mathrm{dn}}, x_i^{\mathrm{up}}]$ *for* $i = 1, 2$ *be two intervals. Then, we define*

$$x_1^* \cup_I x_2^* = [\min_{i=1,2} x_i^{\mathrm{dn}}, \max_{i=1,2} x_i^{\mathrm{up}}].$$

Algorithm 3.5. *Given disjoint intervals* x_0^* *and* x_1^*, *and intervals* u_0^* *and* u_1^*, *we construct intervals* y_0^* *and* y_1^* *such that all solutions* u *to* (3.3) *with initial values equal to any* $u_0 \in u_0^*$ *and any* $u_1 \in u_1^*$, *for any* $x \in x_0^*$ *are guaranteed to exist as positive solutions on* $[x, x']$, *and furthermore satisfy*

$$u(x') \in y_0^*, \qquad u'(x') \in y_1^*,$$

for all $x' \in x_1^*$.

DESCRIPTION: Assume first that $x_0^* < x_1^*$. Say $u_0^* = [u_0^{\mathrm{dn}}, u_0^{\mathrm{up}}]$, and $u_1^* = [u_1^{\mathrm{dn}}, u_1^{\mathrm{up}}]$. Next, run the previous algorithm: first, for $u_0 = u_0^{\mathrm{dn}}$ and $u_1 = u_1^{\mathrm{dn}}$, to obtain intervals w_0^* and w_1^*, and, second, for $u_0 = u_0^{\mathrm{up}}$ and $u_1 = u_1^{\mathrm{up}}$, two obtain intervals z_0^* and z_1^*. Note that if the first algorithm is successful, this implies that all solutions with initial values u_i^{up} and u_i^{dn} for any $x \in x^*$ are well-defined as strictly positive functions all the way up to x', and, by the previous lemma, all other solutions involved will be bounded above and away from zero: this implies that they can all be well-defined as positive functions all the way up to x'. We can then apply the previous lemma again to conclude that we can put

$$y_0^* = w_0^* \cup_I z_0^*, \qquad y_1^* = w_1^* \cup_I z_1^*.$$

If $x_0^* > x_1^*$, then we run the previous algorithm, first, for $u_0 = u_0^{\mathrm{dn}}$ and $u_1 = u_1^{\mathrm{up}}$, to obtain intervals w_0^* and w_1^*, and, second, for $u_0 = u_0^{\mathrm{up}}$ and $u_1 = u_1^{\mathrm{dn}}$, two obtain intervals z_0^* and z_1^*. It is then clear as before, that we can put

$$y_0^* = w_0^* \cup_I z_0^*, \qquad y_1^* = w_1^* \cup_I z_1^*.$$

The Initial Value Problem at 0.

Here we will be concerned with the solution to the Initial Value Problem

$$\begin{cases} u''(x) = x^{-1/2}\, u^{3/2}(x), \\ u(0) = 1, \\ u'(0) = -w, \end{cases} \tag{3.10}$$

for $w > 0$.

In this case, the solution to this problem will be in terms of a function $f \in H^1$ satisfying

$$u(x) = 1 - w\, r\, z^2 + z^3 f(z), \tag{3.11}$$

where $z = (x/r)^{1/2}$ and r is a small positive representable number.

The solution of (3.10) can be viewed as the fixed point of

$$T(u) = 1 + \int_0^x \left(-w + \int_0^t \frac{u^{3/2}(s)}{s^{1/2}}\, ds \right) dt,$$

and again T induces in a trivial way an operator $\tilde{T}$ of which f is its fixed point.

Algorithm 3.6. *We deduce conditions on w, r and α under which $\tilde{T}$ is a contraction in $B(0,\alpha)$, and we compute an upper bound for $\|\tilde{T}\|_{\mathrm{Lip}}$.*

DESCRIPTION: Let $g = \sum a_n z^n$. Consider the operators

$$\begin{aligned} T_1(f) &= -r\, w\, z^2 + z^3 f(z), \\ T_2(g) &= (1 + g(z))^{3/2}, \\ T_3(g) &= 4\, r^{3/2} \sum_{n \geq 0} \frac{a_n z^n}{(n+1)(n+3)}. \end{aligned} \tag{3.12}$$

It is clear that

$$T(u) = 1 - w\, x + z^3 (T_3 \circ T_2 \circ T_1)(f) \tag{3.13}$$

and thus, $\tilde{T} = T_3 \circ T_2 \circ T_1$.

Just as in Algorithm 3.1, T_1 is affine with an isometry as the linear part, $\|T_3\|_{\text{Lip}} \leq 4\, r^{3/2}/3$ and, for T_2, we have

$$\|T_2\|_{\text{Lip}} \leq C_{2.2}\Big(\frac{3}{2}, \gamma_0\Big)\,,$$

where, in this case

$$\sup_{\|f\| \leq \alpha} \|T_1(f)\| \leq r\, w + \alpha \stackrel{\text{def}}{=} \gamma_0\,.$$

We check that $\gamma_0 < 1$; otherwise, the algorithm fails.

Therefore,

$$\|\tilde{T}\|_{\text{Lip}} \leq \frac{4}{3}\, r^{3/2}\, C_{2.2}\Big(\frac{3}{2}, \gamma_0\Big)\,.$$

Then we check that the upper bound for $\|\tilde{T}\|_{\text{Lip}}$ thus obtained is strictly less than 1; otherwise, the algorithm fails.

Next, note that $\|T_3(g)\| \leq (4/3)\, r^{3/2}\, \|g\|$, which implies

$$\Big\|\tilde{T}(0)\Big\| \leq \frac{4}{3}\, r^{3/2} \Big\|(1 - w\, r\, z^2)^{3/2}\Big\| \leq \frac{4}{3}\, r^{3/2}\, K_{2.2}\Big(\frac{3}{2}, w\, r\Big)\,.$$

Then we see that $\tilde{T}$ maps $B(0, \alpha)$ into itself provided

$$\frac{4}{3}\, r^{3/2}\, K_{2.2}\Big(\frac{3}{2}, w\, r\Big) \leq \alpha\, \Big(1 - \|\tilde{T}\|_{\text{Lip}}\Big)\,.$$

Algorithm 3.7. *Given representable w and r we construct a neighborhood*

$$\mathcal{U}(I_0, \ldots, I_N; 0, C_g, 0)$$

such that the solution of (3.10) *is well-defined on* $[0, r]$ *and satisfies*

$$u(x) = 1 - wx + z^3 f(z)\,, \qquad \textit{with} \quad z = \Big(\frac{x}{r}\Big)^{1/2}\,,$$

for $f \in \mathcal{U}$.

DESCRIPTION: Similar to Algorithm 3.2.

Algorithm 3.8. *Given representable w and r, we construct intervals y_0^* and y_1^* such that the solution u of* (3.10) *satisfies*

$$u(r) \in y_0^*\,, \qquad u'(r) \in y_1^*\,.$$

DESCRIPTION: y_0^* can be obtained with a trivial variant of Algorithm 3.3, via Algorithm 3.7.

For y_1^*, note that, if we put

$$u^{3/2}(x) = (T_2 \circ T_1) f(z) = \sum_{n \geq 0} a_n z^n\,,$$

then

$$u'(r) = -w + \int_0^r \frac{u^{3/2}(x)}{x^{1/2}}\,dx = -w + r^{1/2} \sum_{n=0}^{\infty} \frac{2\,a_n}{n+1}\,.$$

Note now that in our representation $z = (x/r)^{1/2}$, we have a neighborhood of type 3 containing $u(x)$ as a function of z. We can thus construct another neighborhood of type 3 such that

$$\sum_{n=0}^{\infty} a_n z^n \in \mathcal{U}(I_0, \ldots, I_N; C_h, C_g; 3)\,.$$

Thus,

$$u'(r) \in -w + r^{1/2} \left(\sum_{n=0}^{N} \frac{2\,I_n}{n+1} \pm \varepsilon \right)\,,$$

whenever

$$|\varepsilon| \geq \frac{2\,C_h}{N+2} + \frac{1}{2}\,C_g\,.$$

Lemma 3.9. *Let u_1 and u_2 be the solutions of* (3.10), *with values w_1 and w_2, $w_1 < w_2$. Then, assuming that $u_{1,2}(x)$ are well-defined and strictly positive for $x \in [0,R]$, we have that $u_1(x) > u_2(x)$ and $u_1'(x) > u_2'(x)$ for $x \in [0,R]$.*

PROOF. Let f_1 and f_2 be associated with u_1 and u_2 as in (3.11). Since

$$u_1(x) \geq 1 - w_1\,x - z^3\,\|f_1\|\,,$$

and

$$u_2(x) \leq 1 - w_2\, x + z^3\, \|f_2\|\ ,$$

for all x small enough, we have that $u_1(x) > u_2(x)$ and thus $u_1''(x) > u_2''(x)$. Since the u_i'' are integrable at the origin, we conclude that

$$u_1'(x) = \int_0^x u_1''(t)\, dt - w_1 > \int_0^x u_2''(t)\, dt - w_2 = u_2'(x)$$

for all x small enough. The lemma now follows from Lemma 3.4.

Algorithm 3.10. *Given representable r and t, and an interval w^*, we construct intervals y_0^* and y_1^* such that any solution u of* (3.10) *for any $w \in w^*$ can be continued to $[0,t]$ and satisfies*

$$u(t) \in y_0^*\,, \qquad u'(t) \in y_1^*\,.$$

DESCRIPTION: Run Algorithm 3.8 twice, once for each endpoint of w^*, to obtain two pairs of intervals w_0^*, w_1^* and z_0^*, z_1^*. Lemma 3.9 then shows that all solutions of (3.10) with $w \in w^*$ are bounded above and away from 0, and can thus be extended as well-defined positive functions over $[0,t]$. Then, Lemma 3.9 again allows us to put

$$y_0^* = w_0^* \cup_I z_0^*\,, \qquad y_1^* = w_1^* \cup_I z_1^*\,.$$

The Initial Value Problem at Infinity.

Here we will be concerned with the solution to the Initial Value Problem

$$\text{(3.14)} \qquad \begin{cases} u''(x) = x^{-1/2}\, u^{3/2}(x)\,, \\ u(\infty) = 0\,, \\ \quad b_1 = b\,, \end{cases}$$

where the last condition is interpreted in the sense of (3.2).

The solution to this problem in this case will be expressed as

$$\text{(3.15)} \qquad u(x) = \frac{144}{x^3}\left(1 + b\, x^{-\alpha} + z^2 f(z)\right)\ ,$$

where $f \in H^1$, $z = R^\alpha x^{-\alpha}$, for some R large. In this case, the operators involved are not so obvious. Define

$$\begin{aligned} T_1(f) &= bR^{-\alpha}z + z^2 f(z)\,, \\ T_2(g) &= (1+g)^{3/2}\,, \\ T_3(g) &= 12\sum_{n=2}^{\infty} \frac{a_n z^{n-2}}{(n\alpha+3)(n\alpha+4)}\,, \end{aligned}$$

where, in the last formula, $g(z) = \sum_{n\geq 0} a_n z^n$. Then, put

$$\tilde{T} = T_3 \circ T_2 \circ T_1\,.$$

We now check that if f is a fixed point of $\tilde{T}$ in H^1, then u defined as in (3.15) solves (3.14). Note first that

$$\frac{u^{3/2}(x)}{x^{1/2}} = \frac{12\cdot 144}{x^5}\,(T_2 \circ T_1)(f)\,,$$

where

$$(T_2 \circ T_1)(f) = \sum_{n=0}^{\infty} a_n z^n\,, \qquad a_0 = 1\,, \quad a_1 = \frac{3}{2}\,bR^{-\alpha}\,.$$

Therefore, since u and its derivatives vanish at ∞,

$$\begin{aligned} u(x) &= \int_x^\infty \int_r^\infty \frac{u^{3/2}(t)}{t^{1/2}}\,dt\,dr \\ &= \frac{144}{x^3}\left(1 + \frac{12\,a_1}{(3+\alpha)(4+\alpha)}z + z^2\tilde{T}(f)\right)\,. \end{aligned}$$

Since $\alpha = (\sqrt{73}-7)/2$ satisfies the equation $(\alpha+3)(\alpha+4) = 18$, u satisfies (3.14).

The problem here is considerably more subtle than in the previous cases, due to the fact that T_3 does not scale with R. As a consequence, contraction properties of $\tilde{T}$ either hold or don't, and taking large R won't help much. We are lucky, however, that the norm of T_2 is essentially 3/2, and that the norm of T_3 is essentially

$$\frac{12}{(2\alpha+3)(2\alpha+4)} < \frac{1}{2}\,,$$

which says that the Lipschitz norm of $\tilde{T}$ will approximately be 3/4. We make this precise now.

Lemma 3.11. *Put $\beta = 0.3$. Assume that $|\bar{b}| = R^{-\alpha}|b| \leq 0.23$. Then $\tilde{T}$ is a contraction in $B(0,\beta)$, and $\|\tilde{T}\|_{\text{Lip}} \leq 0.8652$.*

PROOF (CALCULATOR-ASSISTED). Let f_1, $f_2 \in B(0,\beta)$, and put $\bar{f} = f_1 - f_2$.

$$(T_2 \circ T_1)(f) = 1 + \frac{3}{2}\left(\bar{b}\, z + z^2 f\right) + \frac{3}{8}\left(\bar{b}^2 z^2 + 2\,\bar{b}\, z^3 f + z^4 f^2\right) + \sum_{n\geq 3} \binom{3/2}{n} \left(\bar{b}\, z + z^2 f\right)^n .$$

So,

$$(T_2 \circ T_1)(f_2) - (T_2 \circ T_1)(f_2) = \frac{3}{2}\, z^2 \bar{f} + \frac{3}{4}\,\bar{b}\, z^3 \bar{f} + \frac{3}{8}\left(z^4 (f_1^2 - f_2^2)\right) + \sum_{n\geq 3} \binom{3/2}{n} \left(\left(\bar{b}\, z + z^2 f_1\right)^n - \left(\bar{b}\, z + z^2 f_2\right)^n\right) .$$

Now, since T_3 is linear, bounded, and the sum converges absolutely, we have

$$\tilde{T}(f_1) - \tilde{T}(f_2) = \frac{3}{2}\, T_3(z^2 \bar{f}) + \frac{3}{4}\,\bar{b}\, T_3(z^3 \bar{f}) + \frac{3}{8}\, T_3\left(z^4 (f_1^2 - f_2^2)\right) + \sum_{n\geq 3} \binom{3/2}{n} T_3\left(\left(\bar{b}\, z + z^2 f_1\right)^n - \left(\bar{b}\, z + z^2 f_2\right)^n\right) .$$

Note now that, for any $f \in H^1$, we have

$$\left\| T_3(z^k f) \right\| \leq \frac{12}{(k\alpha + 3)(k\alpha + 4)} \left\| z^k f \right\| ,$$

and that

$$\left(\bar{b}\, z + z^2 f_1\right)^n - \left(\bar{b}\, z + z^2 f_2\right)^n = z^{n+1} h(z) ,$$

$$\left\| \left(\bar{b}\, z + z^2 f_1\right)^n - \left(\bar{b}\, z + z^2 f_2\right)^n \right\| \leq n \left\| f_1 - f_2 \right\| \left(|\bar{b}| + \beta\right)^{n-1} .$$

Thus,

$$\|\tilde{T}\|_{\text{Lip}} \leq \frac{3}{2} \frac{12}{(2\alpha + 3)(2\alpha + 4)} + \frac{3|\bar{b}|}{4} \frac{12}{(3\alpha + 3)(3\alpha + 4)}$$

$$
\begin{aligned}
&+\frac{3}{8}\frac{12\cdot 2\cdot\beta}{(4\alpha+3)(4\alpha+4)}\\
&+\sum_{n\geq 3}\left|\binom{3/2}{n}\right|\frac{12n}{\big((n+1)\alpha+3\big)\big((n+1)\alpha+4\big)}(|\bar b|+\beta)^{(n-1)}\\
&\leq 0.72+.27\,|\bar b|+0.21\,\beta+X+Y\frac{(|\bar b|+\beta)^{20}}{1-|\bar b|-\beta}\\
&\leq 0.8652\,,
\end{aligned}
$$

where we have set

$$
X=\sum_{n=3}^{20}\left|\binom{3/2}{n}\right|\frac{12n}{\big((n+1)\alpha+3\big)\big((n+1)\alpha+4\big)}(|\bar b|+\beta)^{(n-1)}\,,
$$

and

$$
\begin{aligned}
Y &\stackrel{\text{def}}{=} \left|\binom{3/2}{21}\right|\frac{12\cdot 21}{(22\alpha+3)(22\alpha+4)}\\
&\geq \left|\binom{3/2}{n}\right|\frac{12n}{\big((n+1)\alpha+3\big)\big((n+1)\alpha+4\big)}
\end{aligned}
$$

for $n\geq 21$, and we have used

$$
X\leq 0.019\,,\qquad Y\frac{(|\bar b|+\beta)^{20}}{1-|\bar b|-\beta}\leq 9\cdot 10^{-10}\,.
$$

On the other hand,

$$
(T_2\circ T_1)(0)=\sum_{n\geq 0}\binom{3/2}{n}|\bar b|^n z^n\,.
$$

Thus,

$$
\begin{aligned}
\left\|\tilde T(0)\right\| &\leq \sum_{n\geq 2}\left|\binom{3/2}{n}\right|\frac{12\,|\bar b|^n}{(n\alpha+3)(n\alpha+4)}\\
&\leq \frac{3}{16}\,|\bar b|^2+0.0225\,|\bar b|^3+0.0066\,\frac{|\bar b|^4}{1-|\bar b|}\leq 0.01022\,.
\end{aligned}
$$

Therefore,

$$
\|\tilde T(f)\|\leq 0.01022+0.8652\,\beta\leq\beta\,,
$$

and $\tilde{T}$ maps $B(0,\beta)$ into itself.

Algorithm 3.12. *Given b^* (interval) and R (representable), we produce $\mathcal{U}_1$ such that, for any $b \in b^*$, the solution u of* (3.13) *is given by*

$$y(x) = \frac{144}{x^3}\left(1 + b\,x^{-\alpha} + z^2 f(z)\right), \qquad z = R^\alpha x^{-\alpha},$$

with $f \in \mathcal{U}_1$. Here, $\mathcal{U}_1$ depends only on b^*, i.e., it is independent of which particular b in b^* we are considering.

DESCRIPTION: We first check that we are in the hypothesis of Lemma 3.11. In this case, $\tilde{T}$ has a fixed point f, and, as we saw before, y defined as above satisfies the ODE.

In order to obtain bounds for f, we first look for a heuristic guess p: for example, we iterate $\tilde{T}$ (and truncate) a few times, starting with the function 0. Then, since computing rigorously $\tilde{T}p$ for all $b \in b^*$ poses no difficulty in view of Algorithm 2.3, we conclude that

$$\|f - p\| \le \frac{\|\tilde{T}p - p\|}{1 - 0.8652} \le 7.5\,\|\tilde{T}p - p\|, \qquad \text{all } b \in b^*.$$

Note that p is the same for all $b \in b^*$, but $\tilde{T}p$ still depends on b. However, the computation of $\sup_{b\in b^*} \|\tilde{T}p - p\|$ poses no problem, since it is less than or equal to $\|\tilde{T}p - p\|$ in the interval arithmetic sense.

The algorithm fails if the hypothesis of Lemma 3.11 are not met, or if $\|p\| > 0.3$.

Algorithm 3.13. *Given b and R, we produce two intervals u_0^* and u_1^* such that, if u is the solution to* (3.13), *we have*

$$u(R) \in u_0^*, \qquad u'(R) \in u_1^*.$$

DESCRIPTION: First, run Algorithm 3.12 for these values of b and R.

Again, it is easy to obtain u_0^*. Let f be related to u as in (3.15). Then, say

$$\left(1 + b\,x^{-\alpha} + z^2 f(z)\right)^{3/2} = \sum_{n\ge 0} a_n z^n \in \mathcal{U}(I_0, \ldots, I_N; C_h, C_g; 2).$$

Then,

$$\begin{aligned} u'(R) &= -\int_R^\infty \frac{144\cdot 12}{x^5} \sum a_n z^n \, dx \\ &= \frac{-144\cdot 12}{R^4} \sum \frac{a_n}{n\alpha+4} \\ &\in \frac{-144\cdot 12}{R^4} \left(\sum_{n=0}^{N} \frac{I_n}{4+n\alpha} \pm \varepsilon \right), \end{aligned}$$

with

$$|\varepsilon| \le \frac{C_h}{4+(N+1)\alpha} + \frac{C_g}{4+2\alpha}.$$

REMARK. Note that it is enough to run this algorithm for representable values of b, due to the monotonicity of the Thomas-Fermi equation (Lemma 3.4). We omit the trivial details, which are similar to those in Algorithm 3.5.

The Boundary Value Problem.

Next we discuss how to solve the Boundary Value Problem

$$\begin{cases} u''(x) = x^{-1/2} u^{3/2}(x), \\ u(0) = 1, \\ u(\infty) = 0, \end{cases}$$

We first describe how to obtain bounds for w_0.

Lemma 3.14. *Let u be the solution of* (3.3), *with $u_1 < 0$. If*

$$\frac{2u_0^{5/2}}{x_0^{1/2}} \le u_1^2$$

then, there exists a point $t > x_0$ such that u can be extended as a well-defined positive solution of the ODE *to $[x_0, t)$ and, furthermore,* $\inf_{x\in(x_0,t)} u(x) = 0$.

PROOF. Assume the lemma is false. It follows from general ODE considerations that, either u can be extended as a positive well-defined solution of the ODE, or else there exists a T such that $\sup_{x\in(x_0,T)} u(x) = \infty$. Let

$$d = \frac{|u_1|\, x_0^{1/2}}{u_0^{3/2}},$$

and note that in both of the two cases above u extends to a well-defined positive solution of the ODE to $(x_0, x_0 + d)$ and furthermore, $u \leq u_0$ on $[x_0, x_0 + d]$. Indeed, consider two cases:

a) u can be extended as a positive solution of the ODE all the way up to ∞. Then, if $u'(x) < 0$ it is trivial. Otherwise, let $x_1 > x_0$ be the first (and only) zero of u'; this means in particular that $u \leq u_0$ on $[x_0, x_1]$. Then, our claim follows by noting that

$$|u_1| \leq \sup_{(x_0,x_1)} u'' \cdot |x_0 - x_1| \leq u_0^{3/2} x_0^{-1/2} |x_0 - x_1|,$$

which implies $[x_0, x_0 + d] \subset [x_0, x_1]$.

b) u can be extended as a positive solution of the ODE all the way up to T, where it blows up. Since $u'(x_0) < 0$, there exists x_1, such that $x_0 < x_1 < T$ and $u'(x_1) = 0$. As before, x_1 is the first (and only) zero of u', $u \leq u_0$ on $[x_0, x_1]$, and $x_0 + d \leq x_1$.

Then, again, since $u'' \leq x_0^{-1/2} u_0^{3/2}$ on $[x_0, x_0 + d]$, we conclude that

$$u(x) \leq u_0 + u_1(x - x_0) + \frac{u_0^{3/2}}{2\, x_0^{1/2}} (x - x_0)^2, \qquad x \in [x_0, x_0 + d].$$

The lemma then follows by noting that this parabolic bound attains its minimum at exactly $x_0 + d$, and that this minimum is non-positive if the hypothesis in the statement of the lemma is satisfied.

Algorithm 3.15. *Given a representable w, we construct an algorithm that, if successful, will indicate whether $w < w_0$ or $w > w_0$.*

DESCRIPTION: By repeated applications of the previous algorithms, we can determine points x_i and intervals I_i, I_i', for $i = 0, \ldots, n$, for n large, such that the solution to the Thomas-Fermi equation with initial

values $u(0) = 1$, $u'(0) = -w$ satisfies $u(x_i) \in I_i$ and $u'(x_i) \in I_i'$. These algorithms also guarantee us that u does not vanish on $[0, x_n]$.

If, for some i, we have $I_i < I_{i+1}$, or $I_i' > 0$, this implies that for some $r_0 < x_n$, u is increasing and convex on $[r_0, r_0 + \varepsilon)$, and u will either not vanish at ∞, or blow up and cease to exist at a finite R_0. It is then clear by Lemma 3.9 that $w_0 > w$.

On the other hand, we know that if u becomes arbitrary small on $(0,t)$ for some t, then $w_0 < w$. Using the previous lemma, we then know that, if for some i, we have

$$\frac{2\, I_i^{5/2}}{x_i^{1/2}} \leq |I_i'|^2 \,,$$

then we have that $w_0 < w$.

If neither of the above happens, then we quit the algorithm without making any claims for bounds for w_0.

Algorithm 3.16. *Assuming bounds for w_0, and given $x_i \in \mathcal{R}$, we can produce y_i^* and $y_i'^* \in \mathcal{I}$, $i = 0, \ldots, m$, such that*

$$y(x_i) \in y_i^* \,, \qquad y'(x_i) \in y_i'^* \,, \qquad i = 0, \ldots, m \,.$$

DESCRIPTION: Apply Algorithm 3.10 for $r = x_0$, and then iterate Algorithm 3.5 for the x_i. This algorithm will fail if either Algorithm 3.10 or any of the runs of Algorithm 3.5 fails.

In order to ensure success for all algorithms, the choice of the x_i will in practice be rather delicate, as will be explained in Section 7.

Lemma 3.17. *Let u_1 and u_2 be the solutions of (3.14) with $b_1 = a_1$ and $b_1 = a_2$ respectively; then, if $a_1 \leq a_2$ and $u_1 > 0$ on $[M, \infty)$, then we have that $u_1(x) \leq u_2(x)$ and $u_1'(x) \geq u_2'(x)$ for all $x \in [M, \infty)$.*

PROOF. Obviously it is enough to assume $a_1 < a_2$. Let f_1 and f_2 be the functions associated with the u_i as in (3.15), with R common for the two of them, and large (perhaps a lot larger than M). Then,

$$u_1(x) \leq \frac{144}{x^3} \left(1 + z\big(a_1 R^{-\alpha} + z \,\|f_1\|\,\big) \right) ,$$

and

$$u_2(x) \geq \frac{144}{x^3}\left(1 + z\big(a_2 R^{-\alpha} - z\,\|f_2\|\,\big)\right).$$

Now, take R large so

$$a_1 R^{-\alpha} + \|f_1\| < a_2 R^{-\alpha} - \|f_2\|\,,$$
$$a_1 R^{-\alpha} + \|f_1\| < 1\,.$$

This ensures that $0 < u_1(x) < u_2(x)$ for $x > R$ and thus $u_1''(x) < u_2''(x)$ for all $x > R$. Now, note that

$$u_i'(x) = -\int_x^\infty u_i''(t)\,dt\,, \qquad \text{for } i = 1, 2\,,$$

which implies that, not only do we have $0 < u_1(x) < u_2(x)$ for $x > R$, but also $0 > u_1'(x) > u_2'(x)$ for all $x > R$. Finally, if R is larger than M, we apply Lemma 3.4 to guarantee that $0 < u_1(x) \leq u_2(x)$ and $u_1'(x) \geq u_2'(x)$ for $x \in [M, R]$ and thus for all $x \geq M$.

Algorithm 3.18. *Assuming bounds for b_1, we can produce $x_i \in \mathcal{R}$, and y_i^*, $y_i'^* \in \mathcal{I}$, $i = 1, \ldots, m$, such that*

$$y(x_i) \in y_i^*\,, \qquad y'(x_i) \in y_i'^* \qquad i = 1, \ldots, m\,.$$

DESCRIPTION: We choose the x_i in increasing order in i. We apply Algorithm 3.13 and Lemma 3.17 for $R = x_m$, and then iterate -going backwards- Algorithm 3.5 for the x_i. This algorithm will fail if either Algorithm 3.13 or any of the runs of Algorithm 3.5 fails.

REMARK. Strictly speaking, the choice of the x_i above is purely heuristic, and any choice yields a rigorous answer. In practice, most choices of x_i will yield as an answer "`failure`", which, although completely rigorous (after all, no theorem is claimed), is not very useful. As a result, it is important to make a good choice of the x_i. In practice, these x_i will be the same as the one used in Algorithm 3.16, whose choice is explained in Section 7.

Algorithm 3.19. *Given a representable b, and assuming bounds for w_0, we construct an algorithm that, if successful, will indicate whether $b < b_1$ or $b > b_1$.*

Also, assuming bounds for b_1, and given w, we indicate whether $w < w_0$ or $w > w_0$.

DESCRIPTION: Let y be the Thomas-Fermi function, and u be the solution of (3.13). Assuming bounds for w_0, Algorithm 3.16 allows us to produce representable x_i and intervals I_i and I_i', such that $y(x_i) \in I_i$ and $y'(x_i) \in I_i'$. For these x_i, using Algorithm 3.13 and repeated applications of Algorithm 3.5 (going backwards), we can produce intervals J_i and J_i' such that $u(x_i) \in J_i$ and $u'(x_i) \in J_i'$. In this situation we can again guarantee that $u > 0$. Then, if for some i

$$I_i > J_i \qquad \text{or} \qquad I_i' < J_i'$$

then we have $b < b_1$. If, however, we have

$$I_i < J_i \qquad \text{or} \qquad I_i' > J_i'$$

then we have $b > b_1$. We report a failure if

$$I_i \cap J_i \neq \emptyset\,, \qquad I_i' \cap J_i' \neq \emptyset\,,$$

for all i, in which case no relation is claimed between b and b_1.

The rest of the algorithm follows along the same lines.

Note that the last part of the previous algorithm constitutes a refinement of Algorithm 3.16, but it requires bounds for b_1. Also, Algorithm 3.15 allows us to obtain an initial, probably wasteful, bound for w_0. This initial bound allows us to obtain a bound for b_1, which in turn will allow us to improve our initial bound for w_0. Iterating this last algorithm in this way allows us to obtain improved bounds for both w_0 and b_1. The intersection of the bounds produced by Algorithms 3.16 and 3.18 are improved bounds for the Thomas-Fermi function and its derivative at points x_i. These translate immediately to better bounds for the solution of the Thomas-Fermi equation, and related constants.

Algorithm 3.20. *We can produce x_i, $r_i \in \mathcal{R}$, and*

$$\mathcal{U}_i(I_0^i, \ldots, I_N^i; C_{h,i}, C_{g,i}, 2)\,, \qquad i = 1, \ldots, m\,,$$

such that

$$y(x_i + z\, r_i) \in \mathcal{U}_i(I_0^i, \ldots, I_N^i; C_{h,i}, C_{g,i}, 2)\,, \qquad i = 1, \ldots, m\,,$$

and

$$\bigcup_{i=1}^{m}(x_i - r_i, x_i + r_i) = (x_1 - r_1, x_m + r_m) \subset (0, \infty).$$

DESCRIPTION: Our previous remark gives us the x_i, r_i, I_0^i and I_1^i. The rest follows by applying Algorithm 3.2 for every i.

Lemma 3.21. *The following inequalities hold:*

$$\begin{aligned} 1.588071022611278 &\leq w_0 \leq 1.588071022611471, \\ -13.270973847925352 &\geq b_1 \geq -13.270973848125353, \\ 0.486348538043594 &\leq \Omega_c^2 \leq 0.486348538046869, \\ 2.104025280219502 &\leq r_c \leq 2.104025280273837. \end{aligned}$$

Needless to say, the decimal numbers quoted above stand for the exact rational numbers they represent.

PROOF (COMPUTER-ASSISTED). The inequalities for w_0 and b_1 follow by carrying out previous algorithms. The inequality for r_c follows by checking that

$$u'(2.104025280219502) \geq 0 \geq u'(2.104025280273837).$$

The bounds for Ω_c are then trivial.

4. Zone I.

The purpose of this section is to prove (1.2) for all Ω in Zone I, as defined at the end of Section 1. We will do this as follows:

First, we partition Zone I into "fat" intervals $\{W_i\}_{i=1}^n$. Note that the first such interval will have the form $(0, \Omega_\varepsilon]$, for an Ω_ε to be picked (much) later in our proof. In fact, the role of the W_i will change as they approach zero: the larger ones (most of them, by the way) will receive identical treatment. Then, there will be a family of them, rather close to zero, which will receive a sort of special treatment, and then the single $W_1 = (0, \Omega_\varepsilon]$ which will be on its own.

Second, each fat interval W is divided into a finite partition of (lots of) suitably small subintervals Ω^* (except W_1 which will be both a "fat" and "thin" interval at the same time.) Our aim is to produce uniform

bounds for $-F''(\Omega)$ for all $\Omega \in \Omega^*$: for W_1 we will be able to produce only lower bounds, since $-F''$ is unbounded there; for the others, we will be able to produce both upper and lower bounds.

Say $\Omega^* = [z_1, z_2]$ is contained in the fat interval $W = [w_1, w_2]$.

We construct two functions $a(\Omega^*)$ and $b(\Omega^*)$, constant on each subinterval Ω^*, such that

$$r_1(\Omega) < a < b < r_2(\Omega), \qquad \Omega \in \Omega^*.$$

In practice, a and b will be very close to r_1 and r_2 respectively.

Now, we recall Lemma 1.2; our job is then to compute each of the following

$$I_1 = \int_a^b \left(u(r) - \Omega^2\right)^{-3/2} y(r)\, dr, \tag{4.1a}$$

$$I_2 = \lim_{\delta \to 0} \left(\int_{r_1(\Omega)+\delta}^{a} \left(u(r) - \Omega^2\right)^{-3/2} y(r)\, dr - G_1(\Omega)\, \delta^{-1/2} \right), \tag{4.1b}$$

$$I_3 = \lim_{\delta \to 0} \left(\int_{b}^{r_2(\Omega)-\delta} \left(u(r) - \Omega^2\right)^{-3/2} y(r)\, dr - G_2(\Omega)\, \delta^{-1/2} \right), \tag{4.1c}$$

with G_i such that the limit is finite.

The computation of I_1 is done as follows: Break up

$$I_1 = \sum_{i=1}^{n} \int_{t_i}^{t_{i+1}} \left(u(r) - \Omega^2\right)^{-3/2} y(r)\, dr = \sum_{i=1}^{n} J_i(\Omega),$$

where $t_1 = a$ and $t_{n+1} = b$.

Note that each J_i can be computed directly, since it involves only elementary operations. However, computing *all* J_i like that will take a very long time. To remedy this, we do as follows:

First, we take two numbers $\tilde{a}(\Omega^*)$ and $\tilde{b}(\Omega^*)$, constant on each subinterval Ω^*, such that

$$\tilde{a} = t_{i_0}, \qquad \tilde{b} = t_{i_1},$$

with $1 \le i_0$ and $i_1 \le n$. Normally, we will have that $i_0 \le i_1$. It could happen, however, that $i_0 > i_1$ meaning that the computation of the J_i is always done directly, without using the faster method below.

Then, we take t_i for $i = i_0, \dots, i_1$ to be the same for all $\Omega^* = [z_1, z_2] \subset W = [w_1, w_2]$, and we compute once and for all the following numbers:

$$a_{k,i} = \int_{t_i}^{t_{i+1}} \left(u(r) - w_k^2\right)^{-3/2} y(r)\, dr\,,$$

$$b_{k,i} = 3\, w_k \int_{t_i}^{t_{i+1}} \left(u(r) - w_k^2\right)^{-5/2} y(r)\, dr\,,$$

for $k = 1, 2$, and $i = i_0, \dots, i_1$.

Next, note that the functions

$$f_i(w) = \int_{t_i}^{t_{i+1}} \left(u(r) - w^2\right)^{-3/2} y(r)\, dr$$

are increasing and convex on W. Therefore, if $w \in [w_1, w_2]$,

$$\max_{k=1,2} \left(f_i'(w_k)(w - w_k) + f_i(w_k)\right) \le f_i(w)\,,$$

$$f_i(w) \le \frac{f_i(w_1) - f_i(w_2)}{w_1 - w_2}(w - w_1) + f_i(w_1)\,.$$

Thus,

$$\max_{k=1,2} \left(b_{k,i}(\Omega - w_k) + a_{k,i}\right) \le J_i(\Omega)\,,$$

$$J_i(\Omega) \le \frac{a_1 - a_2}{w_1 - w_2}(\Omega - w_1) + a_1\,, \qquad \Omega \in W\,.$$

This gives us intervals $\tilde{J}_i(\Omega^*)$, such that

$$J_i(\Omega) \subset \tilde{J}_i(\Omega^*)\,, \qquad i_0 \le i \le i_1\,, \qquad \Omega \in \Omega^*\,. \tag{4.2}$$

In practice, $\tilde{a}$ and $\tilde{b}$ will be far from r_1 and r_2. They will enclose a region which is safely away from the singularities of the integrand in our formula for F'', for which we can expect (4.2) to be sharp.

For i outside of the range $[i_0, i_1]$, we compute $f_i(z_1)$ and $f_i(z_2)$ directly, and, by our previous remark,

$$J_i(\Omega) \in \tilde{J}_i(\Omega^*) \stackrel{\text{def}}{=} [f_i(z_1), f_i(z_2)]\,.$$

Thus, we have defined $\tilde{J}(\Omega^*)$ for all $i = 1, \ldots, n$, and we conclude that

$$I_1(\Omega) \in \sum_{i=1}^{n} \tilde{J}_i(\Omega^*), \qquad \Omega \in \Omega^*.$$

Computation of $\mathbf{I_2}$.

Consider a small number $\bar{\Omega}_2 \ll \Omega_c$, that we can make coincide with one of the endpoints of the fat intervals W_i. We distinguish two cases: $\Omega > \bar{\Omega}_2$ and $\Omega \leq \bar{\Omega}_2$.

If $\Omega > \bar{\Omega}_2$, we use Algorithm 3.2 to compute $\mathcal{U}_1$ such that

$$u(x) = \Omega^2 + z f(z), \qquad z = \frac{x - r_1(\Omega)}{r}, \qquad f \in \mathcal{U}_1,$$

where $r \geq |a - r_1(\Omega)|$ and $\mathcal{U}_1$ is uniform for all $\Omega \in \Omega^*$. Note that $f(0) > 0$. Also, in order to apply Algorithm 3.2, we need to obtain bounds for $r_1(\Omega)$ and for $u'(r_1)$; the first can be done by obtaining heuristic bounds r_{dn} and r_{up}, and checking that $u(r_{\mathrm{dn}}) \leq \Omega^2$ and $u(r_{\mathrm{up}}) \geq \Omega^2$, which can be easily checked using the information given by Algorithm 3.20. Bounds for $u'(r_1)$ can be obtained using the bounds for r_1 and the information on y_{TF} (hence on u) given by Algorithm 3.20. See the section on implementation for more details. Therefore,

$$\begin{aligned}\int_{r_1(\Omega)+\delta}^{a} \left(u(x) - \Omega^2\right)^{-3/2} y(x)\, dx &= \int_{r_1(\Omega)+\delta}^{a} z^{-3/2} f^{-3/2}(z)\, y(x)\, dx \\ &= \int_{r_1(\Omega)+\delta}^{a} z^{-3/2} \tilde{f}(z)\, dx,\end{aligned}$$

for a new function $\tilde{f}(z) = y(x) f^{-3/2}(z)$, that can also be enclosed in a computable $\mathcal{U}_2$. Note that $\tilde{f}(0) > 0$ also. Thus, if

$$\tilde{f}(z) = \sum_{n \geq 0} a_n z^n \in \mathcal{U}(J_0, \ldots, J_N; C_h, C_g; 1),$$

we see that

$$\begin{aligned}\int_{r_1(\Omega)+\delta}^{a} \left(u(x) - \Omega^2\right)^{-3/2} y(x)\, dx &= \int_{r_1(\Omega)+\delta}^{a} \sum_{n \geq 0} a_n z^{n-3/2}\, dx \\ &= r \sum_{n \geq 0} \frac{a_n}{n - 1/2} z^{n-1/2} \Bigg|_{z=\delta/r}^{z=(a-r_1(\Omega))/r}.\end{aligned}$$

This implies that

$$I_2 = r \sum_{n \geq 0} \frac{a_n}{n - 1/2} \left(\frac{a - r_1(\Omega)}{r} \right)^{n - 1/2}$$
$$\in r \sum_{n=0}^{N} \frac{J_n}{n - 1/2} \left(\frac{a - r_1(\Omega)}{r} \right)^{n - 1/2} \pm \varepsilon ,$$

with

$$|\varepsilon| \leq r \left(\frac{C_h}{N + 1/2} + 2\, C_g \right) .$$

When $\Omega \leq \bar{\Omega}_2$, we proceed as follows:

Consider the change of variables given by $r(t)$, the inverse of u. Then, by the last remark in Lemma 1.2,

$$I_2 = \frac{d}{d\Omega} \left(\Omega \int_{r_1(\Omega)}^{a} \left(u(r) - \Omega^2 \right)^{-1/2} \frac{dr}{r} \right)$$
$$= \frac{d}{d\Omega} \left(\Omega \int_{\Omega^2}^{u(a)} \left(t - \Omega^2 \right)^{-1/2} w(t)\, dt \right) ,$$

for

$$w(t) = \frac{r'(t)}{r(t)} .$$

In order to compute w, we consider the following: let x_0 and ε' be small numbers satisfying

a) $u'(x) > 1 - \varepsilon'$ for $x \in [0, x_0]$.

b) For a sequence $\{\bar{b}_n\} \in l^1$, we have

$$u(x) = x \left(1 + \sum_{n=2}^{\infty} \bar{b}_n\, \bar{x}^{n/2} \right) , \qquad x \leq x_0 .$$

Furthermore, we know that

$$1 + \sum_{n=2}^{\infty} \bar{b}_n\, z^n \in \mathcal{U}(I_0, \ldots, I_m; 0, C_g; 2) , \tag{4.3}$$

with $I_0 = [1,1]$ and $I_1 = [0,0]$. Here, $\bar{x}$ denotes x/x_0, and $\bar{b}_n = b_n\, x_0^{n/2}$.

Define

$$\bar{t} = (t/x_0)^{1/2}\,.$$

We also consider a small number $\eta \leq u(x_0)$. It will be chosen so that (4.10) below holds. We start by obtaining expressions for $u'(r)$ and $u''(r)$ similar to the one for u in (4.3). Note first that (4.3) is equivalent to an expression for $y(x)$. Then, by the Thomas-Fermi equation, and by integration, we have

$$y''(x) = x^{-1/2}\left(\sum_{n=0}^{\infty} \bar{b}_n\, \bar{x}^{\,n/2}\right)^{3/2} = x^{-1/2}\sum_{n=0}^{\infty} y''_n\, \bar{x}^{\,n/2} = x^{-1/2}\, y_{pp}(\bar{x}^{1/2})\,,$$

with $y''_0 = 1$, and

$$y'(x) = \sum_{n=0}^{\infty} y'_n\, \bar{x}^{\,n/2} = y_p(\bar{x}^{1/2})\,, \qquad y'_n = \begin{cases} -w_0\,, & \text{if } n = 0\,, \\ \dfrac{2}{n}\, x_0^{1/2}\, y''_{n-1}\,, & \text{if } n > 0\,, \end{cases}$$

from which we obtain

$$u'(x) = x\, y'(x) + y(x) = \sum_{n=0}^{\infty} u'_n\, \bar{x}^{\,n/2} = u_p(\bar{x}^{1/2})\,,$$

(4.4.a)
$$u'_n = \begin{cases} 1\,, & \text{if } n = 0\,, \\ 0\,, & \text{if } n = 1\,, \\ \bar{b}_n + y'_{n-2}\, x_0\,, & \text{if } n \geq 2\,, \end{cases}$$

$$u''(x) = x\, y''(x) + 2\, y'(x) = u_{pp}(\bar{x}^{1/2}) = \sum_{n=0}^{\infty} u''_n\, \bar{x}^{\,n/2}\,,$$

(4.4.b)
$$u''_n = \begin{cases} -2\, w_0\,, & \text{if } n = 0\,, \\ 2\, y'_n + x_0^{1/2}\, y''_{n-1}\,, & \text{if } n > 0\,. \end{cases}$$

Note that since we know a neighborhood of type 2 that contain y, we can enclose y_{pp} and u_p also in neighborhoods of type 2, and y_p and u_{pp} in neighborhoods of type 3.

We start our analysis understanding $r(t)$. First, a technical algorithm.

Algorithm 4.1. *Given a function* $r(t) = t\,R(\bar{t})$ *with*

$$R(z) \in \mathcal{U}^0(a_0^*, \ldots, a_N^*; C_N, 0; \infty; S)\,, \qquad a_0^* = [1,1]\,,$$

valid for $\bar{t} \in S \subset [0,1]$, *and given a neighborhood* $\mathcal{U}^1$ *in* H^1, *we can compute another neighborhood* $\mathcal{U}_2^0$, *also of type* ∞ *in* C^0, *also valid on* $\bar{t} \in S$, *such that if*

$$f(t) = \sum_{n=0}^{\infty} \bar{c}_n \bar{t}^{\,n} \in \mathcal{U}^1(I_0, \ldots, I_N; C_h, C_g, m)\,,$$

then $G(\bar{t}) \in \mathcal{U}_2^0$ *for*

$$G(\bar{t}) = f\big(r(t)\big) = \sum_{n=0}^{\infty} \bar{c}_n \left(\frac{r(t)}{x_0}\right)^{n/2}\,.$$

We assume that $0 < r(t) \leq x_0$ *for* $\bar{t} \in S$.

DESCRIPTION: Consider any $a_i \in a_i^*$. Put

$$\left(\sum_{n=0}^{N} a_n \bar{t}^{\,n} + \bar{t}^{\,N+1} h(\bar{t})\right)^{\gamma} = \sum_{n=0}^{n_0} a_{n,\gamma}\, \bar{t}^{\,n} + \bar{t}^{\,n_0+1} h(\bar{t}; \gamma, n_0 + 1)\,,$$

with $a_{n,\gamma} \in a_{n,\gamma}^*$, the $a_{n,\gamma}^*$ easily determined intervals,

$$\|h(\bar{t}; \gamma, n_0)\|_{C^0} \leq \varepsilon_{\gamma, n_0}\,.$$

We can see, on one hand, that

$$\begin{aligned}\sum_{n=0}^{N} \bar{c}_n \left(\frac{r(t)}{x_0}\right)^{n/2} &= \sum_{n=0}^{N} \bar{c}_n\, \bar{t}^{\,n} \left(\sum_{k=0}^{N-n} a_{k,n/2}\, \bar{t}^{\,k} + \bar{t}^{\,N+1-n} h\big(\bar{t}; \frac{k}{2}, N-n\big)\right) \\ &= \sum_{n=0}^{N} d_n\, \bar{t}^{\,n} + \bar{t}^{\,N+1} h(t)\,,\end{aligned}$$

with

$$\begin{aligned}|h(t)| &\leq \sum_{n=0}^{N} |\bar{c}_n|\, \varepsilon_{n/2, N-n+1} \\ &\leq \sum_{n=0}^{N} |I_n|\, \varepsilon_{n/2, N-n+1} + C_g \max_{n=m,\ldots,N} \varepsilon_{n/2, N-n+1}\,,\end{aligned}$$

and

$$d_n = \sum_{i+j=n} \bar{c}_i \, a_{j,i/2}$$
$$\in \sum_{i+j=n} I_i \, a^*_{j,i/2} \pm \varepsilon_n \stackrel{\text{def}}{=} d^*_n ,$$

where

$$\varepsilon_n \leq \begin{cases} C_g \sup\limits_{m \leq i \leq n} |a_{n-i,i/2}| , & \text{if } n \geq m , \\ \\ 0 , & \text{otherwise.} \end{cases}$$

On the other hand, since

$$\sum_{n=N+1}^{\infty} |\bar{c}_n| \left(\frac{r(t)}{x_0} \right)^{n/2} \leq (C_g + C_h) \left(\frac{r(t)}{x_0} \right)^{(N+1)/2}$$
$$\leq (C_g + C_h) \, \bar{t}^{\,N+1} \left(1 + \sum_{n=1}^{N} |a_n| + C_N \right)^{(N+1)/2} ,$$

we conclude that

$$f(r(t)) = G(\bar{t}) \in \mathcal{U}^0(d^*_0, \ldots, d^*_N; \tilde{C}_h, 0; \infty) ,$$

with

$$\tilde{C}_h = (C_g + C_h) \left(1 + \sum_{n=1}^{N} |a_n| + C_N \right)^{(N+1)/2}$$
$$+ \sum_{n=0}^{N} |I_n| \, \varepsilon_{n/2, N-n+1} + C_g \max_{n=m,\ldots,N} \varepsilon_{n/2, N-n+1} .$$

Algorithm 4.2. *Given $N \geq 0$, we produce intervals $a^*_2, \ldots, a^*_N$, and a constant C_N, such that*

$$\left| r(t) - t\Big(1 + \sum_{n=2}^{N} a_n \bar{t}^{\,n}\Big) \right| \leq C_N \, t \, \bar{t}^{\,N+1} ,$$

for constants $a_i \in a^*_i$, $i = 2, \ldots, N$, and for $t \leq \eta$.

DESCRIPTION: First, we will construct an inductive procedure to define numbers a_n such that

$$u(r_N(t)) = t\left(1 + O(\bar{t}^{N+1})\right) \qquad \text{as } t \to 0\,, \tag{4.5}$$

where

$$r_N(t) = t\Big(1 + \sum_{n=2}^{N} a_n \bar{t}^{\,n}\Big)\,.$$

By induction. For $N = 1$, let $r_0(t) = t$. Note that $r(t) \geq r_0(t)$, that $r_0(t) \leq x_0$ for $\bar{t} \leq 1$, and that, if $t \leq \eta \leq u(x_0)$ then $\bar{t} \leq 1$.

Therefore, we have

$$u(r_0(t)) = t\Big(1 + \sum_{n\geq 2} b_n t^{n/2}\Big)\,.$$

Thus,

$$|r_0(t) - r(t)| \leq t\,\frac{\left|\sum_{n\geq 2} b_n t^{n/2}\right|}{\inf_{r\in[r_0(t),\,r(t)]} |u'(r)|}\,. \tag{4.6}$$

Since, for t small enough, $r(t) \leq x_0$, the denominator is bounded below by $(1 - \varepsilon')$, and we conclude

$$r_0(t) - r(t) = O(t^2)\,.$$

For general N, we set

$$r_N(t) = t\Big(1 + \sum_{n=2}^{N} a_n \bar{t}^{\,n}\Big)\,,$$

where $a_2, \ldots, a_{N-1}$ satisfy the induction hypothesis.

Note that we have

$$\sum_{n=1}^{\infty} b_n\, r_N(t)^{n/2} - \sum_{n=1}^{\infty} b_n\, r_{N-1}(t)^{n/2} = O(\bar{t}^{\,N+1})\,. \tag{4.7}$$

Thus, if for any real number γ we put

$$\left(1 + \sum_{n=2}^{N-1} a_n \bar{t}^{\,n}\right)^{\gamma} = \sum_{n=0}^{N} a_{n,\gamma}\, \bar{t}^{\,n} + O(\bar{t}^{\,N+1})\,, \tag{4.8}$$

we see that

$$u(r_{N-1}(t)) = t\left(1+\sum_{n=2}^{N-1} a_n \bar{t}^{\,n}\right)$$

$$\cdot\left(\sum_{n=0}^{N} \bar{b}_n \bar{t}^{\,n}\left(\sum_{i=0}^{N} a_{i,n/2}\bar{t}^{\,i} + O(\bar{t}^{\,N+1})\right) + O(\bar{t}^{\,N+1})\right) \tag{4.9}$$

$$= t\left(1+\sum_{n=2}^{N-1} a_n \bar{t}^{\,n}\right)\left(\sum_{n=0}^{N} c_n \bar{t}^{\,n} + O(\bar{t}^{\,N+1})\right),$$

where

$$c_k = \sum_{n=0}^{k} \bar{b}_n\, a_{k-n,n/2}\,, \qquad c_0 = 1\,.$$

By the induction hypothesis, (4.9) is equal to $t(1+O(\bar{t}^{\,N}))$. Thus, using (4.7), we can see that

$$u(r_N(t)) = t\left(1+\sum_{n=2}^{N-1} a_n \bar{t}^{\,n} + a_N \bar{t}^{\,N}\right)\left(\sum_{n=0}^{N} c_n \bar{t}^{\,n} + O(\bar{t}^{\,N+1})\right),$$

where the c_n here are the same as those in (4.9).

Therefore, by putting

$$a_N = -\sum_{k=1}^{N-2} a_{N-k}\, c_k - c_N$$

we get rid of all $\bar{t}^{\,N}$ terms, thus obtaining (4.5).

So far, we have proved the existence of numbers a_n such that (4.5) is satisfied.

This procedure also gives us an algorithm to compute the a_n^*. Indeed, bounds $a_{n,\gamma}^*$ for the $a_{n,\gamma}$ can be computed explicitly, since by the induction hypothesis we already know a_i^*, for $i = 1, \ldots, N-1$. As for the c_k, recalling (4.3),

$$c_k = \sum_{n=0}^{k} \bar{b}_n\, a_{k-n,n/2} \in \sum_{n=0}^{k} I_n\, a_{k-n,n/2}^* \pm \varepsilon_k\,,$$

with

$$|\varepsilon_k| \le \begin{cases} C_g \sup\limits_{k\ge n\ge 2} |a_{k-n,n/2}|\,, & \text{if } k \ge 2\,, \\ 0\,, & \text{if } k \le 1\,. \end{cases}$$

In particular, it follows immediately that $a_2 = -\bar{b}_2 = x_0\, w_0 \in -I_2$.

To obtain a good value for the constant C_N, we proceed as follows: First, check that

$$\eta\Big(1 + \sum_{n=2}^{N} |a_n|\Big) \leq x_0\,. \tag{4.9.a}$$

(See also (4.10) below.) This allows us to invoke Algorithm 4.1, with $S = [0, \sqrt{\eta/x_0}]$, to obtain

$$y(r_N(t)) = f(\bar{t}) \in \mathcal{U}_1^0(\cdots;\cdots;\infty)$$

and thus we can write

$$u\,(r_N(t)) = r_N(t)\, f(\bar{t}) \equiv t\, g(\bar{t})\,,$$

with g belonging to $\mathcal{U}^0(I_0, \ldots, I_N; \tilde{\tilde{C}}_h, 0; \infty; t \leq \eta)$, the product neighborhood of

$$\mathcal{U}_2(a_0^*, \ldots, a_N^*; 0, 0; \infty; t \leq \eta)$$

and $\mathcal{U}_1^0$. Now, since (4.5) implies that $g(\bar{t}) = 1 + O(\bar{t}^{\,N+1})$, we can take $I_0 = [1,1]$ and $I_i = [0,0]$, for $i = 1, \ldots, N$. Note that this is relied crucially on the fact that $\mathcal{U}^0$ is of type ∞; in fact, since

$$g(\bar{t}) \in \mathcal{U}^0(I_0, \ldots, I_N; \tilde{\tilde{C}}_h, 0; \infty; t \leq \eta)\,,$$

we can find constants $p_i \in I_i$ such that

$$\Big| g(\bar{t}) - \sum_{k=0}^{N} p_i \bar{t}^{\,k} \Big| \leq \tilde{\tilde{C}}_h\, \bar{t}^{\,N+1}\,.$$

On the other hand, since $g(\bar{t}) = 1 + O(\bar{t}^{\,N+1})$, we must have $p_0 = 1$ and $p_i = 0$ for $i = 1, \ldots, N$, thus

$$|g(\bar{t}) - 1| \leq \tilde{\tilde{C}}_h\, \bar{t}^{\,N+1}\,.$$

If Algorithm 4.1 had produced a neighborhood of any other type, it would have been harder to conclude this without changing $\tilde{\tilde{C}}_h$. More precisely, if we have $\phi(t)$ defined on $t \in [-1,1]$ such that

$$\Big| \phi(t) - \sum_{n=0}^{N} a_n t^n \Big| \leq C\, t^k \quad \text{and} \quad \phi(t) - \sum_{n=0}^{N} a_n t^n = O(t^{N+1})$$

for $t \in [-1,1]$, we cannot conclude

$$\left|\phi(t) - \sum_{n=0}^{N} a_n t^n\right| \leq C\, t^{N+1}\,,$$

with the same constant C unless $k > N$. Counterexamples with $k \leq N$ are readily available (simply take $N = k = 0$, $\phi(t) = t^3 - t$, $a_0 = 0$; then, $|\phi| \leq 2^{-1/2}$ but $\phi(t)$ is not bounded by $2^{-1/2}|t|$.)

So, we have that

$$\left|u\big(r_N(t)\big) - t\right| \leq \tilde{\tilde{C}}_h\, t\, \bar{t}^{\,N+1}\,, \qquad t \leq \eta\,.$$

Next, note that

$$|r_N(t) - r(t)| \leq \frac{|u(r_N(t)) - t|}{\inf_{0 \leq r \leq \max\{r(t), r_N(t)\}} |u'(r)|}\,.$$

Then, the fact that $\eta \leq u(x_0)$ and (4.9.a) imply that

$$|u'(r)| \geq (1 - \varepsilon')\,, \qquad 0 \leq r \leq \max\{r(t), r_N(t)\}\,, \quad t \leq \eta\,,$$

from which the lemma follows by taking

$$C_N = \frac{\tilde{\tilde{C}}_h}{1 - \varepsilon'}\,.$$

Algorithm 4.3. *We produce a neighborhood* $\mathcal{U}^0(I_0, \ldots, I_N; C_h, 0; \infty)$ *such that*

$$h(t) \stackrel{\text{def}}{=} t\, w'(t) + w(t) = f(\bar{t})\,, \qquad f \in \mathcal{U}\,,$$

for $t \leq \eta$.

DESCRIPTION: We note that

$$r'(t) = \frac{1}{u'\big(r(t)\big)}\,, \qquad r''(t) = -r'(t)^3 u''\big(r(t)\big)\,.$$

Therefore, we check that

$$\eta\Big(1 + \sum_{n=1}^{N} |a_n| + C_N\Big) \leq x_0 \tag{4.10}$$

and as a result of this, we can apply Algorithm 4.1 and obtain neighborhoods of type ∞ in C^0 containing functions f, f_p and f_{pp} such that

$$r(t) = tf(\bar{t}), \qquad r'(t) = f_p(\bar{t}), \qquad r''(t) = f_{pp}(\bar{t}),$$

which are valid for $t \leq \eta$. These functions are obtained by putting

$$f_p(\bar{t}) = \frac{1}{u_p\big((r(t)/x_0)^{1/2}\big)}, \qquad f_{pp}(\bar{t}) = -f_p^3\, u_{pp}\big((r(t)/x_0)^{1/2}\big).$$

Note that with this definition, f and f_p are normalized to be 1 at 0, and $f_{pp}(0) = 2w_0$. Thus,

$$w(t) = \frac{r'(t)}{r(t)} = \frac{1}{t}\,\frac{f_p(\bar{t})}{f(\bar{t})},$$
$$w'(t) = \frac{r''(t)}{r(t)} - \left(\frac{r'(t)}{r(t)}\right)^2 = \frac{1}{t}\,\frac{f_{pp}(\bar{t})}{f(\bar{t})} - \frac{1}{t^2}\left(\frac{f_p(\bar{t})}{f(\bar{t})}\right)^2,$$

and

$$\begin{aligned} h(t) &= t\,w'(t) + w(t) \\ &= t^{-1}\left(\frac{f_p(\bar{t})}{f(\bar{t})} - \left(\frac{f_p(\bar{t})}{f(\bar{t})}\right)^2\right) + \frac{f_{pp}(\bar{t})}{f(\bar{t})} = f_h(\bar{t}), \end{aligned}$$

for a function f_h belonging to an easily computable neighborhood of type ∞ in C^0.

Note that by our normalization, the t^{-1} terms drop out. Furthermore, there are no $t^{-1}\bar{t}$ terms since neither f nor f_p have $\bar{t}$ terms, and in fact $f_h(0) = w_0$, which we can easily see as follows: first, $f(\bar{t}) = 1 + w_0\,t + O(\bar{t}^{\,3})$, $f_p(\bar{t}) = 1 + 2\,w_0\,t^2 + O(\bar{t}^{\,3})$ and $f_{pp}(\bar{t}) = 2\,w_0 + O(\bar{t})$, therefore

$$t^{-1}\left(\frac{f_p(\bar{t})}{f(\bar{t})} - \left(\frac{f_p(\bar{t})}{f(\bar{t})}\right)^2\right) = -w_0 + O(t^{-1}\bar{t}^{\,3}) = -w_0 + O(\bar{t}),$$

and

$$\frac{f_{pp}(\bar{t})}{f(\bar{t})} = 2\,w_0 + O(\bar{t}),$$

thus $f_h(0) = w_0$. Moreover, if we set

$$\frac{f_p(\bar{t})}{f(\bar{t})} - \left(\frac{f_p(\bar{t})}{f(\bar{t})}\right)^2 + t\,\frac{f_{pp}(\bar{t})}{f(\bar{t})} \in \mathcal{U}^0(I_0, \ldots, I_{N+2}; \varepsilon_h, 0; \infty),$$

then

$$f_h \in \mathcal{U}^0(x_0^{-1} I_2, \ldots, x_0^{-1} I_{N+2}; x_0^{-1}\varepsilon_h, 0; \infty; t \leq \eta),$$

valid for $t \leq \eta$.

Now, let

$$f_h(\bar{t}) = \sum_{n=0}^{N} a_n \bar{t}^n + H(\bar{t}),$$

with

$$|H(\bar{t})| \leq \varepsilon_h |\bar{t}|^{N+1}, \qquad t \leq \eta.$$

Finally, then, let δ be a small number such that $u(\delta) \leq \eta$, set $\bar{\Omega}_2 \leq \sqrt{u(\delta)}$, and consider $\Omega \leq \bar{\Omega}_2$ for which we set $a(\Omega) \equiv \delta$:

$$\begin{aligned}
&\frac{d}{d\Omega}\left(\Omega \int_{r_1(\Omega)}^{\delta} \left(u(r) - \Omega^2\right)^{-1/2} \frac{dr}{r}\right) \\
&\qquad = \frac{d}{d\Omega}\left(\Omega^2 \int_1^{\Omega^{-2}u(\delta)} (t-1)^{-1/2}\, w(t\,\Omega^2)\, dt\right) \\
&\qquad = 2\,\Omega \int_1^{\Omega^{-2}u(\delta)} (t-1)^{-1/2}\, h(t\,\Omega^2)\, dt \\
&\qquad\quad - 2\left(u(\delta) - \Omega^2\right)^{-1/2} w(u(\delta))\, u(\delta) \qquad \text{(4.11.a)} \\
&\qquad = 2\,\Omega \sum_{n=0}^{N} a_n x_0^{-n/2} \Omega^n \int_1^{\Omega^{-2}u(\delta)} (t-1)^{-1/2}\, t^{n/2}\, dt \\
&\qquad\quad + \tilde{h}(\Omega) - \left(u(\delta) - \Omega^2\right)^{-1/2} \frac{2\,u(\delta)}{\delta\, u'(\delta)},
\end{aligned}$$

with

$$\begin{aligned}
&\left|\tilde{h}(\Omega)\right| \leq 2\,\Omega^{N+2} \varepsilon_h\, x_0^{-(N+1)/2} \\
&\qquad \cdot \int_1^{\Omega^{-2}u(\delta)} (t-1)^{-1/2}\, t^{(N+1)/2}\, dt. \qquad \text{(4.11.b)}
\end{aligned}$$

At this point, we introduce another small number, Ω_ε, on which we impose, first, the condition

$$u(\delta) \geq 2\,\Omega_\varepsilon^2. \qquad \text{(4.11.c)}$$

Expression (4.11.a) above can be computed easily for all $\Omega \geq \Omega_\varepsilon$. The evaluation of integrals of the type $\int (t-1)^{-1/2} t^\gamma \, dt$ can be done by enclosing the integrand locally in neighborhoods in H^1. We omit the trivial details.

When $\Omega \leq \Omega_\varepsilon$, consider first the following trivial Lemma.

Lemma 4.4. *If $R \geq 2$, then*

1. $\displaystyle\int_1^R (t-1)^{-1/2} t^\gamma \, dt \leq 2 R^{\gamma+1/2}$ *when* $\gamma \geq 0$.
2. $\displaystyle\int_1^R (t-1)^{-1/2} t^\gamma \, dt \geq 1$ *when* $-1 \leq \gamma$.
3. $\displaystyle\int_1^R (t-1)^{-1/2} t^\gamma \, dt = O\Big(R^{\gamma+1/2}\Big)$ *when* $\gamma > -1/2$.

Then, by (4.11.c), *a*), *b*), *c*) and Lemma 4.4,

$$\frac{d}{d\Omega}\left(\Omega \int_{r_1(\Omega)}^{\delta} \big(u(r) - \Omega^2\big)^{-1/2} \frac{dr}{r}\right)$$

is bounded below by

$$(4.12)\quad T_1(\Omega) \stackrel{\text{def}}{=} 4\sqrt{u(\delta)} \sum_{a_n < 0} a_n \left(\frac{u(\delta)}{x_0}\right)^{n/2} - \big(u(\delta) - \Omega^2\big)^{-1/2} \frac{2\,u(\delta)}{\delta\, u'(\delta)},$$

where we have set $a_{N+1} = -\varepsilon_h$.

Computation of I_3.

Here, we also consider two cases: $\Omega \geq \bar{\Omega}_3$ and $\Omega < \bar{\Omega}_3$. The first case is dealt with in a similar manner to I_2. We omit the trivial modifications. The second case is also treated in much the same way, with a few differences coming mainly from the different powers in the asymptotic expansion of u at 0 and at ∞. We include the details, although many of the differences are basically typographical considerations, because conclusions are somewhat different. In particular, as will be noted below, I_3 is mainly responsible for the singularity of F'' at $\Omega = 0$.

Let M be a large number satisfying

a) $u'(M) < 0$ and $u''(x) \geq 0$ on $[M, \infty)$.

b) For a sequence $\{\bar{b}_n\} \in l^1$, we have

$$\text{(4.13.a)} \qquad u(x) = \frac{144}{x^2}\, u_0(x) = \frac{144}{x^2}\left(1 + \sum_{n=1}^{\infty} \bar{b}_n\, \bar{x}^{-n\alpha}\right), \quad x \geq M.$$

Furthermore, we know that

$$\text{(4.13.b)} \qquad 1 + \sum_{n=1}^{\infty} \bar{b}_n z^n \in \mathcal{U}(I_0, \ldots, I_m; 0, C_g; 2), \qquad I_0 = [1,1].$$

Here, $\bar{x}$ denotes x/M, and, as a rule, we set $\bar{b}_n = b_n/M^{n\alpha}$.

c) $|u'(x)| \geq 2 \cdot 144\, x^{-3}(1 - \varepsilon')$, for $x \geq M$.

Define

$$\bar{t} = (M\, t^{1/2}/12)^{\alpha}.$$

We also consider a small number $\eta \leq u(M)$. It will be chosen so that (4.19) below holds.

We start by obtaining expressions for $u'(r)$ and $u''(r)$ similar to the one for u in (4.13.b). Note first that (4.13.b) is equivalent to an expression for $y(x)$. Then, by the Thomas-Fermi equation, and by integration, we have

$$\begin{aligned} y''(x) &= \frac{12 \cdot 144}{x^5}\left(\sum_{n=0}^{\infty} \bar{b}_n\, \bar{x}^{-n\alpha}\right)^{3/2} \\ &= \frac{12 \cdot 144}{x^5} \sum_{n=0}^{\infty} y''_n\, \bar{x}^{-n\alpha}, \qquad y''_0 = 1, \\ y'(x) &= \frac{-3 \cdot 144}{x^4} \sum_{n=0}^{\infty} \frac{4}{4 + n\alpha}\, y''_n\, \bar{x}^{-n\alpha} \\ &= \frac{-3 \cdot 144}{x^4} \sum_{n=0}^{\infty} y'_n\, \bar{x}^{-n\alpha}, \qquad y'_0 = 1, \end{aligned}$$

from which we obtain

$$\text{(4.14.a)} \qquad u'(x) = x\, y'(x) + y(x) = -\frac{2 \cdot 144}{x^3} \sum_{n=0}^{\infty} \frac{3\, y'_n - \bar{b}_n}{2}\, \bar{x}^{-n\alpha}$$

$$= -\frac{2 \cdot 144}{x^3} u_p(\bar{x}^{-\alpha}),$$

$$u''(x) = x\,y''(x) + 2\,y'(x) = \frac{6 \cdot 144}{x^4} \sum_{n=0}^{\infty} (2\,y_n'' - y_n')\,\bar{x}^{-n\alpha}$$

(4.14b)

$$= \frac{6 \cdot 144}{x^4} u_{pp}(\bar{x}^{-\alpha}),$$

for u_p and u_{pp} in H^1, normalized so $u_p(0) = u_{pp}(0) = 1$.

The strategy will be, as in the case for I_2, to change variables to the inverse function of u, $r(t)$.

Algorithm 4.5. *Given a function*

$$r(t) = 12\,t^{-1/2}\,R(\bar{t}),$$

with

$$R(z) \in \mathcal{U}^0(a_0^*, \ldots, a_N^*; C_N, 0; \infty; S), \qquad a_0^* = [1,1],$$

satisfying the hypothesis

$$r(t) \geq M, \qquad \text{whenever } \bar{t} \in S \subset [-1,1],$$

and given

$$g(x) = \sum_{n=0}^{\infty} \bar{c}_n z^n \in \mathcal{U}^1(I_0, \ldots, I_N; C_h, C_g, m), \qquad z = (x/M)^{-\alpha},$$

we compute another neighborhood of type ∞ in C^0, such that, if we set

$$G(\bar{t}) = g\,(r(t)) = \sum_{n=0}^{\infty} \bar{c}_n \left(\frac{r(t)}{M}\right)^{-n\alpha},$$

then

$$G(\bar{t}) \in \mathcal{U}^0(d_0^*, \ldots, d_N^*; C, 0; \infty; S),$$

valid for $\bar{t} \in S$.

DESCRIPTION: Consider $a_j \in a_j^*$, for $j = 0, \ldots, N$. Put

$$\left(\sum_{n=0}^{N} a_n \bar{t}^{\,n} + \bar{t}^{\,N+1} h(\bar{t})\right)^{\gamma} = \sum_{n=0}^{n_0} a_{n,\gamma}\,\bar{t}^{\,n} + \bar{t}^{\,n_0+1} h(\bar{t}; \gamma, n_0 + 1),$$

with $a_{n,\gamma} \in a^*_{n,\gamma}$, and

$$\|h(\bar{t};\gamma,n_0)\|_0 \leq \varepsilon_{\gamma,n_0}\,.$$

We can see, on one hand, that

$$\sum_{n=0}^{N} \bar{c}_n \left(\frac{r(t)}{M}\right)^{-n\alpha} = \sum_{n=0}^{N} \bar{c}_n\, \bar{t}^{\,n} \left(\sum_{k=0}^{N-n} a_{k,-n\alpha}\, \bar{t}^{\,k} + \bar{t}^{\,N+1-n} h(\bar{t};-n\alpha,N-n+1)\right)$$
$$= \sum_{n=0}^{N} d_n\, \bar{t}^{\,n} + \bar{t}^{\,N+1} h(t)\,,$$

with

$$|h(t)| \leq \sum_{n=0}^{N} |\bar{c}_n|\, \varepsilon_{-n\alpha,N-n+1}$$
$$\leq \sum_{n=0}^{N} |I_n|\, \varepsilon_{-n\alpha,N-n+1} + C_g \max_{n=m,\ldots,N} \varepsilon_{-n\alpha,N-n+1}\,,$$

and

$$d_n = \sum_{i+j=n} \bar{c}_i\, a_{j,-i\alpha}$$
$$\in \sum_{i+j=n} I_i\, a^*_{j,-i\alpha} \pm \varepsilon_n \stackrel{\text{def}}{=} d^*_n\,,$$

where

$$\varepsilon_n \leq \begin{cases} C_g \sup\limits_{i=m,\ldots,n} |a_{n-i,-i\alpha}|\,, & \text{if } n \geq m\,, \\ 0\,, & \text{otherwise.} \end{cases}$$

On the other hand, since

$$\left|\sum_{n=N+1}^{\infty} \bar{c}_n \left(\frac{r(t)}{M}\right)^{-n\alpha}\right| \leq (C_g + C_h)\left(\frac{r(t)}{M}\right)^{-(N+1)\alpha}$$
$$\leq (C_g + C_h)\, \bar{t}^{\,N+1} \left(1 - \sum_{n=1}^{N} |a_n| - C_N\right)^{-(N+1)\alpha}\,,$$

we conclude that

$$g(r(t)) = G(\bar{t}) \in \mathcal{U}^0(d_0^*, \dots, d_N^*; \tilde{C}_h, 0; \infty),$$

with

$$\tilde{C}_h = (C_g + C_h)\left(1 - \sum_{n=1}^{N} |a_n| - C_N\right)^{-(N+1)\alpha} + \sum_{n=0}^{N} |I_n|\, \varepsilon_{-n\alpha, N-n+1} + C_g \max_{n=m,\dots,N} \varepsilon_{-n\alpha, N-n+1}.$$

If we cannot check that

$$1 - \sum_{n=1}^{N} |a_n| - C_N > 0$$

the algorithm fails.

Now, we analyze $r(t)$.

Algorithm 4.6. *Given $N \geq 0$, we produce intervals $a_1^*, \dots, a_N^*$, and a constant C_N, such that*

$$\left| r(t) - 12\, t^{-1/2}\left(1 + \sum_{n=1}^{N} a_n \bar{t}^{\,n}\right)\right| \leq C_N\, t^{-1/2}\, \bar{t}^{\,N+1},$$

for constants $a_i \in a_i^*$, $i = 1, \dots, N$, and for $t \leq \eta$.

DESCRIPTION: First, we will construct an inductive procedure to define numbers a_n such that

$$r_N(t) = 12\, t^{-1/2}\left(1 + \sum_{n=1}^{N} a_n \bar{t}^{\,n}\right)$$

satisfies

$$u(r_N(t)) = t\left(1 + O(\bar{t}^{\,N+1})\right), \qquad \text{as } t \to 0. \tag{4.15}$$

By induction. For $N = 0$, let $r_0(t) = 12\, t^{-1/2}$. Note that, since $u(x) < 144\, x^{-2}$, then $r(t) < r_0(t)$. Also, $r_0(t) \geq M$ for $\bar{t} \leq 1$, and if $t \leq \eta \leq$

$u(M)$ then $\bar{t} \leq 1$. Furthermore, $t \leq \eta \leq u(M)$ implies $r(t) \geq M$, which we will need below.

Therefore, we have

$$u(r_0(t)) = t\left(1 + \sum_{n\geq 1} b_n\, 12^{-n\alpha}\, t^{n\alpha/2}\right).$$

Since $u(r(t)) = t$,

$$|r_0(t) - r(t)| \leq t\, \frac{\left|\sum_{n\geq 1} b_n\, 12^{-n\alpha}\, t^{n\alpha/2}\right|}{\inf_{r\in[r(t),\, r_0(t)]} |u'(r)|}.$$

Note that hypotheses *a*) and *c*) imply

$$|u'(r)| \geq |u'(r_0(t))| \geq \frac{t^{3/2}}{6}(1 - \varepsilon'),$$

provided $r \in [r(t), r_0(t)] \cap [M, \infty)$. Our previous remarks, and our assumption on η then imply

$$|r_0(t) - r(t)| \leq C\, t^{(\alpha-1)/2}, \qquad t \leq \eta.$$

Here we have used the fact that $r(t) \geq M$ for $t < \eta$.

For general N, we set

$$r_N(t) = 12\, t^{-1/2} \sum_{n=0}^{N} a_n \bar{t}^{\,n}, \qquad a_0 = 1.$$

where $a_1, \ldots, a_{N-1}$ satisfy the induction hypothesis.

Note that we have

$$\sum_{n=1}^{\infty} b_n\, r_N(t)^{-n\alpha} - \sum_{n=1}^{\infty} b_n\, r_{N-1}(t)^{-n\alpha} = O(\bar{t}^{\,N+1}). \tag{4.16}$$

Thus, if for any real number γ we put

$$\left(\sum_{n=0}^{N-1} a_n\, \bar{t}^{\,n}\right)^{\gamma} = \sum_{n=0}^{N} a_{n,\gamma}\, \bar{t}^{\,n} + O(\bar{t}^{\,N+1}), \tag{4.17}$$

we see that

$$u(r_{N-1}(t)) = t\left(\sum_{n=0}^{N} a_{n,-2}\,\bar{t}^{\,n} + O(\bar{t}^{\,N+1})\right)$$

(4.18)

$$\cdot\left(\sum_{n=0}^{N} \bar{b}_n\,\bar{t}^{\,n}\Big(\sum_{i=0}^{N} a_{i,-n\alpha}\,\bar{t}^{\,i} + O(\bar{t}^{\,N+1})\Big) + O(\bar{t}^{\,N+1})\right)$$

$$= t\left(\sum_{n=0}^{N} a_{n,-2}\,\bar{t}^{\,n} + O(\bar{t}^{\,N+1})\right)\left(\sum_{n=0}^{N} c_n\,\bar{t}^{\,n} + O(\bar{t}^{\,N+1})\right),$$

where

$$c_k = \sum_{n=0}^{k} \bar{b}_n\,a_{k-n,-n\alpha}\,, \qquad c_0 = 1\,.$$

By the induction hypothesis, (4.18) is equal to $t(1 + O(\bar{t}^{\,N}))$. Thus, using (4.16), we can see that

$$u(r_N(t)) = t\left(\sum_{n=0}^{N} a_{n,-2}\,\bar{t}^{\,n} - 2\,a_N\,\bar{t}^{\,N} + O(\bar{t}^{\,N+1})\right)$$

$$\cdot\left(\sum_{n=0}^{N} c_n\,\bar{t}^{\,n} + O(\bar{t}^{\,N+1})\right),$$

for exactly the same c_n as in (4.18).

Therefore, by putting

$$a_N = \frac{1}{2}\sum_{m+n=N} a_{n,-2}\,c_m$$

we get rid of all $\bar{t}^{\,N}$ terms, thus obtaining (4.15).

So far, we have proved the existence of numbers a_n such that (4.15) is satisfied.

This procedure also gives us an algorithm to compute the a_n^*. Indeed, bounds $a_{n,\gamma}^*$ for the $a_{n,\gamma}$ can be computed explicitly, since by the induction hypothesis we already know a_i^*, for $i = 1, \ldots, N-1$. As for the c_k, recalling (4.13.b),

$$c_k = \sum_{n=0}^{k} \bar{b}_n\,a_{k-n,-n\alpha} \in \sum_{n=0}^{k} I_n\,a_{k-n,-n\alpha}^* \pm \varepsilon_k\,,$$

with

$$|\varepsilon_k| \leq \begin{cases} C_g \sup\limits_{2\leq n\leq k} |a_{k-n,-n\alpha}|\,, & \text{if } k \geq 2\,, \\ 0\,, & \text{if } k \leq 1\,. \end{cases}$$

In particular, it is easy to see that $a_1 = \bar{b}_1/2 \in \frac{1}{2} I_1$ (recall (4.13)).

To obtain a good value for the constant C_N, we proceed as follows: By Algorithm 4.5, we can construct a neighborhood $\mathcal{U}_1^0$ such that

$$f(\bar{t}) = u_0\,(r_N(t)) \in \mathcal{U}_1^0(I_0, \ldots, I_N; \tilde{\tilde{C}}_h, 0; \infty; t \leq \eta)\,,$$

(see (4.13.a)) provided $r_N(t) \geq M$ for $t \leq \eta$. At this point then we check that

$$12\,\eta^{-1/2}\left(1 - \sum_{n=1}^{N} |a_n|\right) \geq M\,.$$

See also (4.19) below. If we put

$$g(\bar{t}) = \left(\sum_{n=0}^{N} a_n \bar{t}^{\,n}\right)^{-2} \in \mathcal{U}_2(\infty)\,,$$

then,

$$u\,(r_N(t)) = t\,g(\bar{t})\,f(\bar{t})\,,$$

with

$$F(\bar{t}) = g(\bar{t})\,f(\bar{t}) \in \mathcal{U}^0(\hat{I}_0, \ldots, \hat{I}_N; \hat{C}_h, 0; \infty)$$

and $\mathcal{U}^0$ is the product neighborhood of $\mathcal{U}_1^0$ and $\mathcal{U}_2$. Note that (4.15) implies that $F(\bar{t}) = 1 + O(\bar{t}^{\,N+1})$, therefore, we can take $\hat{I}_1 = [1,1]$ and $\hat{I}_i = [0,0]$, for $i = 1, \ldots, N$. This implies that

$$\left|u\big(r_N(t)\big) - t\right| \leq \hat{C}_h\,t\,\bar{t}^{\,N+1}\,, \qquad t \leq \eta\,.$$

Now, note that

$$|r_N(t) - r(t)| \leq \frac{|u(r_N(t)) - t|}{\inf_{M\leq r\leq \max\{r(t), r_N(t)\}} |u'(r)|}\,.$$

At this point, we check that $a_i < 0$ for all $i = 1, \ldots, N$, which implies that $r_N(t) \leq r_0(t)$. We then conclude, by hypothesis c), that

$$|u'(r)| \geq \left|u'\big(r_0(t)\big)\right| \geq \frac{1}{6}\,t^{3/2}(1 - \varepsilon')\,, \qquad M \leq r \leq \max\{r(t), r_N(t)\}\,,$$

from which the algorithm follows by taking $C_N = 6\,\hat{C}_h/(1-\varepsilon')$.

Now we compute bounds for the derivatives of $r'(t)$, also in the C^0 topology; this will allow us to compute bounds for $w(t) = -r'(t)/r(t)$. We check first that

$$12\,\eta^{-1/2}\left(1 - \sum_{n=1}^{N} |a_n| - \tilde{C}_N\right) \geq M\,, \qquad \tilde{C}_N = \frac{C_N}{12}\,. \tag{4.19}$$

Algorithm 4.7. *We produce a neighborhood*

$$\mathcal{U}^0(I_0, \ldots, I_N; C_h, 0; \infty; t \leq \eta)$$

such that

$$h(t) \stackrel{\text{def}}{=} t\,w'(t) + w(t) = t^{-1}\,f(\bar{t})\,, \qquad f \in \mathcal{U}\,,$$

for $t \leq \eta$, where we define $w(t) = -r'(t)/r(t)$.

DESCRIPTION: We note that

$$r'(t) = \frac{1}{u'\big(r(t)\big)}\,, \qquad r''(t) = -r'(t)^3\,u''\big(r(t)\big)\,.$$

As a result of this, in view of (4.14a) and (4.14b) and Algorithm 4.5, we obtain neighborhoods containing functions f, f_p and f_{pp} such that

$$r(t) = 12\,t^{-1/2} f(\bar{t})\,, \quad r'(t) = -6\,t^{-3/2} f_p(\bar{t})\,, \quad r''(t) = 9\,t^{-5/2} f_{pp}(\bar{t})\,,$$

which are valid for $t \leq \eta$, and where

$$f_p(\bar{t}) = f^3(\bar{t})\,u_p\Big(\big(\frac{r(t)}{M}\big)^{-\alpha}\Big)^{-1}\,,$$
$$f_{pp}(\bar{t}) = f_p^3(\bar{t})\,f^{-4}(\bar{t})\,u_{pp}\Big(\big(\frac{r(t)}{M}\big)^{-\alpha}\Big)\,.$$

Thus,

$$w(t) = \frac{-r'(t)}{r(t)} = \frac{1}{2\,t}\,\frac{f_p(\bar{t})}{f(\bar{t})}\,,$$

$$w'(t) = \left(\frac{r'(t)}{r(t)}\right)^2 - \frac{r''(t)}{r(t)} = \frac{1}{4t^2}\left(\frac{f_p(\bar{t})}{f(\bar{t})}\right)^2 - \frac{3}{4t^2}\frac{f_{pp}(\bar{t})}{f(\bar{t})},$$

and

$$\begin{aligned} h(t) &= t\,w'(t) + w(t) \\ &= t^{-1}\left(\frac{1}{2}\frac{f_p(\bar{t})}{f(\bar{t})} + \frac{1}{4}\left(\frac{f_p(\bar{t})}{f(\bar{t})}\right)^2 - \frac{3}{4}\frac{f_{pp}(\bar{t})}{f(\bar{t})}\right) = \frac{1}{4t} f_h(\bar{t}), \end{aligned}$$

for a function f_h belonging to an easily computable neighborhood in C^0.

Note now that our choice of functions was normalized so that $f(0) = f_p(0) = f_{pp}(0) = 1$, which implies that $f_h(0) = 0$. This is important: it says that if the Thomas-Fermi potential were *equal* to $c\,x^{-3}$, then $F'(\Omega)$ (which still makes sense) would be constant (recall Section 1), and would make Theorem 1.1 completely wrong. But it is not, and one can easily see that

$$f_h(\bar{t}) = -\frac{\alpha^2 b_1}{2 \cdot 12^\alpha} t^{\alpha/2} + O(t^\alpha) \tag{4.20}$$

as follows: recall that

$$r(t) = 12\,t^{-1/2}\left(1 + \frac{\bar{b}_1}{2}\,\bar{t} + O(\bar{t}^2)\right).$$

Note that

$$\begin{aligned} r'(t) &= -6\,t^{-3/2}\left(1 - \frac{(\alpha-1)b_1}{2 \cdot 12^\alpha}\,t^{\alpha/2} + O(t^\alpha)\right), \\ r''(t) &= 9\,t^{-5/2}\left(1 + \frac{(\alpha-1)(\alpha-3)b_1}{6 \cdot 12^\alpha}\,t^{\alpha/2} + O(t^\alpha)\right), \end{aligned}$$

which are easily guessed by termwise differentiation of the expression for $r(t)$, and -not so easily- checked using the formulas for $r'(t)$ and $r''(t)$ above. This yields

$$\begin{aligned} h(t) = \frac{1}{4t}\Bigg(&2\Big(1 - \frac{(\alpha-1)b_1}{2 \cdot 12^\alpha}t^{\alpha/2}\Big)\Big(1 - \frac{b_1}{2 \cdot 12^\alpha}t^{\alpha/2}\Big) \\ &+ \Big(1 - \frac{(\alpha-1)b_1}{12^\alpha}t^{\alpha/2}\Big)\Big(1 - \frac{b_1}{12^\alpha}t^{\alpha/2}\Big) \end{aligned}$$

$$-3\Big(1+\frac{(\alpha-1)(\alpha-3)b_1}{6\cdot 12^\alpha}t^{\alpha/2}\Big)\Big(1-\frac{b_1}{2\cdot 12^\alpha}t^{\alpha/2}\Big)+O(t^\alpha)\Big)$$

which immediately implies (4.20).

Now, let

$$f_h(\bar{t})=\sum_{n=1}^{N}a_n\bar{t}^{\,n}+H(\bar{t})\,,$$

with

$$|H(\bar{t})|\le\varepsilon_h\,|\bar{t}|^{N+1}\,,\qquad t\le\eta\,.$$

Finally, then, let L be a large number such that $u(L)\le\eta$: we set $\bar{\Omega}_3\le\sqrt{u(L)}$, $b(\Omega)=L$ for all $\Omega\le\bar{\Omega}_3$, and, arguing as before, we have

$$\begin{aligned}
I_3&=\frac{d}{d\Omega}\left(\Omega\int_L^{r_2(\Omega)}\big(u(r)-\Omega^2\big)^{-1/2}\,\frac{dr}{r}\right)\\
&=\frac{d}{d\Omega}\left(\Omega\int_{\Omega^2}^{u(L)}\big(t-\Omega^2\big)^{-1/2}\,w(t)\,dt\right)\\
&=\frac{d}{d\Omega}\left(\Omega^2\int_1^{\Omega^{-2}u(L)}(t-1)^{-1/2}\,w(t\,\Omega^2)\,dt\right)\\
&=2\,\Omega\int_1^{\Omega^{-2}u(L)}(t-1)^{-1/2}\,h(t\,\Omega^2)\,dt \qquad\text{(4.21.a)}\\
&\quad-2\,\big(u(L)-\Omega^2\big)^{-1/2}\,w(u(L))\,u(L)\\
&=\frac{1}{2}\,\Omega^{-1}\sum_{n=1}^{N}\frac{a_nM^{n\alpha}}{12^{n\alpha}}\,\Omega^{n\alpha}\int_1^{\Omega^{-2}u(L)}(t-1)^{-1/2}\,t^{\alpha n/2-1}dt\\
&\quad+\tilde{h}(\Omega)+\big(u(L)-\Omega^2\big)^{-1/2}\,\frac{2\,u(L)}{L\,u'(L)}\,,
\end{aligned}$$

with

$$\begin{aligned}
\big|\tilde{h}(\Omega)\big|&\le\frac{1}{2}\,\Omega^{-1+\alpha(N+1)}\varepsilon_h\left(\frac{M}{12}\right)^{(N+1)\alpha}\qquad\text{(4.21.b)}\\
&\quad\cdot\int_1^{\Omega^{-2}u(L)}(t-1)^{-1/2}\,t^{-1+(N+1)\alpha/2}\,dt\,.
\end{aligned}$$

Now, we recall Ω_ε, on which we impose now the extra condition

$$u(L) \geq 2\,\Omega_\varepsilon^2 . \tag{4.22}$$

Both (4.21.a) and (4.21.b) can be computed easily for all $\Omega \geq \Omega_\varepsilon$. The evaluation of integrals of the type $\int (t-1)^{-1/2} t^\gamma \, dt$ can be done by the same method as in the previous section.

When $\Omega \leq \Omega_\varepsilon$, note that the first term in (4.21.a) goes to infinity as $\Omega \to 0$, while all the others remain bounded. We use this to obtain a uniform lower bound for the absolute value of this derivative.

By (4.20) we know that $a_1 > 0$; thus, we have

$$\begin{aligned} \frac{d}{d\Omega}\left(\Omega \int_L^{r_2(\Omega)} \left(u(r) - \Omega^2\right)^{-1/2} \frac{dr}{r}\right) & \\ \geq \frac{1}{2}\,\Omega^{-1+\alpha} a_1 \left(\frac{M}{12}\right)^\alpha & \\ + u(L)^{-1/2} \sum_{\substack{a_n < 0 \\ n \geq 3}} a_n \left(\frac{u(L)\,M^2}{144}\right)^{n\alpha/2} & \\ + \left(u(L) - \Omega^2\right)^{-1/2} \frac{2\,u(L)}{L\,u'(L)} , & \end{aligned} \tag{4.23}$$

where we have put $a_{N+1} = -\varepsilon_h$.

Note that, since the exponent $\gamma = \alpha - 1$ does not fall under the cases considered in Lemma 4.4, (4.23) is only correct provided $a_2 > 0$. Of course, one can try to modify Lemma 4.4 to include the case $n = 2$, but since it so happens that $a_2 > 0$ there is no need. By this we mean that we check $a_2 > 0$: if the check fails, our proof of Theorem 1.1 fails, and we claim no theorem.

Putting together now (4.1.a)-(4.1.c), (4.12), (4.11.c), (4.22) and (4.23), we conclude that, if

$$\Omega^2 \leq \frac{1}{2} \min\{u(L), u(\delta)\} ,$$

then

$$-F''(\Omega) \geq \frac{1}{2}\, a_1 \left(\frac{M}{12}\right)^\alpha \Omega^{-1+\alpha} + T_1(\Omega) + T_2(\Omega) + T_3 , \tag{4.24}$$

for

$$T_2(\Omega) = u(L)^{-1/2} \sum_{a_n<0} a_n \left(\frac{u(L)M^2}{144}\right)^{n\alpha/2} + \left(u(L) - \Omega^2\right)^{-1/2} \frac{2\,u(L)}{L\,u'(L)} < 0\,,$$

$$T_3 = \int_\delta^L \left(y(r)\right)^{-1/2} \frac{dr}{r^{3/2}} > 0\,.$$

The following is a consequence of formulas (4.21.a,b), (4.11.a,b) and Lemma 4.4.

Proposition 4.8. *We have*

$$F''(\Omega) = -c_0\,\Omega^{-1+\alpha} + O(1)\,, \qquad c_0 > 0\,, \qquad \text{as} \quad \Omega \to 0\,.$$

We now organize the main results in this section in the following algorithm.

Algorithm 4.9. *Given representable δ (small) and L (large), and given neighborhoods in C^0 containing the functions $h(t)$ in Algorithm 4.3 and 4.7, valid for $0 < t \le u(\delta)$ and $0 < t \le u(L)$ respectively, we compute strictly positive lower bounds for $-F''(\Omega^*)$ for all thin subintervals Ω^* of Zone I.*

DESCRIPTION: Note that our hypotheses imply that the requirements for the smallness of η and largeness of L have already been checked.

Break up $-F''$ into the three terms in (4.1). I_1 can be computed as described earlier all the way down to $\Omega = 0$.

If $\Omega^* \ge \bar{\Omega}_2$ (similarly for $\bar{\Omega}_3$), I_2 can also be computed as described above. Thus we are left only with the computation of I_2 for $\Omega^* \le \bar{\Omega}_2$ (similarly for I_3). Note that the dicotomy $\Omega_2^* \ge \bar{\Omega}_2$ or $\Omega_2^* \le \bar{\Omega}_2$ can be trivially achieved by choosing $\bar{\Omega}_2$ to be one of the endpoints of the Ω^*.

We assume first that $\Omega^* \ge \Omega_\varepsilon$. We begin by computing bounds for ${\Omega^*}^{-2}u(\delta)$: if we cannot check that these bounds are greater than or equal to 1, we report a failure and quit. Otherwise, we compute bounds for I_2 using (4.11.a) with the error bound for $\tilde{h}$ given by (4.11.b). Note that this procedure will prove, in particular, that $\bar{\Omega}_2$ satisfies

the smallness requirement above. This is of mild importance since the choice of $\bar{\Omega}_2$ will not be explicit in our computer implementation.

When $\Omega \leq \Omega_\varepsilon$, we simply check that the right hand side of (4.24) is strictly positive for $\Omega = \Omega_\varepsilon$. Trivial monotonicity properties will then imply the positivity for $0 < \Omega \leq \Omega_\varepsilon$. We also check that requirements (4.11.c) and (4.22) for Ω_ε are satisfied.

5. Zone II.

The purpose of this section is to prove (1.1) for all Ω close to Ω_c.

Lemma 5.1. *Let $f(z)$ be analytic in $|z - z_0| < R$, continuous up to the boundary, with*

1. $f'(z_0) \neq 0$.
2. $f(z) = w_0$ *if and only if* $z = z_0$.
3. *If* $|z - z_0| = R$, *then* $|f(z) - w_0| \geq T$.

Then, there exists $F(w)$ analytic,

$$F : B(w_0, T) \longrightarrow B(z_0, R),$$

such that $f(F(w)) = w$, for $w \in B(z_0, R)$.

PROOF. Consider the curves

$$\Gamma_s(t) = f\big(\gamma_s(t)\big), \qquad |\gamma_s(t) - z_0| = s, \quad 0 < s < R,$$

with γ_s positively oriented. Condition 2 implies that

$$n(\Gamma_s, w_0) = \frac{1}{2\pi i} \int_{\gamma_s} \frac{f'(z)}{f(z) - w_0}\, dz$$

is continuous in $0 < s < R$, and it is therefore constant.

Condition 1 says that $n(\Gamma_s, w_0) = 1$ for all s small enough. Thus,

$$n(\Gamma_s, w_0) = 1, \qquad 0 < s < R. \tag{5.1}$$

Finally, let $\varepsilon > 0$ be given. Condition 3 implies that $B(w_0, T - \varepsilon)$ does not intersect Γ_s for $R - \varepsilon' < s < R$, for some other ε': indeed, assume not: then, there exists z_n, such that $|z_n - z_0| \to R$ such that

$|f(z_n)-w_0| < T-\varepsilon$. Passing to a subsequence, $z \to z_\infty$ with $|z_\infty - z_0| = R$ and $|f(z_\infty) - w_0| \leq T - \varepsilon$, which contradicts 3.

Therefore, $B(w_0, T-\varepsilon)$ is contained in one of the connected components of the complement of Γ_s for $R - \varepsilon' < s < R$. This implies that the index is constant in w, *i.e.*

$$n(\Gamma_s, w) = n(\Gamma_s, w_0) = 1\,, \quad R - \varepsilon' < s < R\,, \quad w \in B(w_0, T-\varepsilon)\,.$$

Thus, if $\alpha_i(w)$ are the solutions of $f(z) = w$ inside $B(z_0, R)$,

$$\sum n(\alpha_i, \gamma_s) = \frac{1}{2\pi i} \int_{\gamma_s} \frac{f'(z)}{f(z) - w}\, dz = n(w, \Gamma_s) = 1\,,$$

for

$$R - \varepsilon' < s < R\,, \qquad w \in B(w_0, T-\varepsilon)\,,$$

from which, taking $\varepsilon \to 0$, we deduce that there is only one α_i and $f'(\alpha_i) \neq 0$. This implies that f^{-1} exists and is analytic.

Lemma 5.2. *Let $u \in H^1(|z - r_c| \leq R)$, smooth on the boundary of $B(r_c, R)$, of the form*

$$u(x) = \Omega_c^2 - u_2\, R^2\, z^2 + z^3 f(z)\,, \qquad z = \frac{x - r_c}{R}\,, \qquad f(0) = u_3 R^3\,,$$

satisfying

1. $\|f\| \leq h$, $u_2 > 0$ *and* $u_2 R^2 > h$.
2. *For a constant M we have*

$$\left| \frac{d^4}{dx^4} u(x) \right| \leq M\,, \qquad |z| \leq 1\,.$$

Then, $t(x)$ as in (1.4) can be extended analytically to $B(r_c, R)$, and there is an inverse $r(w)$ of $t(x)$, analytic in $|w| < T$ where

$$T \leq \sqrt{u_2 R^2 - h}\,,$$

and

$$\sup_{|w| \leq T} |r'(w)| \leq \frac{2\sqrt{u_2 + hR^{-2}}}{2\,u_2 - 3\,|u_3|\,R - MR^2/6}\,,$$

$$\left| \frac{d^{n+1} r}{dw^{n+1}}(0) \right| \leq n!\, T^{-n}\, \frac{2\sqrt{u_2 + hR^{-2}}}{2\,u_2 - 3\,|u_3|\,R - MR^2/6}\,, \qquad n \geq 0\,.$$

PROOF. First, note that 1. implies that

$$t(x) = z\sqrt{u_2 R^2 - z f(z)}$$

exists as an analytic function in $x \in B(r_c, R)$ (*i.e.*, $|z| \leq 1$), since the radicand never vanishes and a ball is simply connected, and note also that this definition agrees with (1.4) if x is real.

Also, note that $t(x)$ satisfies the hypothesis of the previous lemma in the circle $x \in B(r_c, R)$. Indeed, $t(x) \neq 0$ unless $x = r_c$ (by 1.), and if $|x - r_c| = R$, then $|z| = 1$ and

$$|t(x)| \geq \sqrt{u_2 R^2 - |f(z)|} \geq T\,.$$

Therefore, by the previous lemma, $r(w)$ exists for all $|w| < T$, and we also have $|r(w) - r_c| \leq R$.

Now, note that

$$\sup_{|w|<T} |r'(w)| \leq \sup_{|z|<1} \frac{1}{|t'(x)|} \leq 2 \left(\sup_{|z|\leq 1} \frac{|u(x) - u(r_c)|}{|u'(x)|^2} \right)^{1/2}.$$

Since

$$|u(x) - u(r_c)| \leq u_2 |x - r_c|^2 + \frac{h |x - r_c|^3}{R^3}\,,$$

and

$$\begin{aligned} |u'(x)| &\geq 2\,u_2 |x - r_c| - 3\,|u_3|\,|x - r_c|^2 - \frac{1}{6} M\,|x - r_c|^3 \\ &\geq |x - r_c| \left(2\,u_2 - 3\,|u_3|\,R - \frac{1}{6} M R^2 \right), \end{aligned}$$

we deduce that

$$\frac{|u(x) - u(r_c)|}{|u'(x)|^2} \leq \frac{u_2 + h R^{-2}}{(2\,u_2 - 3\,|u_3|\,R - M R^2/6)^2}\,.$$

The other conclusion follows from Cauchy's inequalities applied to $r'(w)$ on $|w| < T$.

We now switch to the notation of Lemma 1.2.

Algorithm 5.3. *Given $\mathcal{U}_0(I_0, \dots, I_{2N+1}; C_h, 0; \infty)$ and bounds $A^0 < A_0$, we construct a representable T and $\mathcal{U}_1$, such that if*

$$u(x) = \Omega_c^2 - z^2 f(z), \qquad z = \frac{x - r_c}{R},$$

and

$$A^0 \le |y(x)| \le A_0, \qquad |x - r_c| \le R,$$

with $f \in \mathcal{U}_0$, then $w(t) = g(t/T)$, with $g \in \mathcal{U}_1$.

DESCRIPTION: Note that, since y satisfies (3.3) on $B(r_c, R)$, we have the identities

$$y'''(x) = \frac{3}{2}\frac{y^{1/2}\,y'}{x^{1/2}} - \frac{1}{2}\frac{y^{3/2}}{x^{3/2}},$$

$$y''''(x) = \frac{3}{4}\left(\frac{y^{-1/2}\,(y')^2}{x^{1/2}} + \frac{2\,y^2}{x} - \frac{2\,y^{1/2}\,y'}{x^{3/2}} + \frac{y^{3/2}}{x^{5/2}}\right),$$

$$u''''(x) = 4\,y'''(x) + x\,y''''(x),$$

which imply

$$|y''(x)| \le A_2 \stackrel{\text{def}}{=} \frac{A_0^{3/2}}{(r_c - R)^{1/2}}, \tag{5.2a}$$

$$|y'(x)| \le A_1 \stackrel{\text{def}}{=} \frac{\Omega_c^2}{r_c^2} + A_2 R, \tag{5.2b}$$

$$|y'''(x)| \le A_3 \stackrel{\text{def}}{=} \frac{3A_0^{1/2}}{2(r_c - R)^{1/2}} A_1 + \frac{A_0^{3/2}}{2(r_c - R)^{3/2}}, \tag{5.2c}$$

$$|y''''(x)| \le A_4 \stackrel{\text{def}}{=} \frac{3}{4}\left(\frac{A^{0\,-1/2} A_1^2}{(r_c - R)^{1/2}} + \frac{2A_0^2}{r_c - R} + \frac{2A_0^{1/2} A_1}{(r_c - R)^{3/2}} + \frac{A_0^{3/2}}{(r_c - R)^{5/2}}\right), \tag{5.2d}$$

$$|u''''(x)| \le M \stackrel{\text{def}}{=} 4\,A_3 + A_4(r_c + R). \tag{5.2e}$$

Choose representable T and $\tilde{T}$ such that

$$T < \tilde{T} \le \sqrt{I_0 - h}, \qquad h = \sum_{n=1}^{2N+1} |I_n| + C_h,$$

and put $\beta = T/\tilde{T}$. Here we assume that $I_0 > h$ and $\beta < 1$. Otherwise, the algorithm fails.

By the previous lemma, $r'(t)$ can be written as

$$r'(t) = \sum_{n \geq 0} r'_n (t/T)^n ,$$

where r'_n can be computed by power-matching, for $n = 1, \ldots, 2N+1$, because we have $C_g = 0$, and

$$\sum_{n > 2N+1} |r'_n| \leq \frac{2R^{-1}\sqrt{I_0 + h}}{R^{-2}(2|I_0| - 3|I_1|) - MR^2/6} \cdot \frac{\beta^{2N+2}}{1-\beta} . \tag{5.4}$$

A neighborhood of type ∞ and order $2N+2$ containing $r(t)$ can be obtained by integration.

This allows us to construct a neighborhood $\mathcal{U}_1$ of type ∞ and order $2N+1$ containing the function g in the statement of the algorithm, by simply dividing the neighborhood for r' by the neighborhood for r.

Algorithm 5.4. *We compute a bound for F'' in Zone II.*

DESCRIPTION: Note that

$$\begin{aligned} \alpha_n &= \frac{1}{\pi} \int_{-1}^{1} (1-t^2)^{-1/2} t^{2n} \, dt \\ &= \frac{1}{2^{2n+1} i^{2n} \pi} \int_0^{2\pi} \left(e^{i\theta} - e^{-i\theta}\right)^{2n} d\theta \\ &= \binom{2n}{n} 2^{-2n} , \end{aligned} \tag{5.5}$$

so their computation poses no difficulty. Note also that $\alpha_n > \alpha_{n+1}$ for all $n \geq 0$.

Therefore, by (1.8), if we set $\bar{w}_n = T^n w_n$, we can see that

$$\begin{aligned} -\frac{1}{\pi} F'(\Omega) &= \Omega \sum_{n=0}^{\infty} \bar{w}_{2n} \left(\frac{D}{T}\right)^{2n} \alpha_n \\ &= \Omega \sum_{n=0}^{N} \bar{w}_{2n} \left(\frac{D}{T}\right)^{2n} \alpha_n + \Omega \sum_{n>N} \bar{w}_{2n} \left(\frac{D}{T}\right)^{2n} \alpha_n . \end{aligned}$$

The first term is a polynomial in Ω, so we can easily compute its derivative anywhere. In fact, its derivative equals

$$\sum_{n=0}^{N} \bar{w}_{2n}\, \alpha_n \gamma^n + \Omega \sum_{n=1}^{N} \bar{w}_{2n}\, n\, \alpha_n\, \gamma^{n-1} \left(\frac{-2\Omega}{T^2}\right),$$

with

$$\gamma = \left(\frac{D}{T}\right)^2 .$$

Here we check that Zone II is included in the set of Ω that make $\gamma < 1$. Otherwise, we report a failure and we quit the proof.

As for the other term, using Lemma 2.1, and taking

$$C_h \geq \sum_{n>N} |\bar{w}_{2n}| ,$$

we have

$$\begin{aligned}
&\left| \frac{d}{d\Omega} \left(\Omega \sum_{n>N} \bar{w}_{2n} \left(\frac{D}{T}\right)^{2n} \alpha_n \right) \right| \\
&\quad \leq C_h\, \alpha_{N+1}\, \gamma^{N+1} + 2\,\Omega_c^2 \sum_{n>N} \frac{\alpha_n\, |\bar{w}_{2n}|\, n\, \gamma^{n-1}}{T^2} \\
&\quad \leq C_h\, \alpha_{N+1} \left(\gamma^{N+1} + \frac{2\,\Omega_c^2}{T^2} \sup_{n>N} n\, \gamma^{n-1} \right) \\
&\quad \leq \begin{cases} C_h\, \alpha_{N+1}\, \gamma^N \left(\gamma + \dfrac{2(N+1)\Omega_c^2}{T^2} \right), & \text{if } N+1 \geq \dfrac{1}{|\log \gamma|} \\[2ex] C_h\, \alpha_{N+1} \left(\gamma^{N+1} + \dfrac{2\,\Omega_c^2}{e\,T^2\,\gamma\,|\log \gamma|} \right), & \text{in any case.} \end{cases}
\end{aligned}$$

All previous expressions can be easily computed using (5.5). Also, note the slight improvement in the result as a consequence of taking the neighborhoods in the previous algorithm to be of odd order.

This concludes the description of all algorithms needed for the proof of Theorem 1.1. We summarize its computer-assisted proof in the following algorithm.

Algorithm 5.5. (PROOF OF THEOREM 1.1) *We produce a constant c such that Theorem* 1.1 *holds.*

DESCRIPTION: Run Algorithm 3.20 (and algorithms thereof) to obtain all necessary knowledge of the Thomas-Fermi function.

Take $\bar{\Omega}$ as explained at the end of Section 1. to define Zone I and Zone II. As stated earlier in this section, we check that $\gamma(\bar{\Omega}) < 1$. Then, we compute an upper bound for F'' in Zone II, and check that it is strictly negative.

Choose Ω_ε, $\bar{\Omega}_2$ and $\bar{\Omega}_3$ as in Section 4, and a partition consisting of fat subintervals of $[\Omega_\varepsilon, \bar{\Omega}]$ whose endpoints contain both $\bar{\Omega}_2$ and $\bar{\Omega}_3$ and a subpartition of thin intervals Ω_i^*. We compute the numbers $a_{k,i}$ and $b_{k,i}$. Next, we check that F'' is bounded above by a strictly negative number on the interval $(0, \Omega_\varepsilon]$ and on $[\Omega_\varepsilon, \bar{\Omega}]$ as described in Algorithm 4.9.

Theorem 1.1 then follows by taking the maximum of all these -finitely many- strictly negative constants.

6. Some Extensions.

The purpose of this Section is to extend Theorem 1.1 to a neighborhood of the Thomas-Fermi potential, in an appropriate topology. As pointed out earlier, the fact that Theorem 1.1 holds is a rather delicate one. The following theorem shows this precisely.

Theorem 6.1. *Given any two large numbers N and R, and given ε small, there exists a smooth function $f(x)$ such that*

a) $f(x) = y(x)$ *for* $0 \le x \le R$.

b) For all $x \ge R$, *and all* $n \ge 0$, *we have that*

$$\left| \frac{d^n}{dx^n} f(x) - \frac{d^n}{dx^n} y(x) \right| \le \varepsilon \, C_n \, x^{-3-n},$$

and, however, we also have that $F_f(\Omega)$ vanishes at least N times in $(0, \Omega_c)$. (Note that, if R is large enough, Ω_c is independent of f).

The C_n are universal constants. In particular, they are independent of ε, R and f.

PROOF. Note that we can assume R to be as large as we need. By Corollary 1.3, $F_f''(\Omega)$ is bounded in the range $\Omega_\varepsilon \leq \Omega < \Omega_c$. Also, $F_f''(\Omega) = F_y''(\Omega) < 0$ in the same range. From a trivial adaptation of Section 4, it follows that if

$$f(x) = \frac{144}{x^3}\left(1 + b\,x^{-\alpha}\right), \qquad x \geq R,$$

then

$$F_f''(\Omega) = c_1\, b\, \Omega^{-1+\alpha} + O(1), \qquad \alpha = \frac{\sqrt{73}-7}{2} < 1,$$

for $c_1 > 0$, uniformly in Ω. Therefore, taking $b = \varepsilon$, there is $\Omega_1 \ll \Omega_\varepsilon$ such that $F_f''(\Omega_1) > 0$. This gives a function f such that F_f'' has at least one zero. To get more zeros, take R_1 large depending on Ω_1 so that $F_f(\Omega)$, $\Omega \in [\Omega_1, \Omega_c)$, is independent of $f(x)$ outside of $x \in (0, R_1)$. Then, define

$$f(x) = \frac{144}{x^3}\left(1 - \varepsilon\, x^{-\alpha}\right), \qquad x \geq 2\,R_1,$$

and smooth. Then, for Ω_2 small enough $F_f''(\Omega_2) < 0$. This gives us two zeros for F_f''. And so on.

From this theorem it is then clear that if we want Theorem 1.1 to hold, we need a stronger grip of the behavior of the function f at infinity.

The following theorem is just a consequence of the rest of this article. Its proof is computer assisted.

Theorem 6.2. *There exist N large integer, C and x_1 large constants, and $\varepsilon > 0$ and $x_0 > 0$ small, such that if $f(x)$ satisfies*

1. $\|f - y\|_{C^N[x_0, x_1]} \leq \varepsilon$.
2. $f(x) = 1 - w\,x + x^{3/2} g\left((x/x_0)^{1/2}\right)$ *with* $|w - w_0| \leq \varepsilon$, *$g$ analytic and* $\|g - g_0\|_1 \leq \varepsilon$, *where* $y_{TF}(x) = 1 - w_0\,x + x^{3/2} g_0\left(x(x/x_0)^{1/2}\right)$.
3. *Recall formula* (4.13.a). *Then,*

$$f(x) = \frac{144}{x^3}\left(1 + \sum_{n=1}^{\infty} \bar{a}_n\, \bar{x}^{\,-n\alpha}\right), \qquad x \geq R,$$

with

$$\sum_{n=1}^{\infty} |\bar{b}_n - \bar{a}_n| \leq \varepsilon .$$

Here, ε is assumed to be small enough so that our assumptions on f stated at the beginning of Section 1 are satisfied.

Then,

$$F_f(\Omega) \leq c < 0 , \qquad \Omega \in (0, \max |r\, f(r)|) .$$

PROOF (COMPUTER-ASSISTED). Take η any small number. If ε is small enough, hypothesis 2. and 3. imply that formulas (4.11.a,b) and (4.21.a,b) remain valid for $f(x)$ by perturbing the a_n and ε_h by at most η percent. Also, for ε small enough, hypothesis 1. implies that the value of integral I_1 in (4.1.a) remains valid for f also with an error at most η percent. As a consequence, T_1, T_2 and T_3 will change by at most $C\eta$ percent. Therefore,

$$-F_f''(\Omega) \geq \frac{1}{2}\, \tilde{a}_1 \left(\frac{\tilde{M}}{12}\right)^{\alpha} \Omega^{-1+\alpha} - C ,$$

where $\tilde{a}_1$ and $\tilde{M}$ differ from the ones in (4.20) by at most η percent. In fact, we only need $\tilde{a}_1 > a_1/2 > 0$.

Therefore, taking $\Omega_0 \ll \Omega_\varepsilon$, we see that $F''(\Omega) < c < 0$ for $\Omega \in (0, \Omega_0)$.

Now, set

$$\delta = \inf_{\Omega \in (0\,, \Omega_c)} |F_y''(\Omega)| > 0 .$$

(This is the only point where we use a computer-assisted result.) From formula (1.6), it follows that hypothesis 1., for ε small enough, implies that

$$|F_f''(\Omega) - F_y''(\Omega)| \leq \frac{1}{10}\, \delta , \qquad \Omega \in [\Omega_0, \Omega_c) ,$$

which concludes the proof of the theorem.

7. The Implementation.

The aim of this section is to provide details about the way algorithms were implemented. The section will be organized as follows:

1. General remarks; in particular, the choice of several heuristic parameters is of special importance for a successful run of the computer proof: we list the approximate values.

2. The second deals with the computer programs, which can divided into two groups.

   a) One is a general package that performs general arithmetic and functional operations on certain general objects. This basic interval arithmetic package is a variation on the one used in [Se1] and [Se2], which in turn is an adaptation of the one developed by D. Rana in [Ra]. It is too long to present here, but we will give enough information about it so that a similar package can be built with little thought. In particular, we will list all function names with a very brief description of each.

   Such packages are quite common already, and probably they will soon be standard.

   b) The other is a package which takes care of the specific functions needed to prove our theorem. It follows very closely the algorithmic presentation in the present paper. We will list all these programs, preceded by a short explanation for each function, which will relate each of them to the corresponding algorithm in the text above.

7.1. General Remarks.

According to the general package, (see below) we can store functions locally using the neighborhoods in function space

$$\mathcal{U}(I_0, \ldots, I_N; C_h, C_g; k),$$

introduced in Section 2 as follows: say $f(t) = \tilde{f}\big((t-x)/r\big)$, where $\tilde{f} \in \mathcal{U}$. Then, our knowledge of f can be stored as a structure variable, consisting of

1. A pointer to an array of intervals: it is used to store the I_j,
2. Two integers: one has value N, the order of the Taylor approximation. The other has value k, the type of the neighborhood,
3. Two doubles, to store C_h and C_g,
4. Two doubles, to store x and r.

By considering arrays of the structures above, we can store our *global* knowledge of functions as a single structure variable, consisting of a pointer to an array of the structure variables above. As a consequence, objects like the Thomas-Fermi function $y(x)$, are represented as a single variable. This gives a special computational meaning to Algorithm 3.20, the main result of Section 3.

We divide our remarks according to the section they are related to.

Section 3.

In Algorithm 3.20, note that the choice of the x_i and r_i is in principle arbitrary, but in practice, it is very important that they are chosen carefully. Main points to take into account are:

1. All runs of parent algorithms should be successful.
2. Error bounds $C_{h,i}$ and $C_{g,i}$ obtained when we run the algorithm are sensitive to our choice of x_i and r_i. The proof Theorem 1.1 is in turn sensitive to these error bounds. In principle, the smaller the r_i the better. It is important that these error bounds are small enough so that we can prove our theorem.
3. The number m is important also: a large m is a consequence of small r_i which will give small error terms for the $C_{\substack{h \\ g},i}$, but will make the computation of I_1 in Section 4 very slow, maybe too slow to prove our theorem in a finite time. On the other hand, a small m will speed up the computation of I_1, but will yield bad bounds for the $C_{\substack{h \\ g}}, i$. Similar considerations hold for the choice of N.

The choice of N is fixed on a trial and error basis. $N \sim 10$ works. About the x_i and r_i note that their choice was made in Algorithms 3.16 and 3.18. They were picked adaptatively inside the program, in the

sense that if during the execution of Algorithm 3.2 (a parent algorithm), one of the error bounds grows outside a pre-specified range, then we make the next r_i a little smaller. And viceversa, if that error goes below a certain range, then we make the next r_i a little bigger. The error we look at in deciding this is $\|p - \tilde{T}p\|/\|p\|$ in Algorithm 3.2, and we wanted it to be within the bounds $[\sim 10^{-17}, \sim 10^{-15}]$. We chose $x_0 \sim 0.008$ and $r_0 \sim 0.0008$. The radii grow as we leave the origin. In carrying out this procedure we made $x_{i+1} \sim x_i + r_i$. This gave enough overlap between intervals to capture the behavior of our functions all over $(x_0 - r_0, x_m + r_m)$. As a result of this method, we obtained $m \sim 800$ and $x_m \sim 300$. The following remark is of mild interest: when choosing the x_i, we instructed the computer to include the point ~ 2.10 with the idea in mind that r_c is close to this number. Since the information given by Algorithm 3.20 is the only one we would like to use when computing F'', the neighborhoods for $y_{\mathrm{TF}}(x)$ around r_c which would have to be computed in Section 5 turn out a little better.

Section 4.

The actual value for $\bar{\Omega}$ is about 0.6956, only $\sim 10^{-3}$ from Ω_c. Thus, Zone II will turn out to be very small. This is unavoidable using our complex-variable methods for Zone II, since with radii larger than that, we cannot exclude the existence of other zeros of $u(z) - \Omega_c^2$ in the vicinity of r_c in the complex plane.

The first heuristic choice we have to make is the numbers a and b as a function of Ω. We describe the choice of a. The choice of b is similar. In the notation of Algorithm 3.20, choose the x_i closest to r_1 such that

$$\sum_{k=1}^{N} \left|I_k^i\right| \leq \delta \left|I_0^i - \Omega^2\right|, \qquad \delta < 1.$$

This is a trivial prerequisite if we want to understand $\left(u(r) - \Omega^2\right)^{-3/2}$ in H^1. The choice of δ is delicate, though: if it is very close to 1, then the fractional power operation will yield bad error bounds. If it is very small, we will be forced to take x_i far away from r_1. This will hurt the error terms when we compute I_2, and it could even make the computation of I_2 not possible with our method. The problem is that we will be forced to solve ODE's at rather large distances (this

is already dangerous), and, even worse, we will have to take fractional powers of Taylor expansions with large radii: this may be impossible if, for instance, the solution of the ODE has zeros within our large radius in the complex plane. A value of about 0.3 for δ works most of the time, but it needs fixing for some values of Ω. We refer the reader to functions `alim()` and `blim()` in the program listings for the specific choice of δ as a function of Ω.

We continue now with the computation of I_1, the most time-consuming procedure. The division into fat intervals is done with intervals of length $\sim 10^{-3}$. This is large enough so we can cover the all of Zone I $= [0,\bar{\Omega}]$ with not so many of those fat intervals, around 1000 of them, and it is also small enough so that the approximation given by (4.2) in terms of the $a_{k,i}$ and $b_{k,i}$ is good enough. As we approach $\bar{\Omega}$, we made the length of these fat intervals smaller, about 10^{-4}. Note that a reduction in the size of the W_i results in having to compute the a_i and b_i more often, but if we are close to Ω_c, the interval $[r_1(\Omega),r_2(\Omega)]$ is very small anyway which require few i, and thin W_i won't hurt.

Next, we have to choose $\tilde{a}$ and $\tilde{b}$, or, which is equivalent, we have to choose i_0 and i_1. We chose them to be $i_0 = 10$, $i_1 = n - 10$. This works. The choice of the Ω^* is the most delicate. What we did, is give the computer an initial interval of length l and let it compute bounds for I_1 as well as I_2 and I_3: if these bounds are good enough so that we can show that $-F'' > 0$ on Ω^*, we tell the computer happily to take another Ω^* inside W_i, and do the same, until all of W_i is covered with tests. If for some interval we cannot produce the bound $-F'' > 0$, then we tell the computer to subdivide that interval into two halves, and try each half recursively until (hopefully) we finish. The process finished, so we conclude $-F'' > 0$ all over W_i. The length of the Ω^* that work is about $5 \cdot 10^{-5}$, degenerating until about 10^{-7} near $\bar{\Omega}$: note that this will generate *a lot* of computations.

In principle, we could have given the computer as a first try all of the interval W_i, even without hope, but let the computer figure out how much finer to go before getting the desired bounds. This is fine from the rigorous point of view, but we would be wasting a lot of precious time asking the computer to make checks that we are confident are going to fail. It is thus important to grind W_i into finer intervals before feeding the computer with this recursive procedure.

Concerning the computation of I_2 and I_3, they are analogous, and

the only thing worth mentioning is our choice of the following heuristic parameters: $M \sim 291$, $L \sim 295$, $x_0 \sim 0.012$ and $\delta \sim 0.0099$. The degrees of Taylor expansions we chose are 10 for I_3 and 20 for I_2. Also, $\Omega_\varepsilon \sim 10^{-2}$, $\bar{\Omega}_2 \sim 0.0932$ and $\bar{\Omega}_3 \sim 0.03469$. See the implementation comments for functions `secder0()` and `secder1()` for more about the $\bar{\Omega}_{2,3}$.

Section 5.

We finally discuss the peculiarities of Zone II. Recall that the diameter of Zone II is about 10^{-3}. Also, it follows from Section 5 that it is rather easy to obtain good bounds for the value for $-F''(\Omega_c)$. The analysis for Zone II therefore looks unnecessarily complicated, since it would follow from the apparently easy, but in practice deep statement that $|F''(x)|$ is bounded by about 1000. Since it is numerically evident that $|F''(x)|$ is bounded by a number much smaller than 1000, maybe one can obtain a good bound for F''' which would make the analysis of Zone II trivial.

The only parameter of real importance is R. Too large R are bad because, as mentioned before, it forces us to carry our fractional power analysis to large distances. Too small R will force us to take small $\tilde{T}$ and therefore small T, and will result of restricting our knowledge of $w(t) = g(t/T)$ to a too small neighborhood of Ω_c, thus being unable to cover all of Zone II. Our choice was $R \sim 0.462$. Once R is chosen, we take $\tilde{T}$ as large as we can, still satisfying (5.3), and we are left with the choice of T only. Of course, we would like to take T as large as possible, but note that the closer we take T to $\tilde{T}$, the closer β will get to 1, which will give us a bad error estimate in (5.4). This negative effect can be neutralized by taking N, which until now was arbitrary, to be big, so that the power β^{2N+2} makes the right hand side of (5.4) very small. We took $T \sim 0.0605$ and $N \sim 26$, but larger N will be even better. The only problem with large N is that it will force us to invert a polynomial of large degree. Even with our sloppy implementation of the inversion procedure, errors and speed are of negligible importance.

It follows from these remarks that the analysis of Section 5 will work on an interval around Ω_c whose length depends basically on how large we can take R. Without a more refined analysis, our choice of R is imposed on us by the apparent complex solutions of $u(z) = \Omega_c^2$ around r_c, and thus is not subject to improvement. In other words, there is a

good reason for taking $\bar{\Omega} \sim 0.6956$ and not smaller: Zone II is given to us by the problem, not by the computer's ability to compute fast or accurately. As a result, with a slower or less accurate computer, which would not be able to compute $-F''$ in Zone I all the way up to $\bar{\Omega}$, we wouldn't be able to prove our theorem in this way. One would need to perform a real variable analysis to a larger Zone II, in a similar way to the analysis of I_2 and I_3 for $\Omega < \Omega_2, \Omega_3$.

Finally, all programs are written in **C**, and were run on several IBM RS600 simultaneously. As explained later, our problem can be naturally split into several independent processes, making it a very appropriate problem to run on different machines at the same time. Execution took about two days for the programs related to the Thomas-Fermi equation, and about 6 hours for the ones involving the actual computation of F''. Executable files averaged 4Mb each.

7.2. General Purpose Package.

The basic variable types in this package are the following:

```
typedef struct {      double dn;
                      double up;}                INTERVL;
typedef struct {      double b;}                 BND;
typedef struct {      int deg;
                      INTERVL *p; }              POLY;
typedef struct {      POLY p;
                      BND center;
                      BND r;
                      BND g;
                      int k;
                      BND h;}                    RSERIES;
typedef struct {      int n;
                      RSERIES *f;}               GRS;
union convert{        reps r;
                      unsigned long int i[2];
                      };

double mtwo = (double) -2;
double mone = (double) -1;
double zero = (double) 0;
```

```
double half = (double) 0.5;
double one = (double) 1;
double two = (double) 2;
double eight = (double) 8;
BND bmone = {(double) -1.};
BND b0ero = {(double) 0};
BND bquarter = {(double) .25};
BND bhalf = {(double) .5};
BND bone = {(double) 1};
BND btwo = {(double) 2};
BND bthree = {(double) 3};
BND bfour = {(double) 4};
INTERVL imfour = {(double) -4,(double) -4};
INTERVL imthree = {(double) -3,(double) -3};
INTERVL imtwo = {(double) -2,(double) -2};
INTERVL imone = {(double) -1,(double) -1};
INTERVL izero = {(double) 0,(double) 0};
INTERVL ihalf = {(double) 0.5,(double) 0.5};
INTERVL imhalf = {(double) -0.5,(double) -0.5};
INTERVL ione = {(double) 1,(double) 1};
INTERVL itwo = {(double) 2,(double) 2};
INTERVL ithree = {(double) 3,(double) 3};
INTERVL ifour = {(double) 4,(double) 4};
INTERVL ifive = {(double) 5,(double) 5};
INTERVL isixteen = {(double) 16,(double) 16};
INTERVL imsix = {(double) -6,(double) -6};
INTERVL ieight = {(double) 8,(double) 8};
INTERVL ifortyeight = {(double) 48,(double) 48};
int n;
double ln2;
INTERVL iln2 = {0.69314718055994484, 0.69314718055994584};
```

The following are the function descriptions.

`up(r)`. Returns a representable strictly larger than `r`.

`dn(r)`. Returns a representable strictly smaller than `r`.

The functions to follow return variable of type `BND`. Variables `a` and `b` are of type `BND`, `x` is of type `INTERVL` and `m` is of type `int`.

`ucvtib(x)`. Returns an upper bound for `x`.

`lcvtib(x)`. Returns a lower bound for `x`.

`cvtdb(d)`. Converts `d` (a `double`) into `BND`.

`cvtintb(m)`. Converts `m` into `BND`.

`uplusb(a,b)`. Returns an upper bound for the sum of `a` and `b`.

`lplusb(a,b)`. Returns a lower bound for the sum of `a` and `b`.

`neg(a)`. Returns $-$`a`.

`absb(a)`. Returns |`a`|.

`minb(a,b)`. Returns the minimum of `a` and `b`.

`maxb(a,b)`. Returns the maximum of `a` and `b`.

`umultb(a,b)`. Returns an upper bound for the product of `a` and `b`.

`lmultb(a,b)`. Returns a lower bound for the product of `a` and `b`.

`uinvb(a)`. Returns an upper bound for the inverse of `a`.

`linvb(a)`. Returns a lower bound for the inverse of `a`.

`udivb(a,b)`. Returns an upper bound for `a/b`.

`ldivb(a,b)`. Returns a lower bound for `a/b`.

`usquareb(b)`. Returns an upper bound for $\mathtt{b}^2$.

`lsquareb(b)`. Returns a lower bound for $\mathtt{b}^2$.

`upowerb(b,m)`. Returns an upper bound for $\mathtt{b}^m$.

`lpowerb(b,m)`. Returns a lower bound for $\mathtt{b}^m$.

The functions to follow return a variable of type `int`.

`eqb(a,b)`. Returns 1 if `a`$=$`b`, 0 otherwise.

`neqb(a,b)`. Returns 0 if `a`$=$`b`, 1 otherwise.

`lsb(a,b)`. Returns 1 if `a` $<$ `b`, 0 otherwise.

`lseqb(a,b)`. Returns 1 if `a` $\leq$ `b`, 0 otherwise.

`grtb(a,b)`. Returns 1 if `a` $>$ `b`, 0 otherwise.

`grteqb(a,b)`. Returns 1 if `a` $\geq$ `b`, 0 otherwise.

The functions to follow return variable of type `INTERVL`, unless stated otherwise. Variables `x` and `y` are of type `INTERVL`, `d` is `double`,

`m`, `i` and `j` are of type `int` and `b` is of type `BND`.

`cvtbi(b)`. Converts `b` into `INTERVL`.

`cvtdi(d)`. Converts `d` into `INTERVL`.

`cvtinti(m)`. Converts `m` into `INTERVL`.

`plus(x,y)`. Returns an interval containing the true set-theoretic sum of `x` and `y`.

`neg(x)`. Returns $-$`x`.

`iabs(x)`. Returns |`x`|.

`uabs(x)`. Returns an upper bound to |`x`|. The function returns a variable of type `BND`.

`labsi(x)`. Returns a lower bound to |`x`|. Returns a variable of type `BND`.

`iequ(x,y)`. Returns 1 if the arguments are exactly the same, 0 otherwise.

`ienlarge(x,b)`. Returns an interval containing all points at distance at most `b` from `x`.

`mult(x,y)`. Returns an interval containing the true set-theoretic product of `x` and `y`.

`divi(x,y)`. Returns an interval containing the true set-theoretic division of `x` by `y`. If $0 \in$ `y`, then we abort the program.

`inv(x)`. Returns an interval containing the true set-theoretic inverse of `y`. If $0 \in$ `y`, then we abort the program.

`square(x)`. Returns an interval containing the true set-theoretic square of `x`.

`power(x,m)`. Returns an interval containing the true set-theoretic power $\mathtt{x}^m$.

`intersect(x,y)`. Returns `x` $\cap$ `y`, which also belongs to $\mathcal{I}$.

`iunion(x,y)`. Returns `x` $\cup_I$ `y` $\in \mathcal{I}$, the smallest interval containing the union of both arguments.

`ration(i,j)`. Returns an interval containing `i/j`.

`iexp(x)`. Returns an interval containing e^x. This can be easily con-

structed using the Taylor expansion for the exponential.

`ilog(x)`. Returns an interval containing log(`x`). This can be easily constructed using the Taylor expansion for the exponential, in the case $x \in [1/2,1)$, and the general case follows trivially after we obtain upper and lower bounds for $\log 2$. These bounds can be obtained heuristically, and then checked using the function `iexp()`. Alternatively, bounds for $\log 2$ are available in the literature, which are better than the ones we could check with `iexp()`; we preferred our way since we simply don't know whether those bounds in the literature are rigorous. This is somewhat wasteful, since `iexp()` is rather conservative (not much). But it did not affect our proof in any noticeable way.

The functions to follow return variables of type `POLY`, unless said otherwise. Arguments starting with `p` are also of type `POLY`, `m` is of type `int`, `x`, `y` and `a` are of type `INTERVL`, and variables starting with `b` are `BND`.

`make_poly(m)`. Returns a `POLY` of degree `m` with zero coefficients.

`polycopy(p)`. Returns a `POLY` identical to `p`.

`coeff(p,m)`. Returns `p.p[m]`, an `INTERVL`.

`coeffmult(p1,p2,m)`. Returns the `m`'th coefficient in the algebraic product (in the interval arithmetic sense) of `p1` and `p2`.

`polysca(p,a)`. Returns bounds for `a`·`p`.

`evalpoly(p,a)`. Returns an `INTERVL` containing the algebraic evaluation of `p` at `a`.

`polynorm(p)`. Returns a `BND`, which is an upper bound for the sum of the absolute value of the coefficients of `p`.

`polyplus(p1,p2)`. Returns bounds for the algebraic sum of the arguments.

`polymult(p1,p2)`. Returns bounds for the algebraic product of the arguments.

`polyscale(p,a)`. Returns bounds for the polynomial in x given by `p`($\mathtt{a} \cdot x$).

`polyder(p)`. Returns bounds for the algebraic derivative of `p`.

`polyinteg(p)`. Returns bounds for the algebraic integral of `p`.

`polycomp(p1,p2)`. Returns bounds for the algebraic composition `p1(p2)`.

`coeffcomp(p1,p2,m)`. Returns bounds for `m`'th coefficient in the algebraic composition `p1(p2)`.

`polyinv(p)`. Returns bounds for the first `p.deg`+1 Taylor coefficients of the functional inverse p^{-1} such that $p \circ (p^{-1}) = \mathrm{Id}$.

The following functions return variables of type `RSERIES`, with same radius, center, order and type as the arguments, unless stated otherwise. Variable names continue with the same type, except that those starting with `s` and `r` are now of type `RSERIES`.

`rs(x,r,i)`. Returns a `RSERIES`, with `.center=x`, `.r=r` and `.p.deg` = `i`. Polynomial coefficients are zero.

`rscopy(r)`. Returns a `RSERIES` identical to `r`.

`ichcoo(a,r)`. Returns bounds for (`r.center` − `a`)/`r.r`.

`geomrs(x,y,b1,m,b2)`. Returns a neighborhood for $1/(xt + y)$ with center at `b1`, radius `b2` and degree `m`.

`rstrunc(s,m)`. Returns a `RSERIES` of degree `m` which contains `s`. It aborts if `m` > `s.p.deg`.

`rsplusc(r,a)`. Returns bounds for `r`+`a`.

`rsca(r,a)`. Returns bounds for `a`·`r`.

`rsplus(r,s)`. Returns bounds for the sum of the arguments. The order and type are the smaller of those of the arguments. It is assumed without check that the center of the arguments are identical.

`rsminus(r,s)`. Returns bounds for `r`−`s`. The order and type are the smaller of those of the arguments. It is assumed without check that the center of the arguments are identical.

`rsmult(r,s)`. Returns bounds for the product of the arguments. The order and type are the smaller of those of the arguments. It is assumed without check that the center of the arguments are identical.

`rseval(r,a)`. Returns an `INTERVL` with bounds for `r`(`a`).

`rsinteg(r)`. Returns a neighborhood for the functions $\int_{\mathtt{r.center}}^{x} f(t)\,dt$

where $f \in$ r, and $|x - \texttt{r.center}| < \texttt{r.r}$. The polynomial order and type of the output is one more than those of the argument.

rsdint(r,a,b). Returns a INTERVL with bounds for $\int_b^a f(t)\,dt$ where $f \in$ r.

rsoverx(s). Returns bounds for $\texttt{s}(x)/(x - \texttt{s.center})$. It is assumed here without check that s(s.center)=0 exactly and the type of s is at least 1.

rslog(x,y,b1,m,b2). Returns a neighborhood for $\log(\texttt{x}t + \texttt{y})$ with center at b1, radius b2 and degree m.

ievders(s,x). Returns an INTERVL containing bounds for the derivative of s at x.

ievnders(s,x,m). Returns an INTERVL containing bounds for the m'th derivative of s at x.

bl1nrs(s). Returns a BND which contains an upper bound for $\|s\|_1$.

rstimesx(s). Returns bounds for the functions $x \cdot \texttt{s}(x)$.

frac22(x,b). Returns a BND with an upper bound for $C_{2.2}(x, b)$, as in Section 2.

frak22(x,b). Returns a BND with an upper bound for $K_{2.2}(x, b)$.

rsmatpower(s,*r). Returns the argument *r, a pointer to an array of RSERIES, containing bounds for all powers of s, from 0 to s.p.deg.

rspower(r,x). Returns bounds for $\texttt{r}^x$.

polypower(p,x). Returns a POLY containing bounds for the Taylor approximation of degree p.deg of $\texttt{p}^x$.

The functions to follow perform operations on variables of type GRS, represented by arguments starting with gs.

Functions

```
grscopy(gs)
grsmult(gs1,gs2)
grseval(gs,x)
grsdereval(gs,x)
grstimesx(gs)
grspower(grs,x)
```

perform the corresponding operations as their RSERIES counterparts on each RSERIES member of their structures.

`grsdint(gs,x,y)`. Returns bounds for the integral from `x` to `y` of all global functions in `gs`. Note here that the role the of `x` and `y` is reversed with respect to `rsdint()`.

`grs(n)`. This function simply returns a `GRS` with `.n` member equal to `n`, and with space allocated for `n+1` variables of type `RSERIES`. Note that the further allocation needed *in* the `POLY` member of `RSERIES` is not done here. This should be done using either `rs()` or `make_poly()` above.

`grsintpt(gs,i)`. Returns a `double`, a heuristic choice for the "middle" point between the centers of the `i`'th and `i+1`'th members of `gs`. If we denote the centers by x_1 and x_2, and corresponding radii by r_1 and r_2, this function returns approximately the number

$$\frac{r_1 \cdot x_2 + r_2 \cdot x_1}{r_1 + r_2} .$$

`grsloc(gs,x)`. Another heuristic function. Returns an integer representing the member of the structure `gs` which best captures the behavior of `gs` near `x`, *i.e.,* the one that minimizes (in a heuristic way), the output of `ichcoo(x, ... )` above.

The functions with names equal to the above followed by an `f` perform the same operations, plus: they destroy the arguments containing pointers by freeing the memory they have allocated.

In addition to these functions, we also have the following, which are of an entirely heuristic nature. They are designed to make the heuristic guesses of p in Algorithms 3.2 and similar. They manipulate polynomials, this time defined simply as arrays of `SIZE+1` variables of type `double` (we took `SIZE=50` in our programs); we will denote such variables here with names starting with `po`. They also use the additional `extern` variable `DEGREE`, smaller than `SIZE` at all times, intended to allow us to vary the effective degree of these polynomials inside the programs. They do not return any variables a values, only as arguments. All operations they perform are floating-point.

`pzer(po)`. Initializes `po` to 0.

`pcopy(po1,po2)`. Copies `po1` into `po2`.

`pnorm(po1)`. Returns a `double` with a floating-point approximation to $\|\texttt{po1}\|_1$.

`psub(po1,po2,po3)`. Puts in `po3` the algebraic difference of `po1` and `po2`.

`pprod(po1,po2,po3)`. Puts in `po3` the algebraic product of `po1` and `po2` truncated to `DEGREE`.

`pinte(p01,po2)`. Puts in `po2` the algebraic integral of `po1`, truncated to `DEGREE`.

`psca(po1,x,po2)`. Puts in `po2` the product of `po1` by the `double x`.

`pscale(po1,x,po2)`. Puts in `po2` the scaled polynomial $\texttt{po1}(\texttt{x} \cdot t)$.

`myprpower(po1,x,po2)`. Takes `po1` to the power `x` and puts the result in `po2`.

Last, but not least, we also need functions that give the *decimal* expansion of rationals bounds for representable numbers. This is required, for instance, to be able to state Lemma 3.21 in the form we did, rather than in a form where the bounds claimed are given in the harder to visualize hexadecimal form. The construction of such functions, while not trivial, is not too hard and we omit the details.

7.3. Aperiodicity Programs.

The following is a brief itemized explanation of the computer programs included at the end of this paper.

Throughout the programs, we will use the external variables

```
extern GRS Y, YRS, U;
extern INTERVL Le, UL, De, UD, C1, W, RC, BC, ALPHA;
extern RSERIES HINF, H0;
```

The variable `ALPHA` will contain consist of bounds for $(\sqrt{73} - 7)/2$ computed once and for all at the beginning of each program. The variables `De` and `UD` correspond to δ and $u(\delta)$ of Section 4, and `Le` and

UL correspond to variables L and $u(L)$ also in Section 4, and they are introduced in functions `h_at_0()` and `hinf()` respectively. The rest will be explained below.

`lipreg()`. Implements Algorithm 3.1, returning the Lipschitz norm if we can show that it is less than 1, and returning 1 otherwise.

`vtffx()`. Implements Algorithm 3.3. Algorithm 3.2, which is needed for the execution of the former, is implemented explicitly inside the function.

`supervtffx()`. Implements Algorithm 3.5. Note that in this function, as well as in `vtffx()`, the values and derivatives of the solution of the ODE are returned as arguments, while the `double` that the function returns as value is an approximation to $\|p - \tilde{T}p\|$, which, as pointed out before, will be used in deciding how much to increase or decrease the next choice of r_i.

`vtffxi()`. Implements Algorithm 3.2. Gives neighborhoods of type 2. This function (and `vtffxi2()` below) returns a neighborhood valid for all centers in the interval `xin`. As a consequence, no `.center` of type `BND` can be naturally specified in the `RSERIES` it returns; we assigned the value 0 (a better choice would be `NaN`). This means that we cannot manipulate the outcome of this particular function with any general-purpose function which would attempt to make use of the structure member `.center`. All such manipulations should be done explicitly taking into account that the centers are contained in `xin`.[1]

`tfypoly()`. See below.

`vtffxi2()`. Implements Algorithm 3.2. Gives neighborhoods of type ∞. The power matching scheme is done in `tfypoly()`

`lip0()`. Implements Algorithm 3.6, returning the Lipschitz norm if we can show that it is less than 1, and returning 1 otherwise.

`vtff0()`. Implements Algorithm 3.8. Again, Algorithm 3.7 is built in.

`y_at_0()`. Implements Algorithm 3.7.

`tfw()`. Implements Algorithm 3.15. In our description of this Algo-

[1] One could attempt to define a new variable type in which centers are of type `INTERVL`. Since this is the only place in which we use this special choice, we decided not to do it in this particular situation.

rithm above, the x_i and r_i are given. From the logical point of view, this is true, but from the computational point of view, the x_i and r_i are produced within `tfw()`.

The successive rigorous bounds we find for w_0 are printed in hexadecimal form as they are obtained. The reason for this is that it takes a long time to run each iteration. In this way, one has rigorous bounds for w_0 even if the function does not finish (due to computer shut down, or impatience on our part).

`tff(w)`. This function implements Algorithm 3.16, where the `INTERVL w` has as endpoints a rigorous upper and lower bound for w_0. As in `tfw()`, the choice of the x_i and r_i is done inside the function. The values y_i and y_i' are returned as a single variable of type `GRS`.

`tfrs(y)`. Implements Algorithm 3.20. The argument `y`, which is of type `GRS`, is the output of `tff(w)`.

`Omega(u)`. The argument `u`, a `GRS` variable, contains bounds for $u_{\mathrm{TF}}(x) = x \cdot y_{\mathrm{TF}}(x)$. The function then returns an interval $[r^{\mathrm{dn}}, r^{\mathrm{dn}}]$ which is guaranteed to contain r_c. Recall that r_c is uniquely defined by the identity $u'(r_c) = 0$. Thus, we first look in a heuristic manner for r^{dn} and r^{up}, and conclude that they are valid bounds after checking that $u'(r^{\mathrm{dn}}) \geq 0$ and $u'(r^{\mathrm{up}}) \leq 0$, which we can easily do using the general purpose interval arithmetic package, namely, function `grsdereval()`.

The heuristic construction of the interval is done via a bisection method, slightly modified so that the interval produced is optimal, in the sense that any representable $r > r^{\mathrm{dn}}$, the bounds we obtain for $u'(r)$ would not be strictly positive (similarly for r^{up}).

`tfprint()`. This is a bookeeping function. It prints the output of `tfw()`, bounds for w_0 and $-b_1$, the output of `tff()` (Algorithm 3.16), `tfrs()` (Algorithm 3.18), together with a `GRS` variable containing u_{TF}, the bounds for r_c produced in `Omega()`, and bounds for Ω_c^2. The bounds for u_{TF}, and Ω_c^2 are easily obtained via the general purpose functions `grstimesx()` and `grseval()`.

The print out is done in hexadecimal form, so that it can be printed

on a file and read rigorously for later use.

`tfread()`. This function simply reads the output of `tfprint()`.

`r0(w)`. Given an `INTERVL w`, this function produces an interval $[a,b]$ containing all solutions $r \leq r_c$ to the equation $u(r) = w^2$, for all values w contained in `w`. The bounds are, first, obtained heuristically using bisection (as in `Omega()`), and seen to be correct by checking that an interval containing $u(b)$ is entirely to the right of (*i.e.* larger than or equal to) $\mathtt{w}^2$, and an interval containing $u(a)$ is entirely to the left of $\mathtt{w}^2$.

`r1(w)`. Same as before, except that the solutions we are looking for are $u(r) = w^2$ for $r \geq r_c$. Both functions produce optimal intervals, in the sense described in `Omega()`. This function only returns a true bound when $\mathtt{w} \geq 10^{-20}$. This is perfectly fine, since it is only invoked for $\mathtt{w} \geq \Omega_\varepsilon$, and we check that $\Omega_\varepsilon > 10^{-20}$ (if fact, $\Omega_\varepsilon \sim 0.01$).

`vtfinf()`. Implements Algorithm 3.13, for representable values of b=`-a0` and R=`t`. As usual, Algorithm 3.12 is implemented inside.

`tfc1()`. We use our bounds for w_0, which `tff()` transforms for bounds for y_{TF}, and, for a representable `ctest` we return 1 if we can guarantee that $b_1 \leq -\mathtt{ctest}$, -1 if $b_1 \geq -\mathtt{ctest}$ and 0 if we cannot guarantee any inequality.

`getc1()`. Organizes `tfc1()` to implement Algorithm 3.19. The bounds for $-b_1$ are stored in the external variable `C1`. As in `tfw()`, instead of returning an interval value for our bounds at the end, this function prints the successive rigorous bounds it obtains in hexadecimal form.

`refiney()`. This function implements Algorithm 3.18, with the extra obvious feature that instead of obtaining y_i^* and y'^*_i alone, it takes care of comparing them with the old bounds we had, given by function `tff()` and stored in `Y`, and takes the intersection of them. This requires the x_i to be the same as before, which poses no problem, of course.

`refine_numbers()`. Similar to `tfprint()`, except that, once bounds for b_1 are computed, it takes care of using them to improve the bounds for y_{TF} before printing them out.

`tfw2(w)`. According to Algorithm 3.19, once new bounds for b_1 are obtained, and the corresponding bounds for y_{TF} are obtained, one can attempt to improve the bounds for w_0. This function takes care of this, by returning 1 if, using the scheme in Algorithm 3.19, we can show that

$w_0 \geq$ `w`, -1 if $w_0 \leq$ `w`, and 0 if we cannot claim any inequality.

`mygetw()`. This function simply organizes the previous one.

`refineY()`. This function implements Algorithm 3.16 again. The difference with `tff()` is only a programming one, since the bounds this function computes are assumed to be refinements of previous ones. Thus, the x_i need not be recomputed.

`rstfu2(x,r,y)`. This function implements a trivial variant of Algorithm 3.2, in the type ∞ case, except that instead of returning bounds for y_{TF} alone (which are returned in the pointer variable `y`), it returns $r \cdot y_{\mathrm{TF}}(r)$ also. As pointed out before, multiplication by r, which is implemented as a general purpose routine `rstimesx()`, is not available here, since `vtffxi2()` returns a `RSERIES` without a meaningful `.center` structure member, needed in `rstimesx()`.

`yinf(t,m)`. This function implements Algorithm 3.12. The `int` variable `m` represents the order of the expansion we want.

`expandY()`. Given the variable `Y` of type `RSERIES` containing bounds for y_{TF} and y'_{TF} at certains points x_i, this function returns another `RSERIES` with the same bounds at the same x_i, plus the trivial bounds $y_{\mathrm{TF}} \in [0,1]$, $y'_{\mathrm{TF}} \in [-2,0]$ at other points x_i, chosen heuristically inside it. This is justified since if at the stage we use this function we already know that $w_0 \leq 2$, which we do because by the time we use this function we would have already run `mygetw()`. Obviously, this bound can be replaced by any other we know to be true by the time we run this function, and it will probably have no effect on the final answer, since the information `expandY()` produces will most probably never be used until `refineY()` has already improved it to a quite sharp bound. After doing this, it also destroys `Y` by freeing the memory allocated to it. Note that this function has a purely administrative role.

The functions below refer to the algorithms presented in Section 4. In the explanation to follow, we use the notation introduced there.

`secder0(w,a)`. Computes I_2 in Section 4, for the thin interval `w` and $a =$ `a`. If `w` $\leq \bar{\Omega}_2$, it simply invokes `secder0_sp()` below. Note that $\bar{\Omega}_2$ is implicitly defined by the first `if` statement in this function.

`secder1(w,a)`. Same as before, but for I_3 this time.

`dermatrix()`. Computes the numbers $a_{k,i}$ and $b_{k,i}$ involved in the

computation of I_1 in Section 4. They are stored in the polynomial part of RSERIES variables.

`secdermat(w,a1,a2,der1,der2,i)`. Uses the numbers $a_{k,i}$ and $b_{k,i}$ (given in `a1`, `a2`, `der1` and `der2` respectively) to compute $\tilde{J}_i$(`w`). The choice of the t_i is made using `grsintpt()`.

`secderdir()`. This function computes J_i directly, *i.e.* computes bounds for the functions f_i in Section 4 involved in the computation of I_1, without use of the numbers $a_{k,i}$ and $b_{k,i}$. This function is intended to compute these f_i for representable arguments.

`super_secderdir()`. This function does the same as the previous, but for interval values of the argument. In other words, for the thin interval under consideration $[z_1, z_2]$, this function uses the previous one to compute the $f_i(z_1)$ and $f_i(z_2)$, and then sets $J_i = f_i(z_1) \cup_I f_i(z_2)$. Recall that this is justified due to the monotonicity of the f_i.

`secder_help()`. This function uses the previous one to compute bounds for I_1. It selects the i_0 and i_1, computes the $\tilde{J}_i$ either directly or using the $a_{k,i}$ and $b_{k,i}$, and adds them up together.

`alim()`. This function selects heuristically the number a (as a function of Ω) involved in the break up of I into the I_i ($i = 1, 2, 3$) in (4.1).

`blim()`. Same as before, but for b.

`secder2()`. This function organizes the previous ones to produce bounds for $-F''$ in a thin interval Ω.

`supersecder2()`. Given an interval Ω, it runs the previous function to check whether $-F''$ is strictly positive. If we can check that it is, it reports a success and returns the bounds. If it is not, it subdivides the interval and tries each half recursively. Note that the fact that this function eventually finishes implies that $-F''$ is strictly positive on the original interval.

`supersupersecderdn(w,vup,dup)`. (At this stage, the reader will probably notice our lack of imagination in picking names for all the functions involved in this proof.) Given a fat interval `w`, and numbers $a_{2,i}$ (stored in `vup`), and $b_{2,i}$ (stored in `dup`), corresponding to the upper endpoint of `w`, this function computes the missing $a_{1,i}$ and $b_{1,i}$ corresponding to the lower endpoint of `w`; then, it breaks `w` into thin subintervals using the heuristic variable `step`, and then invokes the previous function to check that $-F''$ is strictly positive in each thin subinterval.

Once this is done, we know that $-F''$ is strictly positive all over `w`. Before returning, this function replaces the arguments `vup` and `dup` by the values of the $a_{1,i}$ and $b_{1,i}$ corresponding to the lower end of `w`. The reason for this will be explained in the next function.

`secderdn(r,step)`. Given representable `r` and `step`, this function computes the a_i and b_i corresponding to r, constructs the fat interval $w_1 = [\mathtt{r} - \mathtt{step}, \mathtt{r}]$, and gives them to the previous function (note that these are exactly the arguments it needs) to do its job. When `supersupersecderdn()` is finished, it returns to us the a_i and b_i corresponding to the lower end of w_1; then, we construct the new fat interval $w_2 = [\mathtt{r} - 2 \cdot \mathtt{step}, \mathtt{r} - \mathtt{step}]$, and we give it to `supersupersecderdn()` again. Note that the a_i and b_i that we need now are exactly the ones returned to us by `supersupersecderdn()`. And so on. Note that in the construction of the w_i we used expressions of the type $\mathtt{r} - i \cdot \mathtt{step}$; there is no need to make these computations rigorous, as long we make sure that the lower endpoint of each interval is exactly the upper endpoint of the next, which is trivial to arrange.

We do not include any stopping criterion for this function, rather, we instruct it to print in exact hexadecimal form each fat interval on which we can guarantee that $-F''$ is strictly positive. The reason for this is that it takes a very long time to do each fat interval; thus, we prefer to let several computers run (say six of them) on complementary ranges, and stop them as they redundantly start to get into each other's territory.

`supersupersecderup()`. Same as `supersupersecderdn()`, but designed to go up, rather than down.

`secderup()`. "Up" version of `secderdn()`.

`hinf()`. Implements Algorithm 4.7. All sub-algorithms are explicitly implemented inside as needed. `Le` is chosen here.

`h_at_0()`. Implements Algorithm 4.3. Also, sub-algorithms are implemented inside. `De` is chosen inside also.

`printh()`. Runs the two previous functions, and types the output in hexadecimal form for later use.

`tfint1(alpha,x)`. This function computes bounds for

$$\int_1^{1+x} (t-1)^{-1/2} t^\alpha \, dt,$$

for $x < 1$. It does it by Taylor-expanding the integrand around 1.

`tfint2(alpha,a,b)`. Computes rough bounds for

$$\int_a^b (t-1)^{-1/2} t^\alpha \, dt$$

by bounding t^α and integrating $(t-1)^{-1/2}$.

`tfint3(alpha,a,b)`. Computes rough bounds for

$$\int_a^b (t-1)^{-1/2} t^\alpha \, dt$$

by bounding $(t-1)^{-1/2}$ and integrating t^α.

`tfint4(alpha,a,b)`. Computes precise bounds for

$$\int_a^b (t-1)^{-1/2} t^\alpha \, dt$$

by Taylor-expanding the integrand. This function is to be used when we require precision and are willing to give up speed. The two previous ones are intended for a fast, rather inaccurate answer.

`tfintf1(alpha,x)`. This function computes rough bounds for

$$\int_1^{1+x} (t-1)^{-1/2} t^\alpha \, dt,$$

for all `x`. It does it by using `tfint1()` around 1, and using `tfint2()` (or `tfint3()`) in several other small intervals away from 0.

`tfintf2(alpha,x)`. As the previous function, but using `tfint4()` instead for precise, slow bounds.

`secder0_sp(w)`. Computes I_2 when $\Omega_\varepsilon \leq \mathtt{w} \leq \bar{\Omega}_2$, as explained in Section 4. It checks that $\mathtt{w}^2 \leq u(\delta)$: otherwise, it aborts the program. Thus, if the program eventually ends without abortions, we are guaranteed that $\bar{\Omega}_2^2 \leq u(\delta)$.

`secder1_sp()`. Computes I_3 when $\Omega_\varepsilon \leq \Omega \leq \bar{\Omega}_3$. The same comments as `secder0_sp()` apply.

`secder0_speps()`. Computes $T_1(\Omega)$, as in Section 4.

`secder1_speps()`. Computes $T_2(\Omega)$, as in Section 4. As before, it also checks that $\mathtt{w}^2 \leq u(L)$: otherwise, it aborts the program. Thus, if the program eventually ends without abortions, we are guaranteed that $\bar{\Omega}_3^2 \leq u(L)$.

`secder_eps()`. Computes T_3.

The next functions are related to Section 5. We also use the same notation used there.

`rtpoly()`. See below.

`tfwz(t,x)`. Implements Algorithm 5.3. The neighborhood $\mathcal{U}_0$ is computed using `vtffxi2()`, where $R =$ `x`. The value `t` is T in the statement of the algorithm, which we check it is less than or equal to $\tilde{T}$. The power-matching scheme is done in `rtpoly()`.

`sdinv(w0,w)`. As in Algorithm 5.4, computes bounds for $-F''$ in the interval `w0`, using `w`, the output of `tfwz()`.

`super_sdinv(a,ww)`. Given an interval `a` (which will be all of Zone II), and `ww`, the output of `tfwz()`, this function subdivides `a` recursively until it checks, using the previous function, that $-F''$ is strictly positive in each subinterval of `a`. When this function exits, we know that $-F''$ is strictly positive all over `a`.

In order to display how the previous functions can be used to prove Theorem 1.1, we conclude the present discussion with a list of the final programs used in proving our theorem. In doing this, we omit the trivial, but lengthy, statements such as those dealing with variable declarations.

The following obtains -from scratch- bounds for w_0, which are printed in exact hexadecimal form.

```
W.dn = (double) 1;
W.up = (double) 2;
tfw(W,0.0);
```

Note that it looks as if in the previous program we are assuming the apparently trivial bounds $[1,2]$ for w_0 before we start. In fact, we are not, since these initial bounds are used only to make heuristic choices where to look. The only thing to bear in mind, is that the choices we will be using will be in $[1,2]$. Therefore, once we exit the program, we only have to check that there is at least one of those choices for which we were able to conclude that it bounds w_0 from above, and that there is one of those choices that bounds w_0 from below. Once we know this, the final bounds obtained will be true bounds for w_0.

Using the bounds for w_0 obtained before, we can now obtain bounds for y_{TF} and use these to obtain bounds for b_1.

```
W=readivalio();
printivalio(W);
Y=tff(W);
Y=expandY();
printgrsio(Y);
fflush(stdout);
C1.up = (double) 14;
C1.dn = (double) 13;
getc1();
```

In the previous program, note that without the statement `expandY()`, the points at which we have bounds for `Y` may not be very many, and when trying to solve the ODE backwards, starting at the largest x_i stored inside `Y`, we may run into trouble, since, as pointed out in Section 3, we need large x_i to be able to solve the ODE around ∞. Note also that the bounds stored in `Y` are printed out in exact hexadecimal form, since we will be using them in all programs to follow.

Next, we organize the output of the previous program so that it contains, first, the bounds for w_0, next, the bounds for $-b_1$, and then the bounds for y_{TF} contained in `Y`, all in exact hexadecimal form. Then, this output can be used as input for the following program, which will use the new bounds for $-b_1$ to obtain improved bounds for y_{TF}, stored in `Y`. It will also compute the corresponding `YRS` and `U`.

```
W=readivalio();
C1 = readivalio();
Y=readgrsio();
refine_numbers();
```

The next program refines our bounds for w_0, b_1 and y_{TF}, as described in Algorithm 3.19. The output is printed out with same format as usual in exact hexadecimal form.

```
tfread();
mygetw();
refineY();
getc1();
refine_numbers();
```

A comment concerning the previous program. Since the bounds that the previous procedures yield are quite sharp, the computer may have to solve ODE's with initial values close to the critical ones that cause the solutions to vanish, but only very slowly. As a result, when trying to check bounds for w_0 with `mygetw()`, some choices of `wtest` may yield a failure of some of the ODE-solving algorithms, which will cause the previous program to be aborted. The thing to do in this case is to take whatever bounds were successfully obtained by `mygetw()`, use them to replace the old bounds for w_0 written in some file, and restart the previous program without using `mygetw()`. To achieve even greater accuracy, one may also rerun the previous program (after replacing the bounds for w_0 with the new ones) with a different choice of the heuristic parameter `t` in `mygetw()`. These comments extend also to b_1 and `getc1()`, although those occurrences are very unlikely in this case.

Once we are happy with all the bounds for the Thomas-Fermi data, we run the following program, which will write in exact hexadecimal form the neighborhoods for $h(t)$ in Algorithm 4.7, for $h(t)$ in Algorithm 4.3, and bounds for L, $u(L)$, δ and $u(\delta)$.

```
tfread();
printh();
```

The program to follow will check that $-F''$ is strictly positive for $\Omega_\varepsilon \leq \Omega \leq \Omega_c$.

```
tfread();
readh();
i1 = ratpower(BC,1,2);
i1.dn = 0.6956; /*This amounts to setting Zone II=[0.6956,Ωc]*/
res = tfwz(0.0605,0.462);
super_sdinv(i1,res);
/*At this point we know that -F"">0 on Zone II*/
secderdn(i1.dn,1.e-4);
```

This program can be complemented with programs of the type

```
tfread();
readh();
secderup(x,t);
```

or

```
tfread();
readh();
secderdn(x,t);
```

for `double` values of `x` and `t`, which can run on separate computers.

Note that these three last programs will tell us in exact hexadecimal form which fat intervals W are guaranteed to satisfy $-F'' > 0$ on W, but will never stop trying to get W closer and closer to 0. The thing to do is, as long as we see that we have checked all intervals inside $[\sim 10^{-2}, \Omega_c]$, halt the program, set Ω_ε equal to the lower end of the last interval checked, and run the following last program:

```
tfread();
readh();
i1 = readivalio();
b1 = minb(uabs(UL),uabs(UD));
b1 = ldivb(b1,btwo);
printbd(b1);
if(lsb(b1,ucvtib(square(i1)))){
printf("error");
abort();
}
i3 = secder1_speps(i1);
i4 = secder0_speps(i1);
i2 = secder_eps();
i2 = plus(i2,neg(plus(i3,i4)));
i3 = poweri(i1,plus(imone,ALPHA));
i3 = mult(i3,divi(HINF.p.p[1],itwo));
i3 = mult(i3, poweri(divi(cvtbi(HINF.center),cvtinti(12)),ALPHA));
i2 = plus(i2,i3);
if(lseqb(lcvtib(i2),b0ero)){
printf("error");
abort();
}
else printf("PROVED");
```

This program checks that $T_1(\Omega) + T_2(\Omega) + T_3 > 0$ for all $\Omega \in$ `i1`, after checking that $\Omega_\varepsilon =$ `i1.up` satisfies (4.11c) and (4.22). As a result, any $\Omega_\varepsilon \in$ `i1` would finish the proof. In our case, `i1.dn=i1.up`=lower end of the last thin interval for which we successfully run `supersuper-secderdn()`.

8. Appendix: The programs.

```
#include <math.h>
#include <stdio.h>
#include "extern.inc"
INTERVL r0(), r1();
RSERIES yinf(), y_at_0(), vtffxi();



INTERVL
secder0(w,a)                                                    secder0
INTERVL w, a;                                                       11
{
        INTERVL yw0, yw1, x, ith;
        INTERVL secder0_sp(), coo, coo12, sol;
        BND b;
        RSERIES y, u, y2;
        int i;

        coo = U.f[1].p.p[0];
        if(lsb(ucvtib(square(w)),lcvtib(coo)))return(secder0_sp(w));    20

        ith = divi(cvtinti(-3),cvtinti(2));
        x = r0(w);
        yw0 = grseval(YRS,x);
        yw1 = grsdereval(YRS,x);
        y = vtffxi( x, a, yw0, yw1);

        u = rscopy(y);
        for(i = y.p.deg; i>=1; --i)
                u.p.p[i] = plus(mult(y.p.p[i],x),mult(y.p.p[i-1], cvtbi(y.r)));  30
        u.p.p[0] = mult(y.p.p[0] ,x);
        u.g = umultb(y.g,uplusb(ucvtib(x),y.r));
        u.h = umultb(y.h,uplusb(ucvtib(x),y.r));
        u.h = uplusb(u.h,umultb(uabs(y.p.p[y.p.deg]),y.r));
        u.p.p[1] = intersect(u.p.p[1],mult(grsdereval(U,x),cvtbi(u.r)));

        y2 = rs(y.p.deg-1,y.center,y.r);
        for(i=0; i <= y2.p.deg; ++i) y2.p.p[i] = u.p.p[i+1];
        y2.g = u.g, y2.h = u.h;
                                                                    40
        y2 = rspowerf(y2,divi(ith,itwo));
        y2 = rsmultf(y2,y2);
        y2 = rsmultf(y2,y);

        sol = izero;
```

```
        coo = divi(iabs(plus(a,neg(x))),cvtdi(y.r.b));
        if(coo.up > 1) coo.up = (double) 1;
        coo12 = iexp(divi(ilog(coo),imtwo));
        if(coo12.dn < 1) coo12.dn = (double) 1;
        for(i = y2.p.deg; i >= 0; i-- ){                                    50
                sol = mult(sol,coo);
                sol = plus(sol,divi(y2.p.p[i],plus(cvtinti(i),neg(ihalf))));
        }


        sol = mult(sol, coo12);
        b =  uplusb(umultb(y2.g,btwo),udivb(y2.h,
             lplusb(cvtintb(y2.p.deg+1),negb(bhalf))));
        sol = mult(ienlarge(sol,b), cvtdi(y2.r.b));
                                                                            60

        freep(u.p), freep(y2.p);
        return(sol);
}

INTERVL
secder1(w,a)                                                        secder1
INTERVL w, a;
{
        INTERVL yw0, yw1, x, ith;                                           70
        INTERVL secder1_sp(), coo, coo12, sol;
        BND b;
        RSERIES y, u, y2;
        int i;

        coo = U.f[U.n-1].p.p[0];
        if(lsb(ucvtib(square(w)),lcvtib(coo)))return(secder1_sp(w));

        ith = divi(cvtinti(-3),cvtinti(2));
        x = r1(w);                                                          80
        yw0 = grseval(YRS,x);
        yw1 = grsdereval(YRS,x);
        y = vtffxi( x, a, yw0, yw1);

        u = rscopy(y);
        for(i = y.p.deg; i>=1; --i)
                u.p.p[i] = plus(mult(y.p.p[i],x),mult(y.p.p[i-1], cvtbi(y.r)));
        u.p.p[0] = mult(y.p.p[0] ,x);
        u.g = umultb(y.g,uplusb(ucvtib(x),y.r));
        u.h = umultb(y.h,uplusb(ucvtib(x),y.r));                            90
        u.h = uplusb(u.h,umultb(uabs(y.p.p[y.p.deg]),y.r));
        u.p.p[1] = intersect(u.p.p[1],mult(grsdereval(U,x),cvtbi(u.r)));
```

```
        y2 = rs(y.p.deg-1,y.center,y.r);
        for(i=0; i <= y2.p.deg; ++i) y2.p.p[i] = neg(u.p.p[i+1]);
        y2.g = u.g, y2.h = u.h;                                          100

        y2 = rspowerf(y2,divi(ith, itwo));
        y2 = rsmultf(y2, y2);




        y2 = rsmultf(y2,y);
        sol = izero;
        coo = divi(iabs(plus(a,neg(x))),cvtdi(y.r.b));                   110
        if(coo.up > 1) coo.up = (double) 1;
        coo12 = iexp(divi(ilog(coo),imtwo));
        if(coo12.dn < 1) coo12.dn = (double) 1;
        for(i = y2.p.deg; i >= 0; i-- ){
                sol = mult(sol,coo);
                if(i % 2 )
                        sol = plus(sol,divi(y2.p.p[i],plus(cvtinti(-i),ihalf)));
                else
                        sol = plus(sol,divi(y2.p.p[i],plus(cvtinti(i),neg(ihalf))));
        }                                                                120


        sol = mult(sol, coo12);
        b =  uplusb(umultb(y2.g,btwo),udivb(y2.h,
             lplusb(cvtintb(y2.p.deg+1),negb(bhalf))));
        sol = mult(ienlarge(sol,b), cvtdi(y2.r.b));

        freep(u.p), freep(y2.p);
        return(sol);
}                                                                        130
```

dermatrix

```
dermatrix(w, sec, thi)
INTERVL w;
RSERIES *sec, *thi;
{
        INTERVL mw2, ifh;
        INTERVL  t1, t2;
        RSERIES rsw1, rsw;                                               140
        int i, i0, i1;

        ifh = divi(cvtinti(-5),cvtinti(2));
```

```
        mw2 = neg(square(w));

        i0 = grsloc(U,r0(w));
        i1 = grsloc(U,r1(w));

        i = i0+5;
        t1 = cvtdi(grsintpt(U,i));                                        150
        while(i <= i1-5){
                ++i;
                if(i== 60 *(i/60 ))
                        printf("%d points done\n", i);
                fflush(stdout);

                rsw = rscopy(U.f[i]);
                rsw.p.p[0] = plus(rsw.p.p[0],mw2);
                t2 = cvtdi(grsintpt(U,i));
                rsw1 = rspower(rsw,divi(ifh,ifour));                      160
                rsw1 = rsmultf(rsw1,rsw1);
                rsw1 = rsmultf(rsmult(rsw1,YRS.f[i]),rsw1);
                thi->p.p[i] = mult(mult(ithree,w),rsdint(rsw1,t2,t1));
                rsw1 = rsmultf(rsw1,rsw);
                sec->p.p[i] = rsdintf(rsw1,t2,t1);
                t1 = t2;
        }


                                                                          170
}


INTERVL
secdermat(w, a1, a2, der1, der2, i)                          secdermat
INTERVL w;
RSERIES a1, a2;
RSERIES der1, der2;
int i;
{                                                                         180
        INTERVL i1, v1, v2, de1, de2, sol;

        v1 = a1.p.p[i];
        v2 = a2.p.p[i];
        de1 = der1.p.p[i];
        de2 = der2.p.p[i];

        i1 = plus(v1,neg(v2));
        i1 = divi(i1,plus(cvtbi(a1.center),neg(cvtbi(a2.center))));
        i1 = mult(i1,plus(w,neg(cvtbi(a1.center))));                      190
        i1 = plus(i1,v1);
        sol.up = i1.up;
```

```
        i1 = mult(de1,plus(w,neg(cvtbi(a1.center))));
        i1 = plus(i1,v1);
        sol.dn = i1.dn;

        i1 = mult(de2,plus(w,neg(cvtbi(a2.center))));
        i1 = plus(i1,v2);
        sol.dn = maxm(sol.dn,i1.dn);                                    200
        if(sol.up < sol.dn){
                printf("SECDERMAT: negative interval !!!\n");
                printival(sol);
                printf("w:");
                printival(w);
                printf("\ncenter1:");
                printbd(a1.center);
                printf("v1:");
                printival(v1);
                printf("d1:");                                          210
                printival(de1);
                printf("\ncenter2:");
                printbd(a2.center);
                printf("v2:");
                printival(v2);
                printf("d2:");
                printival(de2);
                fflush(stdout);
                abort();
        }                                                               220

        return(sol);
}

INTERVL
secderdir(mw2, t1, t2, i)                                    secderdir
INTERVL  mw2, t1, t2;
int i;
{
        RSERIES rsw;                                                    230

        rsw = rscopy(U.f[i]);
        rsw.p.p[0] = plus(rsw.p.p[0],mw2);
        rsw = rspowerf(rsw,ration(-3,4));
        rsw = rsmultf(rsw,rsw);
        rsw = rsmultf(rsw,rscopy(YRS.f[i]));
        return(rsdintf(rsw,t2,t1));
}

INTERVL                                                              240
```

super_secderdir

```
super_secderdir(mw2, t1,t2,i)
INTERVL t1, t2;
INTERVL  mw2;
int i;
{
        return(iunion(secderdir(cvtdi(mw2.up),t1,t2,i),
            secderdir(cvtdi(mw2.dn),t1,t2,i)));
}
```

250

secder_help

```
INTERVL
secder_help(w, a, b, v1, v2, de1, de2)
RSERIES v1, v2, de1, de2;
INTERVL w, a, b;
{
        INTERVL mw2;
        INTERVL sol, t1, t2;
        int i, i0, i1;

        mw2 = neg(square(w));                                    260

        sol = izero;
        i0 = grsloc(U,a);
        if(U.f[i0].center.b < a.dn)++i0;
        i1 = grsloc(U,b);
        if(U.f[i1].center.b > b.up)--i1;

        t1 = a;
        for(i= i0; i <= i1-1; ++i){
                t2 = cvtdi(grsintpt(U,i));                       270
                if(
                        i == i0 ||
                    v1.p.p[i].up == (double) 0 ||
                    v2.p.p[i].up == (double) 0
                    ){
                        sol = plus(sol, super_secderdir(mw2,t1, t2,i));
                }
                else sol = plus(sol,secdermat(w,v1, v2, de1, de2,i));
                t1 = t2;
        }                                                        280
        sol = plus(sol, super_secderdir(mw2,t1, b,i));

        return(sol);
}
```

alim

```
double
alim(w)
```

```
INTERVL w;
{                                                                    290
        int j, i;
        double rat, diff;
        RSERIES rsw;
        INTERVL x, w2;


        rat = 3.0;
        if(w.dn > 0.136 && w.up < .140) rat = 2.0;
        x = U.f[1].p.p[0];
        if(lsb(ucvtib(square(w)),lcvtib(x))){                        300
                if(De.up == De.dn) return(De.up);
                else printf("De error\n");
                fflush(stdout);
                abort();
        }

        w2 = square(w);

        i = 0;
        while(U.f[i].p.p[0].dn < w2.up){                             310
                ++i;
        }

        diff = 1.0;
        --i;




        while(diff > 0.0){                                           320
                ++i;
                rsw = U.f[i];
                diff = (w2.dn - rsw.p.p[0].dn)/rat;
                for(j=1; j <= rsw.p.deg; ++j)
                        diff += fabs(rsw.p.p[j].up);
        }
        --i;


                                                                     330


        return(grsintpt(U,i));
}


double
```

```
blim(w)                                                          blim
INTERVL w;
{                                                                340
        int j, i;
        double rat, diff;
        RSERIES rsw;
        INTERVL w2;


        w2 = U.f[U.n-1].p.p[0];
        if(lsb(ucvtib(square(w)),lcvtib(w2))){
                if(Le.up == Le.dn) return(Le.up);
                else printf("Le error\n");                       350
                fflush(stdout);
                abort();
        }


        if(w.up < .09){
                w2 = r1(w);
                return(w2.dn-16.0);
        }
                                                                 360
        w2 = square(w);

        i = U.n;
        while(U.f[i].p.p[0].dn < w2.up)--i;

        diff = 1.0;
        ++i;
        rat = 2.1;
        if(w.dn > 0.695) rat = 1.5;
        if(w.up < 0.2388 && w.dn > .1) rat = 1.8;                370
        if(w.up < 0.1668 && w.dn > .1) rat = 1.6;
        /*else if(w.dn > 0.136 && w.up < .140) rat = 1.8;*/
        while(diff > 0.0){
                --i;
                rsw = U.f[i];
                diff = (w2.dn - rsw.p.p[0].dn)/rat;
                for(j=1; j <= rsw.p.deg; ++j)
                        diff += fabs(rsw.p.p[j].up);
        }
                                                                 380


        return(grsintpt(U,i));
}

INTERVL
```

secder2

```
secder2(w, v1, v2, de1, de2)
RSERIES v1, v2, de1, de2;
INTERVL w;
{
        INTERVL sol3, sol1, sol2;
        double i1, i2;

        i1 = alim(w);
        i2 = blim(w);

        sol1 = secder1(w,cvtdi(i2));

        sol2 = secder_help(w,cvtdi(i1),cvtdi(i2), v1, v2, de1, de2);
        sol3 = secder0(w,cvtdi(i1));
        return(plus(plus(sol1,sol2),sol3));
}
```

supersecder2

```
INTERVL
supersecder2(w, v1, v2, de1, de2)
RSERIES v1, v2, de1, de2;
INTERVL w;
{
        INTERVL sol, w1, w2;
        sol = secder2(w, v1, v2, de1, de2);
        if(sol.dn <= (double) 0){
                w1.up = w.up;
                w1.dn = 0.5*(w.up+w.dn);
                w2.up = w1.dn;
                w2.dn = w.dn;
                return(iunion(supersecder2(w1, v1, v2, de1, de2),
                    supersecder2(w2, v1, v2, de1, de2)));
        }
        return(sol);
}
```

supersupersecderdn

```
INTERVL
supersupersecderdn(w, vup, dup)
INTERVL w;
RSERIES *vup, *dup;
{
        INTERVL w1, sol;
        RSERIES v2, de2;
        double step = 5.e-5;

        if(w.dn > 0.672)step = 8.e-6;
        if(w.dn > 0.688)step = 3.e-6;
        if(w.dn > 0.693)step = 1.5e-6;
```

```
        if(w.dn > 0.694)step = 4.e−7;
        if(w.dn > 0.6956)step = 2.e−7;

        v2 = rs(U.n,cvtdb(w.dn),b0ero);
        de2 = rs(U.n,cvtdb(w.dn),b0ero);
        dermatrix(cvtdi(w.dn),&v2, &de2);                                    440
        w1.up = w.up;
        w1.dn = w1.up − step;
        sol = supersecder2(w1, *vup, v2, *dup, de2);
        w1.up = w1.dn;
        w1.dn −= step;
        while(w1.dn > w.dn + step/5.0){
                sol = iunion(sol,supersecder2(w1, *vup, v2, *dup, de2));
                w1.up = w1.dn;
                w1.dn −= step;
        }                                                                    450
        w1.dn = w.dn;
        sol = iunion(sol, supersecder2(w1, *vup, v2, *dup, de2));
        freep(vup−>p), freep(dup−>p);
        *vup = v2;
        *dup = de2;
        return(sol);
}


secderdn(r, step)                                                  secderdn
double r, step;                                                      461
{
        RSERIES v, de;
        INTERVL w;

        v = rs(U.n,cvtdb(r),b0ero);
        de = rs(U.n,cvtdb(r),b0ero);
        dermatrix(cvtdi(r),&v, &de);
        w.dn = r;
        while(w.dn > 0){                                                     470
                w.up = w.dn;
                w.dn −= step;
                printival(supersupersecderdn(w, &v, &de));
        }
}

INTERVL
supersupersecderup(w, vup, dup)                          supersupersecderup
INTERVL w;
RSERIES *vup, *dup;                                                  480
{
        INTERVL w1, sol;
```

```
        RSERIES v2, de2;
        double step = 0.00005;

        if(w.dn > 0.672)step = 8.e-6;
        if(w.dn > 0.688)step = 3.e-6;
        if(w.dn > 0.693)step = 1.5e-6;
        if(w.dn > 0.694)step = 4.e-7;
        if(w.dn > 0.6956)step = 2.e-7;                                    490

        v2 = rs(U.n,cvtdb(w.up),b0ero);
        de2 = rs(U.n,cvtdb(w.up),b0ero);
        dermatrix(cvtdi(w.up),&v2, &de2);
        w1.dn = w.dn;
        w1.up = w1.dn + step;
        sol = supersecder2(w1, *vup, v2, *dup, de2);
        w1.dn = w1.up;
        w1.up += step;
        while(w1.up < w.up - step/5.0){                                   500
                sol = iunion(sol,supersecder2(w1, *vup, v2, *dup, de2));
                w1.dn = w1.up;
                w1.up += step;
        }
        w1.up = w.up;
        sol = iunion(sol, supersecder2(w1, *vup, v2, *dup, de2));
        freep(vup->p), freep(dup->p);
        *vup = v2;
        *dup = de2;
        return(sol);                                                      510
}

secderup(r, step)                                                  secderup
double r, step;
{
        RSERIES v, de;
        INTERVL w;

        v = rs(U.n,cvtdb(r),b0ero);
        de = rs(U.n,cvtdb(r),b0ero);                                      520
        dermatrix(cvtdi(r),&v, &de);
        w.up = r;
        while(w.up > 0){
                w.dn = w.up;
                w.up += step;
                printival(supersupersecderup(w, &v, &de));
        }
}

POLY                                                                      530
```

rtpoly

```
rtpoly(u)
POLY u;
{
        POLY res, res2;
        int i;

        res = make_poly(u.deg-2);
        for(i=0; i <= res.deg; ++i) res.p[i] = neg(u.p[i+2]);
        res = polypowerf(res,ihalf);
        res2 = make_poly(res.deg+1);                                    540
        for(i=0; i <= res.deg; ++i) res2.p[i+1] = res.p[i];
        freep(res);
        return(polyinv(res2));
}
```

tfwz

```
RSERIES
tfwz(t, x)
double t, x;
{
        INTERVL i2, a0l, a4, i1, rmr, a0, a1, a2, a3;                   550
        BND m, tt, h;
        RSERIES u, y, rstfu2(), r, rp;
        int oldeg;

        /**** t = 0.0605, and x = 0.462 will do the job for w=0.6956 ****/


        oldeg = DEGREE;
        DEGREE = SIZE -1;
        r.k = rp.k = 0;                                                 560
        r.h = b0ero;
        r.g = b0ero;
        rp.h = b0ero;
        rp.g = b0ero;

        u = rstfu2(RC,x, &y);
        rmr = plus(RC,cvtbi(negb(u.r)));
        a0l = ienlarge(y.p.p[0],bl1nrsf(rsplusc(y,neg(y.p.p[0]))));
        a0 = cvtbi(bl1nrs(y));
        a2 = divi(ratpower(a0,3,2),ratpower(rmr,1,2));                  570
        a1 = plus(divi(BC,square(RC)),mult(a2,cvtbi(u.r)));

        a3 = mult(a1,mult(ration(3,2),ratpower(a0,1,2)));
        a3 = plus(a3,divi(ratpower(a0,3,2),mult(itwo,rmr)));
        a3 = divi(a3,ratpower(rmr,1,2));



        a4 = divi(ratpower(a0,3,2),rmr);
```

```
a4 = plus(a4, mult(mult(itwo,a1),poweri(a0,ihalf)));
a4 = divi(a4,rmr);
a4 = plus(a4,divi(square(a1),poweri(a0l,ihalf)));
a4 = divi(a4,poweri(rmr,ihalf));
a4 = plus(a4, mult(itwo,divi(square(a0),rmr)));
a4 = mult(a4, ration(3,4));


m = uabs(plus(mult(ifour,a3),mult(a4,plus(RC,cvtbi(u.r)))));

r.p = rtpoly(polyscale(u.p,inv(cvtbi(u.r))));
r.r = cvtdb(t);

tt = labsi(u.p.p[2]);
i1 = neg(divi(u.p.p[2],square(cvtbi(u.r))));
i2 = u.p.p[2];
u.p.p[0] = u.p.p[1] = u.p.p[2] = izero;
h = bl1nrs(u);
tt = lplusb(tt,negb(h));
if(lseqb(tt,b0ero)){
        printf("TFWZ: BAD !!!\n");
        fflush(stdout);
        abort();
}
tt =  lcvtib(ratpower(cvtbi(tt),1,2));

h = udivb(h,lmultb(u.r,u.r));
rp.h = ucvtib(poweri(cvtbi(uplusb(uabs(i1),h)),ihalf));
i2 = plus(iabs(mult(i2,itwo)),neg(iabs(mult(ithree,u.p.p[3]))));
i2 = divi(i2,square(cvtbi(u.r)));
i2 = ienlarge(i2,umultb(udivb(m,cvtintb(6)),upowerb(u.r,2)));
rp.h = umultb(btwo,udivb(rp.h,lcvtib(i2)));

if(lseqb(rp.h,b0ero)){
        printf("TFWZ: radius in the expansion for U
                ' is too large !!!\n");
        fflush(stdout);
        abort();
}

r.h = udivb(cvtdb(t),tt);
if(lseqb(bone,r.h)){
        printf("TFWZ: too large t !!!\n");
        fflush(stdout);
        abort();
}
r.h = udivb(upowerb(r.h,r.p.deg),lplusb(bone,negb(r.h)));

rp.h = umultb(rp.h,r.h);
```

```
        r.h = udivb(rp.h,cvtintb(r.p.deg+1));
        r.h = umultb(r.h, cvtdb(t));                                    630

        rp.p = polyder(r.p);
        rp.p = polyscalef(rp.p,cvtbi(r.r));
        r.p = polyscalef(r.p,cvtbi(r.r));
        r.p.p[0] = RC;
        rp.r = r.r;
        r.center = b0ero;
        rp.center = b0ero;

        r = rsmultf(rp,rspowerf(r,imone));                              640
        DEGREE = oldeg;

        return(r);
}


INTERVL
sdinv(w0, w)                                                        sdinv
INTERVL w0;
RSERIES w;                                                          650
{
        int i;
        INTERVL  a[30], sol2, sol, g;
        BND b;
        double t;
        int m;


        t = w.r.b;
        a[0] = ione;                                                    660
        a[1] = ihalf;
        for(i=0; i <= 29; ++i)
                a[i] = divi(ichoose(cvtinti(2*i),i),power(cvtinti(2),2*i));

        g = plus(BC,neg(square(w0)));
        g = divi(g,square(cvtdi(t)));

        if(grteqb(ucvtib(g),bone)){
                printf("SDINV: Omega value tries to escape domain
                                of convergence !!!. \n");                670
        }

        if(w.p.deg % 2) m = (w.p.deg-1)/2;
        else m = (w.p.deg)/2;

        sol  = mult(w.p.p[2*m],a[m]);
        for(i=m-1; i >= 0; --i)
```

```
                sol = plus(mult(sol,g),mult(w.p.p[2*i],
                    a[i]));
        sol2 = mult(mult(cvtinti(m),w.p.p[2*m]),a[m]);
        for(i=m-1; i >= 1; --i)
                sol2 = plus(mult(sol2,g),
                    mult(mult(cvtinti(i),w.p.p[2*i]),a[i]));

        sol = plus(sol,mult(sol2,neg(mult(itwo,square(divi(w0,cvtbi(w.r)))))));

        b = uabs(inv(ilog(g)));
        if(grteqb(cvtintb(m+1),b)){
                b = umultb(cvtintb(2*(m+1)),ucvtib(divi(BC,square(cvtbi(w.r))))));
                b = umultb(uplusb(uabs(g),b),
                    upowerb(uabs(g),m));
        }
        else{
                b = umultb(btwo,ucvtib(divi(BC,square(cvtbi(w.r)))));
                b = udivb(b,labsi(mult(mult(iexp(ione),g),ilog(g))));
                b = uplusb(upowerb(uabs(g),m+1),b);
        }

        b = umultb(w.h, umultb(uabs(a[m+1]), b));
        return(ienlarge(sol,b));
}
```

super_sdinv

```
super_sdinv(a, ww)
INTERVL a;
RSERIES ww;
{
        INTERVL sol;
        RSERIES  w;

        w = rstrunc(ww,53);
        sol = sdinv(a, w);
        if(sol.dn > (double) 0){
                printf("SUCCESS for ");
                printival(a);
                printivalio(a);
                fflush(stdout);
        }

        else {
                printf("trying...\n");
                fflush(stdout);
                sol.up = a.up;
                sol.dn = .5*(a.dn+a.up);
                super_sdinv(sol,w);
```

```
                a.up = sol.dn;
                super_sdinv(a,w);
        }
}                                                                    730


RSERIES
hinf()                                                               hinf
{
        RSERIES y, yp, ypp,  uP, upp, r, rp, rpp, h, rm2, rma, *rpow;
        RSERIES rw, rm1;
        INTERVL i1;
        BND err, unr;
                                                                     740


        double t;
        int m;
        int i, j, k;
        unsigned size;

        t = U.f[U.n-60].center.b;
        m = 10;
                                                                     750
        Le = cvtdi(295.0);
        UL = grseval(U,Le);

        y = yinf(t, m);
        y.center = cvtdb(t);
        y.r = b0ero;
        ypp = rspower(y, ration(3,2));
        yp = rs(y.p.deg, y.center, y.r);
        uP = rs(y.p.deg, y.center, y.r);
        upp = rs(y.p.deg, y.center, y.r);                            760

        for(i=1; i <= ypp.p.deg; ++i){
                yp.p.p[i] = mult(divi(ypp.p.p[i],plus(ifour,
                    mult(cvtinti(i),ALPHA))),ifour);
                uP.p.p[i] = divi(plus(y.p.p[i],mult(imthree,yp.p.p[i])),
                    imtwo);
                upp.p.p[i] = plus(mult(ypp.p.p[i],itwo),neg(yp.p.p[i]));
        }
        yp.g = umultb(ypp.g, udivb(bfour, lplusb(bfour,lcvtib(
            mult(itwo,ALPHA)))));                                    770
        yp.h = umultb(ypp.h, udivb(bfour, lplusb(bfour,lmultb(cvtintb(
            ypp.p.deg+1),lcvtib(ALPHA)))));

        uP.g = udivb(uplusb(y.g,umultb(yp.g,bthree)),btwo);
        uP.h = udivb(uplusb(y.h,umultb(yp.h,bthree)),btwo);
```

```
upp.g = uplusb(yp.g,umultb(ypp.g,btwo));
upp.h = uplusb(yp.h,umultb(ypp.h,btwo));

if(grtb(bllnrs(uP),bhalf)){                                          780
        printf("HINF: Derivative is not bounded below by 1/2\n ");
        abort();
}

if(grtb(bllnrs(upp),bone)){
        printf("HINF: u is not convex\n");
        abort();
}

r = rs(y.p.deg, y.center, y.r);                                      790

r.p.p[0] = ione;
r.p.p[1] = divi(y.p.p[1],itwo);
r.p.deg = 1;
while (r.p.deg < y.p.deg ){
        rma = rspower(r,neg(ALPHA));
        rpow =  (RSERIES *)calloc(size=r.p.deg+1,sizeof(RSERIES));
        rsmatpower(rma, rpow);
        rw = rs(r.p.deg+1,r.center,r.r);
        for(j=1; j <= r.p.deg+1; ++j){                               800
                err= b0ero;
                for(k=1; k <= j; ++k){
                        if(k <= r.p.deg)
                                i1 = rpow[k].p.p[j-k];
                        else
                                i1 = ione;
                        rw.p.p[j] = plus(mult(i1,y.p.p[k]),rw.p.p[j]);
                        if(k > 1) err = maxb(err, uabs(i1));
                }
                rw.p.p[j] = ienlarge(rw.p.p[j],umultb(err,y.g));     810
        }


        rw.p.p[0] = ione;
        r.p.deg++;
        rm2 = rspower(r, imtwo);
        for(j=0; j <= r.p.deg; ++j){
                r.p.p[r.p.deg] = plus(r.p.p[r.p.deg],
                    mult(rm2.p.p[j], rw.p.p[r.p.deg-j]));
        }                                                            820
        r.p.p[r.p.deg] = divi(r.p.p[r.p.deg], itwo);

        for(j=0; j <= r.p.deg-1; ++j) freep(rpow[j].p);
        freep(rm2.p), freep(rma.p), free((char *)rpow), rpow = NULL;
```

```
        freep(rw.p);
}

rma = rspower(r,neg(ALPHA));
rm2 = rspower(r,imone);
rm2 = rsmultf(rm2,rm2);                                          830
rpow =  (RSERIES *)calloc(size=r.p.deg+1,sizeof(RSERIES));
rsmatpower(rma, rpow);
rw = rs(r.p.deg,r.center,r.r);

rw.p.p[0] = ione;
err = b0ero;
for(j=1; j <= r.p.deg; ++j){
        for(k=r.p.deg-j+1; k <= r.p.deg; ++k)
                rpow[j].h = uplusb(rpow[j].h,uabs(rpow[j].p.p[k]));
        err = maxb(err, rpow[j].h);                              840
        for(k = r.p.deg; k >= j; --k)
                rpow[j].p.p[k] = rpow[j].p.p[k-j];
        for(k = 0; k < j; ++k)
                rpow[j].p.p[k] = izero;
        rw = rsplusf(rw, rsca(rpow[j],y.p.p[j]));
}
rw.h = uplusb(rw.h,umultb(err,y.g));

for(j=2; j <= rw.p.deg; ++j){
        err = b0ero;                                             850
        for(k=2; k <= j; ++k){
                err = maxb(err,uabs(rpow[k].p.p[j]));
        }
        rw.p.p[j] = ienlarge(rw.p.p[j],umultb(y.g,err));
}



r.p.p[0] = izero;
unr = bllnrs(r);                                                 860
if (unr.b >= (double) 1){
        printf("HINF: error no. 1\n");
        abort();
}
unr = lplusb(bone,negb(unr));
unr = uabs(poweri(cvtbi(unr),mult(cvtinti(-r.p.deg-1),ALPHA)));
rw.h = uplusb(rw.h, umultb(y.g, unr));
r.p.p[0] = ione;

rw = rsmult(rw,rm2);                                             870
r.h = rw.h;
freep(rw.p);
```

```
freep(rm2.p), freep(rma.p);
for(j=0; j <= r.p.deg; ++j)freep(rpow[j].p);

rma = rspower(r,neg(ALPHA));
rm2 = rsmult(r,r);
rm2 = rsmultf(rm2,rscopy(r));
rsmatpower(rma, rpow);                                          880
rw = rs(r.p.deg,r.center,r.r);
rpp = rs(r.p.deg,r.center,r.r);

rw.p.p[0] = ione;
err = b0ero;
for(j=1; j <= r.p.deg; ++j){
        rpp.h = rpow[j].h;
        for(k=r.p.deg-j+1; k <= r.p.deg; ++k){
                rpp.h = uplusb(rpp.h,uabs(rpow[j].p.p[k]));
        }                                                       890
        err = maxb(err, rpp.h);

        for(k = r.p.deg; k >= j; --k)
                rpp.p.p[k] = rpow[j].p.p[k-j];
        for(k = 0; k < j; ++k)
                rpp.p.p[k] = izero;
        rw = rsplusf(rw,
             rsca(rpp,uP.p.p[j]));

}                                                               900
rw.h = uplusb(rw.h,umultb(err,uP.g));

for(j=2; j <= rw.p.deg; ++j){
        err = b0ero;
        for(k=2; k <= j; ++k){
                err = maxb(err,uabs(rpow[k].p.p[j-k]));
        }
        rw.p.p[j] = ienlarge(rw.p.p[j],umultb(uP.g,err));
}
                                                                910
r.p.p[0] = izero;
unr = bllnrs(r);
if (unr.b >= (double) 1){
        printf("HINF: error no. 3\n");
        fflush(stdout);
        abort();
}

unr = lplusb(bone,negb(unr));
                                                                920
err = labsi(divi(cvtinti(12),poweri(UL,ihalf)));
if(lsb(lmultb(err,unr),cvtdb(t))){
```

```
            printf("HINF: wrong choice of L\n");
            printf("err = ");
            printbd(err);
            printf("unr = ");
            printbd(unr);
            printf(" UL = ");
            printival(UL);
            printf("M = %e",t);                                          930
            printrs(r);
            abort();
}

unr = uabs(poweri(cvtbi(unr), mult(cvtinti(-r.p.deg-1),ALPHA)));
rw.h = uplusb(rw.h,umultb(unr,uplusb(uP.h, uP.g)));
rw.g = b0ero;
r.p.p[0] = ione;

                                                                         940
rp = rsmultf(rspowerf(rw,imone),rm2);
rp.p.p[0] = ione;
freep(rpp.p), freep(rma.p);

rm1 = rspower(r, imone);
rm2 = rsmult(rm1,rm1);
rm2 = rsmultf(rm2,rm2);

rw = rs(r.p.deg, r.center, r.r);
rw.p.p[0] = ione;                                                        950
err = b0ero;
for(j=1; j <= r.p.deg; ++j){
        for(k=r.p.deg-j+1; k <= r.p.deg; ++k)
                rpow[j].h = uplusb(rpow[j].h,uabs(rpow[j].p.p[k]));
        err = maxb(err,rpow[j].h);
        for(k = r.p.deg; k >= j; --k)
                rpow[j].p.p[k] = rpow[j].p.p[k-j];
        for(k = 0; k < j; ++k)
                rpow[j].p.p[k] = izero;
        rw = rsplusf(rw,                                                 960
             rsca(rpow[j],upp.p.p[j]));
}

rw.h = uplusb(rw.h,umultb(err,upp.g));



for(j=2; j <= rw.p.deg; ++j){
        err = b0ero;
        for(k=2; k <= j; ++k){                                           970
                err = maxb(err,uabs(rpow[k].p.p[j]));
```

```
            }
            rw.p.p[j] = ienlarge(rw.p.p[j],umultb(upp.g,err));
        }


        rw.h = uplusb(rw.h,umultb(unr,uplusb(upp.h, upp.g)));
        rw.g = b0ero;

        rpp = rsmultf(rm2,rw);                                          980
        rpp = rsmultf(rsmult(rp,rp),rpp);
        rpp = rsmultf(rpp, rscopy(rp));
        rpp.p.p[0] = ione;


        rw = rsmult(rp,rm1);
        h = rsplusf(rsca(rw,itwo),
            rsplusf(rscaf(rsmult(rpp,rm1),imthree),rsmult(rw,rw)));

        freep(y.p), freep(yp.p), freep(ypp.p);                          990
        freep(uP.p), freep(upp.p);
        freep(r.p), freep(rp.p), freep(rpp.p);
        freep(rm1.p), freep(rw.p);
        for(j=0; j <= r.p.deg; ++j) freep(rpow[j].p);
        free((char *)rpow), rpow = NULL;


        return(h);
}                                                                       1000


RSERIES
h_at_0()                                                                h_at_0
{
        RSERIES y, yp, ypp,  uP, upp, r, rp, rpp, h, rma, *rpow;
        RSERIES rw, rm1;
        INTERVL sc, i1;
        BND err, unr;
        int i, j, k;                                                    1010
        unsigned size;
        double t;
        int m = 20;

        De = cvtdi(0.0099);
        UD = grseval(U,De);

        i = 0;
        t = 0.012;                                                      1020
```

```
y = y_at_0(t, m);
ypp = rspower(y, ration(3,2));
yp = rs(y.p.deg+1, y.center, y.r);
uP = rs(y.p.deg, y.center, y.r);
yp.p.p[0] = neg(W);
ypp.p.p[0] = ione;

sc = poweri(cvtdi(t),ihalf);
for(i=1; i <= yp.p.deg; ++i)
        yp.p.p[i] = mult(divi(mult(ypp.p.p[i-1],itwo),cvtinti(i)),sc);   1030
yp.g = umultb(ypp.g, btwo);
yp.g = umultb(yp.g, ucvtib(sc));
yp.g = udivb(yp.g,bthree);
yp.h = udivb(ypp.h, ldivb(cvtintb(yp.p.deg+1),btwo));
yp.h = umultb(yp.h, ucvtib(sc));

for(i=2; i <= y.p.deg; ++i)
        uP.p.p[i] = plus(y.p.p[i],mult(yp.p.p[i-2],cvtdi(t)));
uP.g = uplusb(y.g,umultb(yp.g,cvtdb(t)));
for(i=y.p.deg-1; i <= yp.p.deg; ++i)                                     1040
        uP.h = uplusb(uabs(yp.p.p[i]),uP.h);
uP.h = uplusb(y.h,umultb(uplusb(yp.h,uP.h),cvtdb(t)));

if(grtb(bllnrs(uP),bhalf)){
        printf("H_AT_0: Derivative too small !!!\n");
        abort();
}
uP.p.p[0] = ione;

upp = rsca(yp,itwo);                                                     1050
for(i=1; i <= upp.p.deg; ++i)
        upp.p.p[i] = plus(upp.p.p[i], mult(ypp.p.p[i-1],sc));
upp.g = uplusb(upp.g,umultb(ypp.g,ucvtib(sc)));
upp.h = uplusb(upp.h,umultb(ypp.h,ucvtib(sc)));


r = rs(y.p.deg, y.center, y.r);

r.p.p[0] = ione;
r.p.p[2] = neg(y.p.p[2]);                                                1060
r.p.deg = 2;
while (r.p.deg < y.p.deg ){
        rma = rspower(r,ihalf);
        rpow =  (RSERIES *)calloc(size=r.p.deg+1,sizeof(RSERIES));
        for(j=0; j <= r.p.deg; ++j) rpow[j] = rs(r.p.deg,r.center,r.r);
        rsmatpower(rma, rpow);
        rw = rs(r.p.deg+1,r.center,r.r);
        for(j=1; j <= r.p.deg+1; ++j){
                err= b0ero;
```

```
                for(k=1; k <= j; ++k){                                  1070

                        if(k <= r.p.deg)
                                i1 = rpow[k].p.p[j-k];
                        else
                                i1 = ione;

                        rw.p.p[j] = plus(mult(i1,y.p.p[k]),rw.p.p[j]);
                        if(k > 1) err = maxb(err, uabs(i1));
                }
                rw.p.p[j] = ienlarge(rw.p.p[j],umultb(err,y.g));        1080
        }


        rw.p.p[0] = ione;
        r.p.deg++;
        for(j=0; j < r.p.deg; ++j){
                r.p.p[r.p.deg] = plus(r.p.p[r.p.deg],
                    mult(r.p.p[j], rw.p.p[r.p.deg-j]));
        }
        r.p.p[r.p.deg] = neg(r.p.p[r.p.deg]);                           1090

        for(j=0; j <= r.p.deg-1; ++j) freep(rpow[j].p);
        freep(rma.p), free((char *)rpow), rpow = NULL;
        freep(rw.p);
}

rma = rspower(r,ihalf);
rpow = (RSERIES *)calloc(size=r.p.deg+1,sizeof(RSERIES));
for(j=0; j <= r.p.deg; ++j) rpow[j] = rs(r.p.deg,r.center,r.r);
rsmatpower(rma, rpow);                                                  1100
rw = rs(r.p.deg,r.center,r.r);

rw.p.p[0] = ione;
err = b0ero;
for(j=1; j <= r.p.deg; ++j){
        for(k=r.p.deg-j+1; k <= r.p.deg; ++k)
                rpow[j].h = uplusb(rpow[j].h,uabs(rpow[j].p.p[k]));
        err = maxb(err,rpow[j].h);

        for(k = r.p.deg; k >= j; --k)                                   1110
                rpow[j].p.p[k] = rpow[j].p.p[k-j];
        for(k = 0; k < j; ++k)
                rpow[j].p.p[k] = izero;
        rw = rsplusf(rw,rsca(rpow[j],y.p.p[j]));
}
rw.h = uplusb(rw.h,umultb(err,y.g));
for(j=2; j <= rw.p.deg; ++j){
        err = b0ero;
```

```
        for(k=2; k <= j; ++k){
                err = maxb(err,uabs(rpow[k].p.p[j]));
        }
        rw.p.p[j] = ienlarge(rw.p.p[j],umultb(y.g,err));
}

unr = bl1nrs(r);
unr = uabs(poweri(cvtbi(unr),divi(cvtinti(r.p.deg+1),itwo)));
rw.h = uplusb(rw.h, umultb(y.g, unr));

rw = rsmultf(rw,rscopy(r));

r.h = umultb(rw.h,btwo);
freep(rw.p);

freep(rma.p), free((char *)rpow), rpow = NULL;

rma = rspower(r,ihalf);
rpow =  (RSERIES *)calloc(size=r.p.deg+1,sizeof(RSERIES));
for(j=0; j <= r.p.deg; ++j) rpow[j] = rs(r.p.deg,r.center,r.r);
rsmatpower(rma, rpow);
rw = rs(r.p.deg,r.center,r.r);
rpp = rs(r.p.deg,r.center,r.r);

rw.p.p[0] = ione;
err = b0ero;
for(j=1; j <= r.p.deg; ++j){
        rpp.h = rpow[j].h;
        for(k=r.p.deg-j+1; k <= r.p.deg; ++k)
                rpp.h = uplusb(rpp.h,uabs(rpow[j].p.p[k]));
        err = maxb(err,rpp.h);
        for(k = r.p.deg; k >= j; --k)
                rpp.p.p[k] = rpow[j].p.p[k-j];
        for(k = 0; k < j; ++k)
                rpp.p.p[k] = izero;
        rw = rsplusf(rw,
             rsca(rpp,uP.p.p[j]));
}
rw.h = uplusb(umultb(err, uP.g), rw.h);


for(j=2; j <= rw.p.deg; ++j){
        err = b0ero;
        for(k=2; k <= j; ++k){
                err = maxb(err,uabs(rpow[k].p.p[j-k]));
        }
        rw.p.p[j] = ienlarge(rw.p.p[j],umultb(uP.g,err));
}
```

```
unr = bl1nrs(r);
err = umultb(ucvtib(UD),unr);                                    1170
if(grtb(err,cvtdb(t))){
        printf("H_at_0: wrong choice of D\n");
        abort();
}
unr = uabs(poweri(cvtbi(unr), divi(cvtinti(r.p.deg+1),itwo)));
rw.h = uplusb(rw.h,umultb(unr,uplusb(uP.h, uP.g)));
rw.g = b0ero;


rp = rspowerf(rw,imone);                                         1180
rp.p.p[0] = ione;
freep(rpp.p), freep(rma.p);

rm1 = rspower(r, imone);

rw = rs(r.p.deg, r.center, r.r);
rw.p.p[0] = upp.p.p[0];
err = b0ero;
for(j=1; j <= r.p.deg; ++j){
        for(k=r.p.deg-j+1; k <= r.p.deg; ++k)                    1190
                rpow[j].h = uplusb(rpow[j].h,uabs(rpow[j].p.p[k]));
        err = maxb(err,rpow[j].h);
        for(k = r.p.deg; k >= j; --k)
                rpow[j].p.p[k] = rpow[j].p.p[k-j];
        for(k = 0; k < j; ++k)
                rpow[j].p.p[k] = izero;
        rw = rsplusf(rw,rsca(rpow[j],upp.p.p[j]));
}
rw.h = uplusb(rw.h,umultb(err,upp.g));
                                                                 1200

for(j=3; j <= rw.p.deg; ++j){
        err = b0ero;
        for(k=3; k <= j; ++k){
                err = maxb(err,uabs(rpow[k].p.p[j]));
        }
        rw.p.p[j] = ienlarge(rw.p.p[j],umultb(upp.g,err));
}

rw.h = uplusb(rw.h,umultb(unr,uplusb(upp.h,upp.g)));             1210
rw.g = b0ero;

rpp = rsmultf(rsmult(rp,rp),rw);
rpp = rsmultf(rpp, rsca(rp,imone));
rpp.p.p[0] = mult(itwo,W);
```

```
        rw = rsmult(rp,rm1);
        h = rsminusf(rsmult(rw,rw), rw);                                    1220

        h.p.deg = h.p.deg - 2;
        for(i=0; i <= h.p.deg; ++i)h.p.p[i] = divi(h.p.p[i+2],cvtdi(t));
        h.h = udivb(h.h,cvtdb(t));

        h = rsminusf(rsmult(rpp,rm1), h);

        freep(y.p), freep(yp.p), freep(ypp.p);
        freep(uP.p), freep(upp.p);
        freep(r.p), freep(rp.p), freep(rpp.p);                              1230
        freep(rm1.p);
        for(j=0; j <= r.p.deg; ++j) freep(rpow[j].p);
        free((char *)rpow), rpow = NULL;

        return(rstruncf(h,9));
}

void
printh()                                                                printh
{                                                                       1241
        RSERIES h;
        printrsio(hinf());
        h = h_at_0();
        printrsio(h);
        printivalio(Le);
        printivalio(UL);
        printivalio(De);
        printivalio(UD);
}                                                                       1250

void
readh()                                                                 readh
{
        HINF = readrsio();
        H0 = readrsio();
        Le = readivalio();
        UL= readivalio();
        De = readivalio();
        UD = readivalio();                                              1260
}

INTERVL
tfint1(alpha,x)                                                         tfint1
```

```
INTERVL alpha, x;
{
        INTERVL sol;
        int m, i;
        BND err;
                                                                    1270
        m = 100;
        if( x.up >= one || x.dn < zero ){
                printf("TFINT1: error\n");
                fflush(stdout);
                abort();
        }

        sol = izero;
        err = udivb(btwo,cvtintb(2*m+3));
        for(i=m; i >= 1; i--){                                      1280
                sol = mult(plus(sol,inv(plus(cvtinti(i),ihalf))),
                     divi(mult(x,plus(alpha,cvtinti(-i+1))),cvtinti(i)));
                err = udivb(
                     umultb(umultb(uabs(x),err),uplusb(uabs(alpha),
                     cvtintb(i-1))),
                     cvtintb(i));
        }

        sol = mult(plus(sol, itwo), poweri(x,ihalf));
        err = umultb(err,udivb(umultb(uplusb(cvtintb(m),negb(lcvtib(alpha))),1290
             uabs(ratpower(x,3,2))),cvtintb(m+1)));
        return(ienlarge(sol, err));
}


INTERVL
tfint2(alpha, a, b)                                                 tfint2
INTERVL a, b, alpha;
{                                                                   1300
        INTERVL sol;

        sol = poweri(iunion(a,b),alpha);
        sol = mult(sol, mult(itwo,plus(b,neg(a))));
        sol = divi(sol,
             plus(poweri(plus(b,imone),ihalf),poweri(plus(a,imone),ihalf)));

        return(sol);
}
                                                                    1310
INTERVL
tfint3(alpha, a, b)                                                 tfint3
```

```
INTERVL a, b, alpha;
{
        INTERVL sol;

        sol = mult(poweri(plus(iunion(a,b),imone),ihalf),plus(alpha,ione));
        sol = divi(plus(poweri(b,plus(alpha,ione)),
             neg(poweri(a,plus(alpha,ione)))),sol);
        return(sol);                                                    1320
}

INTERVL
tfint4(alpha, a, b)                                                     tfint4
INTERVL a, alpha, b;
{
        RSERIES rw1, rw2;
        BND x, r;

        x.b = (a.dn+b.up)/2.0;                                          1330
        r = maxb(uplusb(x,negb(lcvtib(a))),uplusb(negb(x),ucvtib(b)));
        rw1 = rslinpower(ione,imone,neg(ihalf),8,x,r);
        rw2 = rslinpower(ione,izero,alpha,8,x,r);
        return(rsdintf(rsmultf(rw1,rw2),b,a));
}

INTERVL
tfintf1(alpha,x)                                                        tfintf1
INTERVL alpha, x;
{                                                                       1340
        INTERVL sol, a, b;
        double step;

        if(x.up < 1.8)return(tfint1(alpha,plus(x,imone)));

        b.up = b.dn = 1.8;
        sol = tfint1(alpha, plus(b,imone));

        step = x.up/100.0;
        while(b.up < x.dn){                                             1350
                a = b;
                b.up = b.dn = b.dn + step;
                a = tfint2(alpha,a,b);
                sol = plus(sol,a);
        }

        return(plus(sol,tfint2(alpha,b,x)));
}
INTERVL
tfintf2(alpha,x)                                                        tfintf2
```

```
INTERVL alpha, x;                                                    1361
{
        INTERVL sol, a, b;
        double step;

        if(x.up < 1.8)return(tfint1(alpha,plus(x,imone)));

        b.up = b.dn = 1.8;
        sol = tfint1(alpha, plus(b,imone));
                                                                     1370
        step = x.up/100.0;
        while(b.up < x.dn){
                a = b;
                b.up = b.dn = b.dn + step;
                a = tfint4(alpha,a,b);
                sol = plus(sol,a);
        }

        return(plus(sol,tfint2(alpha,b,x)));
}                                                                    1380
```

secder0_sp

```
INTERVL
secder0_sp(w)
INTERVL w;
{
        INTERVL a, sol, uLwm2, m, upL, ta;
        int i;
                                                                     1390
        m = cvtbi(H0.r);
        upL = grsdereval(U,De);
        uLwm2 = divi(UD,square(w));
        if(uLwm2.dn < one){
                printf("SECDER0_sp: Omega is too large\n");
                abort();
        }
        ta = poweri(m,neg(ihalf));

        sol = izero;                                                 1400

        a = divi(cvtinti(H0.p.deg+1),itwo);
        sol = izero;
        sol = ienlarge(sol,umultb(uabs(mult(w,ta)),
            umultb(H0.h,uabs(tfintf1(a,uLwm2)))));
        for(i=H0.p.deg; i >= 0; --i){
                a = divi(cvtinti(i),itwo);
                if(i > 1) a = tfintf1(a,uLwm2);
                else a = tfintf2(a,uLwm2);
```

```
                                                                    1410
            sol = mult(plus(mult(H0.p.p[i],a),mult(sol,ta)),w);
        }

        sol = plus(divi(mult(imtwo,UD),mult(mult(De,upL),poweri(
            plus(UD,neg(square(w))),ihalf))), mult(sol,itwo));
        return(sol);
}


INTERVL                                                             1420
secder1_sp(w)                                                secder1_sp
INTERVL w;
{
        INTERVL wa, a, sol, a2, uLwm2, m, upL, ta;
        int i;

        m = cvtbi(HINF.center);
        upL = grsdereval(U,Le);
        uLwm2 = divi(UL,square(w));
        if(uLwm2.dn < one){                                         1430
                printf("SECDER1_sp: Omega is too large\n");
                abort();
        }
        a2 = divi(ALPHA,itwo);
        wa = poweri(w,ALPHA);
        ta = mult(wa, poweri(divi(m,cvtinti(12)),ALPHA));


        a = plus(mult(cvtinti(HINF.p.deg+1),a2),imone);
        sol = izero;                                                1440
        sol = ienlarge(sol,
            umultb(uabs(ta),umultb(HINF.h,uabs(tfintf1(a,uLwm2)))));
        for(i=HINF.p.deg; i >= 1; --i){
                a = plus(mult(cvtinti(i),a2),imone);
                if( i > 1) a = tfintf1(a,uLwm2);
                else a = tfintf2(a,uLwm2);
                sol = mult(plus(sol,mult(HINF.p.p[i],a)),ta);
        }
        sol = divi(divi(sol,itwo),w);
        sol = plus(divi(mult(itwo,UL),mult(mult(Le,upL),poweri(     1450
            plus(UL,neg(square(w))),ihalf))), sol);
        return(sol);
}


INTERVL
```

secder0_speps(w) secder0_speps
INTERVL w;
{
INTERVL r, sol;
int i;

r = poweri(divi(UD,cvtbi(H0.r)),ihalf);
sol = cvtbi(H0.h);
for(i=H0.p.deg; i >= 0; --i){
sol = mult(sol,r);
if(H0.p.p[i].dn < (double) 0)
sol = plus(sol,iabs(H0.p.p[i]));
}
sol = mult(sol,poweri(UD,ihalf));
sol = mult(sol,cvtinti(4));
sol = plus(divi(mult(itwo,UD),mult(mult(De,grsdereval(U,De)),poweri(
plus(UD,neg(square(w))),ihalf))),sol);
return(sol);
}

INTERVL
secder1_speps(w) secder1_speps
INTERVL w;
{
INTERVL r, sol, i1;
int i;

r = mult(poweri(UL,ihalf),divi(cvtbi(HINF.center),cvtinti(12)));
r = poweri(r,ALPHA);

sol = mult(cvtbi(HINF.h),r);
if(HINF.p.p[2].dn < (double) 0){
printf("SECDER1_SPEPS: second term negative.\n");
abort();
}
for(i=HINF.p.deg; i >= 2; --i){
if(HINF.p.p[i].dn < (double) 0)
sol = plus(sol,iabs(HINF.p.p[i]));
sol = mult(sol,r);
}
sol = mult(sol,r);
sol = divi(sol, poweri(UL,ihalf));

i1 = mult(Le,iabs(grsdereval(U,Le)));
i1 = mult(i1,poweri(plus(UL,neg(square(w))),ihalf));
sol = plus(divi(mult(itwo,UL), i1), sol);

```
        return(sol);
}
                                                                    1510

INTERVL
secder_eps()                                                  secder_eps
{
        GRS yw;

        yw = grstimesx(grstimesx(grstimesx(YRS)));
        yw = grspowerf(yw,neg(ihalf));
        return(grsdintf(yw,De,Le));
}                                                                   1520

#include <math.h>
#include <stdio.h>

BND
lipreg(u0,u1,x0,r,a)                                            lipreg
INTERVL u0, u1, x0;
BND r, a;
{
        BND g1, sol, g0, c1, c2;                                    1530

        g0 = umultb(r,uabs(u1));
        g1 = udivb(g0,lcvtib(u0));
        g0 = udivb(uplusb(g0,a),lcvtib(u0));
        c1 = umultb(frak22(neg(ihalf),udivb(r,lcvtib(x0))),ucvtib(ratpower(x0,
             -1,2)));
        c2 = umultb(ucvtib(ratpower(u0,1,2)),
             frac22(ration(3,2),g0));

        c1 = udivb(umultb(usquareb(r),c1),btwo);                    1540
        sol = umultb(c1,c2);
        c2 = umultb(ucvtib(ratpower(u0,3,2)),
             frak22(ration(3,2),g1));

        c1 = umultb(c1, c2);
        c2 = lmultb(a,lplusb(bone,negb(sol)));
        if (lseqb(c1,c2)) return(sol);
        else{
                return(bone);
        }                                                           1550
}


BND
```

lip0

```
lip0(w,r,a)
INTERVL w;
BND r, a;
{
        BND g1, sol, g0, c1, c2;                                        1560

        g1 = umultb(r,uabs(w));
        g0 = uplusb(g1,a);
        sol = frac22(ration(3,2),g0);
        c2 = frak22(ration(3,2),g1);
        c1 = udivb(umultb(bfour,ucvtib(ratpower(cvtbi(r),3,2))),cvtintb(3));

        sol = umultb(sol, c1);

        c1 = umultb(c1,c2);                                             1570
        c2 = lmultb(a,lplusb(bone,negb(sol)));

        if (lseqb(c1,c2)) return(sol);
        else{
                return(bone);
        }
}


                                                                        1580
```

vtffx

```
double
vtffx(xin, xout, u0, u1,y0, y1)
double xin, u0, u1;
INTERVL xout, *y0, *y1;
{
        RSERIES y, yp, ypp, rsw;
        INTERVL ithreehalfs, r;
        BND bn, eps;
        int i;                                                          1590
        double p[SIZE], pw[SIZE];

        pzer(p), pzer(pw);
        p[0] = u0, p[1] = u1*fabs(.5*(xout.dn+xout.up)-xin);

        pw[0] = xin;
        pw[1] = fabs(.5*(xout.up+xout.dn)-xin);
        myprpower(pw, -0.5, pw);

        for(i=1; i <= DEGREE; ++i){                                     1600
                myprpower(p,1.5,p);
                pprod(p,pw,p);
                pinte(p,p);
                pinte(p,p);
```

```
                psca(p,(.5*(xout.up+xout.dn)-xin)*(.5*(xout.up+xout.dn)-xin),p);
                p[0] = u0, p[1] = u1*fabs(.5*(xout.up+xout.dn)-xin);
        }

        r = iabs(plus(xout,cvtdi(-xin)));
                                                                        1610
        y = rs(DEGREE,cvtdb(xin),cvtdb(r.up));
        for(i=2; i<= DEGREE; ++i) y.p.p[i] = cvtdi(p[i]);

        eps = bl1nrs(y);
        bn = lipreg(cvtdi(u0), cvtdi(u1), cvtdi(xin), ucvtib(r), eps);
        while(grteqb(bn,bone)){
                eps = maxb(eps, cvtdb(eps.b * 1.1));
                bn = lipreg(cvtdi(u0), cvtdi(u1), cvtdi(xin), ucvtib(r), eps);
        }
                                                                        1620
        y.p.p[0] = cvtdi(u0);
        y.p.p[1] = mult(cvtdi(r.up),cvtdi(u1));

        rsw = rs(DEGREE,cvtdb(xin),cvtdb(r.up));
        rsw.p.p[0] = cvtdi(xin);
        rsw.p.p[1] = cvtdi(r.up);
        rsw = rspowerf(rsw,neg(ihalf));

        ithreehalfs = divi(ithree,itwo);
        yp = rspower(y,ithreehalfs);                                    1630
        ypp = rsintegf(rsintegf(rsmultf(yp,rscopy(rsw))));

        yp = rscopy(y);
        yp.p.p[0] = izero;
        yp.p.p[1] = izero;
        eps = bl1nrsf(rsminusf(yp,ypp));

        y.g = udivb(eps,lplusb(bone, negb(bn)));
        y.k = 2;
        rsw.k = 2;                                                      1640
        *y0 = rseval(y,xout);

        *y1 = rsevalf(rsintegf(rsmultf(rspower(y,ithreehalfs),rsw)), xout);


        *y1 = plus(*y1,cvtdi(u1));
        freep(y.p);
        return(eps.b);
}
                                                                        1650
```

double

supervtffx(xin, xout, u0, u1,y0, y1) supervtffx

```
double xin;
INTERVL xout, u0, u1;
INTERVL *y0, *y1;
{
        INTERVL yw0, yw1;
        double eps;
                                                                    1660
        if(xin <= xout.dn){
                eps = vtffx(xin, xout, u0.up, u1.up, &yw0, &yw1);
                y0->up = yw0.up, y1->up = yw1.up;
                eps += vtffx(xin, xout, u0.dn, u1.dn, &yw0, &yw1);
                y0->dn = yw0.dn, y1->dn = yw1.dn;
                return(eps);
        }
        else if(xin >= xout.up){
                eps = vtffx(xin, xout, u0.dn, u1.up, &yw0, &yw1);
                y0->dn = yw0.dn, y1->up = yw1.up;                   1670
                eps += vtffx(xin, xout, u0.up, u1.dn, &yw0, &yw1);
                y0->up = yw0.up, y1->dn = yw1.dn;
                return(eps);
        }
        else{
                printf("SUPERVTFFX: case not considered\n\n\n");
                fflush(stdout);
                abort();
        }
}                                                                   1680

RSERIES
vtffxi(xin, xout, u0, u1)                                          vtffxi
INTERVL xin, xout, u0, u1;
{
        RSERIES y, yp, ypp, rsw;
        INTERVL ithreehalfs, r;
        BND bn, eps;
        int i;
        double p[SIZE], pw[SIZE];                                   1690

        pzer(p), pzer(pw);
        p[0] = .5*(u0.dn+u0.up);
        p[1] = .5*(u1.up+u1.dn)*fabs(.5*(xout.up+xout.dn)
             -0.5*(xin.up+xin.dn));

        pw[0] = 0.5*(xin.up+xin.dn);
        pw[1] = fabs(.5*(xout.up+xout.dn)-0.5*(xin.up+xin.dn));
        myprpower(pw, -0.5, pw);
                                                                    1700
        for(i=1; i <= DEGREE; ++i){
                myprpower(p,1.5,p);
```

```
                pprod(p,pw,p);
                pinte(p,p);
                pinte(p,p);
                psca(p,(.5*(xout.up+xout.dn)-0.5*(xin.up+xin.dn))
                    *(.5*(xout.up+xout.dn)-0.5*(xin.up+xin.dn)),p);
                p[0] = .5*(u0.dn+u0.up);
                p[1] = .5*(u1.up+u1.dn)*fabs(.5*(xout.up+xout.dn)
                    -0.5*(xin.up+xin.dn));                                  1710
        }

        r = iabs(plus(xout,neg(xin)));

        y = rs(DEGREE,b0ero,cvtdb(r.up));
        y.p.p[0] = u0;
        y.p.p[1] = mult(cvtdi(r.up),u1);
        for(i=2; i<= DEGREE; ++i) y.p.p[i] = cvtdi(p[i]);

        rsw = rs(DEGREE,b0ero,cvtdb(r.up));                                 1720
        rsw.p.p[0] = xin;
        rsw.p.p[1] = cvtdi(r.up);
        rsw = rspowerf(rsw,neg(ihalf));

        ithreehalfs = divi(ithree,itwo);
        yp = rspower(y,ithreehalfs);
        ypp = rsintegf(rsintegf(rsmultf(yp,rsw)));

        yp = rscopy(y);
        yp.p.p[0] = izero;                                                  1730
        yp.p.p[1] = izero;

        eps = bl1nrs(yp);
        bn = lipreg(u0, u1, xin, ucvtib(r), eps);
        while(grteqb(bn,bone)){
                eps = maxb(eps,cvtdb(eps.b * 1.1));
                bn = lipreg(u0, u1, xin, ucvtib(r), eps);
        }

        eps = bl1nrsf(rsminusf(yp,ypp));                                    1740
        y.g = udivb(eps,lplusb(bone, negb(bn)));
        y.k = 2;
        return(y);
}
POLY
tfypoly(u0, u1, r, i)                                                   tfypoly
INTERVL u0, u1, r;
int i;
{
        POLY poly, res;                                                     1750
        POLY rsw;
```

```
        int j, k;
        poly = make_poly(i);
        poly.p[0] = u0, poly.p[1] = u1;

        rsw = make_poly(i);
        rsw.p[0] = r;
        rsw.p[1] = ione;
        rsw = polypowerf(rsw,neg(ihalf));
                                                                    1760
        for(k=2; k <= i; ++k){
                res = make_poly(k-1);
                for(j=0; j<=k-2; ++j) res.p[j] = poly.p[j];
                res = polypowerf(res, ration(3,2));
                res.p[k-2] = coeffmult(res,rsw,k-2);
                poly.p[k] = divi(res.p[k-2], cvtinti(k*(k-1)));
                freep(res);
        }

        freep(rsw);                                                 1770
        return(poly);
}

RSERIES
vtffxi2(xin, xout, u0, u1)                                          vtffxi2
INTERVL xin, xout, u0, u1;
{
        RSERIES y, yp, ypp, rsw;
        INTERVL ithreehalfs, r;
        BND bn, eps;                                                1780
        int i;
        POLY poly;


        r = iabs(plus(xout,neg(xin)));
        poly = polyscalef(tfypoly(u0, u1, xin, DEGREE), cvtdi(r.up));

        y = rs(DEGREE,b0ero,cvtdb(r.up));
        y.p.p[0] = u0;
        y.p.p[1] = mult(cvtdi(r.up),u1);                            1790
        for(i=2; i<= DEGREE; ++i)
                y.p.p[i] = cvtdi(.5*(poly.p[i].up+poly.p[i].dn));

        rsw = rs(DEGREE,b0ero,cvtdb(r.up));
        rsw.p.p[0] = xin;
        rsw.p.p[1] = cvtdi(r.up);
        rsw = rspowerf(rsw,imhalf);

        ithreehalfs = divi(ithree,itwo);
        yp = rspower(y,ithreehalfs);                                1800
```

```
        ypp = rsintegf(rsintegf(rsmultf(yp,rsw)));

        yp = rscopy(y);
        yp.p.p[0] = izero;
        yp.p.p[1] = izero;

        eps = bl1nrs(yp);
        bn = lipreg(u0, u1, xin, ucvtib(r), eps);
        while(grteqb(bn,bone)){
                eps = maxb(eps,cvtdb(eps.b * 1.1));                    1810
                bn = lipreg(u0, u1, xin, ucvtib(r), eps);
        }

        eps = bl1nrsf(rsminusf(yp,ypp));
        y.h = udivb(eps,lplusb(bone, negb(bn)));
        for(i=2; i<= DEGREE; ++i)
                y.p.p[i] = iunion(y.p.p[i], poly.p[i]);

        return(y);
}                                                                      1820

void
vtff0(w,t,y0, y1)                                                      vtff0
double w;
BND t;
INTERVL *y0, *y1;
{
        RSERIES y, yp, ypp;
        INTERVL sc;
        INTERVL ithreehalfs;                                           1830
        BND eps, b0;
        double p[SIZE];
        int i, j;

        pzer(p);
        p[0] = 1.0, p[2] = -w;
        for(i=0; i <= DEGREE; ++i){
                myprpower(p,1.5,p);
                for(j=DEGREE; j >= 3; --j)
                        p[j] = 4.0*p[j-3]/(j*(j-2));                   1840
                p[0] = 1.0, p[1]=zero, p[2]= -w;
        }
        pscale(p,sqrt(t.b),p);
        y = rs(DEGREE,b0ero,t);
        for(i=3; i<= DEGREE; ++i) y.p.p[i] = cvtdi(p[i]);
        b0 = bl1nrs(y);
        y.p.p[0] = ione;
        y.p.p[2] = mult(cvtdi(-w),cvtbi(t));
```

```
        ithreehalfs = divi(ithree,itwo);                                    1850
        yp = rspower(y,ithreehalfs);
        ypp = rs(DEGREE+3, b0ero, t);
        for (i=0; i <= DEGREE; ++i) ypp.p.p[i+3]= divi(yp.p.p[i],
            divi(cvtinti((i+1)*(i+3)),cvtinti(4)));
        ypp.h = umultb(yp.h,
            udivb(cvtintb(4),cvtintb((DEGREE+2)*(DEGREE+4))));
        sc = iexp(mult(divi(cvtinti(3),cvtinti(2)),ilog(cvtbi(t))));
        ypp = rscaf(ypp, sc);
        ypp.p.p[0] = y.p.p[0];
        ypp.p.p[1] = y.p.p[1];                                              1860
        ypp.p.p[2] = y.p.p[2];

        eps = btwo;
        while(grteqb(eps,bone)){
                eps = lip0(cvtdi(w),t,b0);
                b0 = maxb(b0,cvtdb(b0.b * 1.01));
        }
        b0 = eps;
        eps = bl1nrsf(rsminus(y,ypp));
        y.g = umultb(eps, lplusb(bone,negb(b0)));                           1870
        *y0 = izero, *y1 = izero;
        for(i=0; i<=y.p.deg; ++i){
                *y0 = plus(*y0,y.p.p[i]);
                *y1 = plus(*y1,divi(yp.p.p[i],divi(cvtinti(i+1),itwo)));
        }
        *y0 = ienlarge(*y0,uplusb(y.g,y.h));
        ypp = rspower(y,ithreehalfs);
        sc = iexp(mult(ihalf,ilog(cvtbi(t))));

        ypp.g = udivb(ypp.g,btwo);                                          1880
        ypp.h = udivb(ypp.h,cvtintb(2*(ypp.p.deg+1)));
        *y1 = mult(ienlarge(*y1,uplusb(ypp.g,ypp.h)), sc);
        *y1 = plus(*y1,neg(cvtdi(w)));
        freep(y.p), freep(yp.p), freep(ypp.p);
}


RSERIES
y_at_0(tb,m)                                                        y_at_0
double tb;                                                          1891
int m;
{
        RSERIES y, yp, ypp;
        INTERVL sc;
        INTERVL ithreehalfs;
        BND t, eps, b0;
        double w, p[SIZE];
```

```
        int olddeg, i, j;
                                                                    1900
        t = cvtdb(tb);
        sc = iexp(mult(divi(ithree,itwo ),ilog(cvtbi(t))));
        olddeg = DEGREE;
        DEGREE = m;
        if (m >= SIZE){
                printf("Y_AT_0: DEGREE %d is not possible\n",m);
                abort();
                fflush(stdout);
        }
                                                                    1910
        pzer(p);
        w = .5*(W.up+W.dn);
        p[0] = 1.0, p[2] = -w*tb;
        for(i=0; i <= DEGREE; ++i){
                myprpower(p,1.5,p);
                for(j=DEGREE; j >= 3; --j)
                        p[j] = 2.0*(sc.up+sc.dn)*p[j-3]/(j*(j-2));
                p[0] = 1.0, p[1]=zero, p[2]= -w*tb;
        }
        y = rs(DEGREE,b0ero,t);                                     1920
        for(i=3; i<= DEGREE; ++i) y.p.p[i] = cvtdi(p[i]);
        b0 = bl1nrs(y);
        y.p.p[0] = ione;
        y.p.p[2] = neg(mult(W,cvtbi(t)));

        ithreehalfs = divi(ithree,itwo);
        yp = rspower(y,ithreehalfs);
        ypp = rs(DEGREE+3, b0ero, t);
        for (i=0; i <= DEGREE; ++i) ypp.p.p[i+3]= divi(yp.p.p[i],
            divi(cvtinti((i+1)*(i+3)),ifour));                      1930
        ypp.h = umultb(yp.h,udivb(cvtintb(4),cvtintb
                ((DEGREE+2)*(DEGREE+4))));
        ypp = rscaf(ypp, sc);

        yp = rscopy(y);
        ypp.p.p[0] = y.p.p[0] = izero;
        ypp.p.p[1] = y.p.p[1] = izero;
        ypp.p.p[2] = y.p.p[2] = izero;

        eps = bl1nrsf(rsminusf(y,ypp));                             1940
        while(grteqb(lip0(W,t,b0),bone)) b0 = maxb(b0, cvtdb(b0.b*1.1));
        yp.g = udivb(eps, dengeob(lip0(W,t,b0)));
        DEGREE = olddeg;
        return(yp);
}
```

```
INTERVL
tfw(w, tol)                                                          tfw
INTERVL w;                                                           1951
double tol;
{
        BND b;
        INTERVL i1, i2, i3, i4, r;
        double eps, wtest, x;

        b.b = 0.008;
        while(w.up-w.dn>tol){
                printf("\n\nBOUNDS FOR w:\n");                       1960
                printival(w);
                printivalio(w);
                fflush(stdout);

                wtest = .5*(w.up+w.dn);
                vtff0(wtest, b, &i1, &i2);
                r.up = b.b;
                x = 0.0008;
                while(i2.dn<= 0 && r.up < 250.0){
                        r.dn = r.up, r.up = r.dn+x;                  1970
                        if(r.dn < 2.104025275
                            && r.up > 2.104025275)
                                r.up = 2.104025275;
                        eps = supervtffx(r.dn, cvtdi(r.up),
                            i1, i2, &i1, &i2);

                        i3 = divi(cvtinti(5),cvtinti(2));
                        i3 = iexp(mult(i3,ilog(i1)));
                        i3 = divi(i3,iexp(mult(ihalf,ilog(cvtdi(r.up)))));
                        i4 = square(i2);                             1980
                        i3 = mult(itwo,i3);
                        if(i3.up <= i4.dn) i2.dn = 1.0;

                        if(eps    > 5.e-17) x *= 0.7;
                        if(eps    > 5.e-16) x *= 0.5;
                        if(eps    < 1.e-17) x *= 1.2;
                        x = minm(x,0.3);
                        if(r.up > 10.0) x = minm(x,0.5);

                }                                                    1990
                if(r.up >= 250.0){
                        return(w);
                }
                else if(i3.up <= i4.dn) w.up = wtest;
                else w.dn = wtest;
        }
```

```
        return(w);
}
```

2000

```
GRS
tff(w)                                                          tff
INTERVL w;
{
        GRS y;
        BND b;
        unsigned size;
        INTERVL i1[1000], i2[1000], i3, i4 , i5[1000], i6[1000];
        double eps, x, r[1000];                                 2010
        int count = 0;
        int count2 = 0;

        b.b = 0.008;
        vtff0(w.up, b, &i1[0], &i2[0]);
        vtff0(w.dn, b, &i5[0], &i6[0]);
        r[count] = b.b;
        x = 0.0008;
        i3.up = 1.0, i4.dn = zero;
        while(i3.up > i4.dn || i6[count].dn < 0){               2020
                if(i3.up <= i4.dn || i6[count].dn >= 0) ++count2;
                r[count+1] =  r[count]+x;
                if(r[count] < 2.10402528 && r[count+1] > 2.10402528)
                        r[count+1]=maxm(r[count],2.10402528);
                if(i3.up > i4.dn )
                        eps = supervtffx(r[count], cvtdi(r[count+1]),
                            i1[count], i2[count], &i1[count+1], &i2[count+1]);
                if(i6[count].dn < 0)
                        eps += supervtffx(r[count], cvtdi(r[count+1]),
                            i5[count], i6[count], &i5[count+1], &i6[count+1]);2030

                ++count;
                if(i3.up > i4.dn ){
                        i3 = divi(cvtinti(5),cvtinti(2));
                        i3 = iexp(mult(i3,ilog(i1[count])));
                        i3 = divi(i3,iexp(mult(ihalf,ilog(cvtdi(r[count])))));
                        i4 = square(i2[count]);
                        i3 = mult(itwo,i3);
                }
                                                                2040
                if(eps   > 1.e-16) x *= 0.7;
                if(eps   > 1.e-15) x *= 0.5;
                if(eps   < 2.e-17) x *= 1.2;
                x = minm(x,4.0);
                if(r[count] > 200.0) x = minm(x,1.0);
```

```
                if(r[count] < 10.0) x = minm(x,0.05);
                if(r[count] > 250.0) x = minm(x,0.3);
        }

        count -= count2;                                                    2050
        y.n = count+1;
        y.f = (RSERIES *)calloc(size=count+2,sizeof(RSERIES));
        y.f[0] = rs(1,b0ero,bone);
        y.f[0].p.p[0] = ione;
        y.f[0].p.p[1] = w;
        for(count=1; count <= y.n; ++count){
                y.f[count] = rs(1,cvtdb(r[count-1]),bone);
                y.f[count].p.p[0].up = i5[count-1].up;
                y.f[count].p.p[0].dn = i1[count-1].dn;
                y.f[count].p.p[1].up = i6[count-1].up;                      2060
                y.f[count].p.p[1].dn = i2[count-1].dn;
        }
        return(y);
}


GRS
tfrs(y)                                                                  tfrs
GRS y;
{                                                                           2070
        int i;
        GRS sol;
        unsigned size;
        /*char *calloc();*/
        INTERVL cvtdi(), x1, x2;

        sol.f = (RSERIES *)calloc(size=y.n-1,sizeof(RSERIES));
        sol.n = y.n-2;
        for (i=0; i <= sol.n; ++i){
                x1 = y.f[i+1].p.p[0];                                       2080
                x2 = y.f[i+1].p.p[1];
                sol.f[i] = vtffxi(cvtdi(y.f[i+1].center.b),
                    cvtdi(y.f[i+2].center.b),x1,x2);
                sol.f[i].center = y.f[i+1].center;
        }

        return(sol);
}

INTERVL                                                                     2090
Omega(u)                                                              Omega
GRS u;
{
```

```
        int d;
        INTERVL ievders(), sol, derw;
        double otest, otest1, otest2;

        sol.dn = (double) 2;
        sol.up = (double) 3;
                                                                    2100
        for(;;){
                otest = .5*(sol.up+sol.dn);
                if(otest == sol.up || otest == sol.dn)
                        return(sol);
                derw = grsdereval(u,cvtdi(otest));
                if(derw.up <= zero) sol.up = otest;
                else if(derw.dn >= zero) sol.dn = otest;
                else
                {
                        otest1 = otest;                             2110
                        d = 1;
                        while(d ){
                                otest2 = .5*(otest+sol.up);
                                if(otest2 == otest || otest2 == sol.up)d = 0;
                                derw = grsdereval(u,cvtdi(otest2));
                                if(derw.up <= zero ) sol.up = otest2;
                                else otest = otest2;
                        }

                        otest = otest1;                             2120
                        d = 1;
                        while(d ){
                                otest2 = .5*(otest+sol.dn);
                                if(otest2 == otest || otest2 == sol.dn)d = 0;
                                derw = grsdereval(u,cvtdi(otest2));
                                if(derw.up >= zero ) sol.dn = otest2;
                                else otest = otest2;
                        }

                        return(sol);                                2130
                }
        }
}

void
tfprint()                                                   tfprint
{
        GRS grstimesx();
        INTERVL  grseval();
                                                                    2140
        printivalio(W);
        printivalio(C1);
```

```
        Y = tff(W);
        printgrsio(Y);
        fflush(stdout);

        YRS = tfrs(Y);
        printgrsio(YRS);
        fflush(stdout);
                                                                    2150
        U = grstimesx(YRS);
        printgrsio(U);
        printivalio(RC = Omega(U));
        fflush(stdout);

        printivalio(BC = grscval(U,RC));
        fflush(stdout);
}

void                                                                2160
tfread()                                                        tfread
{
        int i;


        W = readivalio();
        C1 = readivalio();
        Y = readgrsio();
        YRS = readgrsio();
        for(i=0; i <= YRS.n; ++i)YRS.f[i].k = 2;                    2170
        U = readgrsio();
```

Acknowledgments. We wish to express our deepest gratitude to R. de la Llave: in addition to stimulating conversations, he taught us everything we know about computer-assisted proofs, gave us useful advice concerning the presentation of the paper, and went through the excruciating pain of checking our computer programs. We are also grateful to D. Rana for providing us with his interval arithmetic package. Finally, we thank the Department of Mathematics of the University of Texas at Austin for their help with the electronic distribution of the computer programs.

References.

[Ar] Arnold, V., *Mathematical methods of Classical Mechanics.* Springer.

[EKW] Eckmann, J. P., Koch, H. and Wittwer, P., A computer assisted proof of universality in area preserving maps. *Memoirs, Amer. Math. Soc.* **289** (1984).

[EW] Eckmann, J. P. and Wittwer, P., *Computer methods and Borel summability applied to Feigenbaum's equation.* Lecture Notes in Math. **227** (1985).

[FL] Fefferman, C. and Llave, R., Relativistic stability of matter, I. *Revista Mat. Iberoamericana* **2** (1986), 119-213.

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

[FS2] Fefferman, C. and Seco, L., Eigenvalues and eigenfunctions of ordinary differential operators. To appear in *Advances in Math.*

[FS3] Fefferman, C. and Seco, L., The eigenvalue sum for a one-dimensional potential. To appear in *Advances in Math.*

[FS4] Fefferman, C. and Seco, L., The density in a one-dimensional potential. To appear in *Advances in Math.*

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

[FS6] Fefferman, C. and Seco, L., The density in a three-dimensional radial potential. To appear in *Advances in Math.*

[FS7] Fefferman, C. and Seco, L., On the Dirac and Schwinger corrections to the ground-state energy of an atom. To appear in *Advances in Math.*

[KM] Kaucher, E. W. and Miranker, W. L., *Self-validating numerics for Function Space problems.* Academic Press, 1984.

[HKSW] Helffer, B., Knauf, A., Siedentop, H. and Weikard, R., On the absence of a first-order correction for the number of bound states of a Schrödinger

operator with Coulomb singularity. To appear in *Comm. Partial Diff. Equations.*

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

[LL] Lanford, O. and Llave, R., Solution of the functional equation for critical circle mappings with golden rotation number. In preparation.

[Ll] Llave, R. Computer assisted bounds in stability of matter. *Computer Aided Proofs in Analysis*, IMA Series in Math. and Appl. **28**, Springer, 1989.

[Lo] Lohner, R., Einschließung der Lösung gewöhnlicher Anfangs- und Randwertaufgaben und Anwendungen. Dissertation, Universität Karlsruhe (TH), 1988.

[Mo] Moore, R. E., *Methods and applications of Interval Analysis.* S.I.A.M., 1979.

[Ra] Rana , D., Proof of accurate upper and lower bounds for stability domains in denominator problems. Thesis, Princeton University, 1987.

[Se1] Seco, L., Lower bounds for the ground state energy of atoms. Thesis, Princeton University, 1989.

[S2] Seco, L., Computer assisted lower bounds for atomic energies. *Computer Aided Proofs in Analysis*, IMA Series in Math. and Appl. **28** (1989), 241-251.

[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.

Recibido: 15 de abril de 1.993

Charles L. Fefferman*
Department of Mathematics
Princeton University
Princeton NJ 08544, U.S.A.

Luis A. Seco
Department of Mathematics
California Institute of Technology
Pasadena CA 91125, U.S.A.

* Partially supported by a NSF grant at Princeton University

Revista Matemática Iberoamericana
Vol. 9, N.° 3, 1993

Sur les mesures de Wigner

Pierre Louis Lions et Thierry Paul

dédié à M. R. Dautray

Résumé. Nous étudions les propriétés de la transformée de Wigner pour des fonctions arbitraires dans L^2 ou pour des noyaux hermitiens du type "matrices densité". Et nous introduisons des limites de ces transformées de Wigner pour des suites de fonctions dans L^2, limites qui correspondent à la limite semi-classique en Mécanique Quantique. Les mesures obtenues ainsi, que nous appelons mesures de Wigner, possédent diverses propriétés mathématiques que nous établissons. En particulier, nous démontrons qu'elles satisfont, dans des cadres linéaires (équation de Schrödinger) ou nonlinéaires (équation de Hartree dépendant du temps), des équations de transport du type Liouville ou Vlasov.

Abstract. We study the properties of the Wigner transform for arbitrary functions in L^2 ir for hermitian kernels like the so-called density matrices. And we introduce some limits of these transforms for sequences of fonctions in L^2, limits that correspond to the semi-classical limit in Quantum Mechanics. The measures we obtain in this way, that we call Wigner measures, have various mathematical properties that we establish. In particular, we prove they satisfy, in linear situations (Schrödinger equations) or nonlinear ones (time-dependent Hartree equations), transport equations of Liouville or Vlasov type.

SOMMAIRE

I. Introduction.

En 1932, Wigner [45] a introduit une transformation que l'on peut écrire ainsi

$$(1) \qquad W(x,\xi) = (2\pi)^{-N} \int_{\mathbb{R}^N} e^{-i\xi\cdot y}\, \psi\big(x+\frac{y}{2}\big)\, \psi^*\big(x-\frac{y}{2}\big)\, dy\,,$$

pour tout $(x,\xi) \in \mathbb{R}^N \times \mathbb{R}^N$. Cette opération transformant une fonction ψ arbitraire dans $L^2(\mathbb{R}^N)$ en une fonction sur l'espace des phases (ici $\mathbb{R}^N_x \times \mathbb{R}^N_\xi$) est appelée *transformée de Wigner*. Dans (1) et dans tout ce qui suit, z^* désigne le conjugué d'un nombre complexe z. L'application qui à ψ associe W est bien évidemment quadratique mais devient linéaire dès lors que l'on introduit la *matrice densité* associée à ψ à savoir

$$(2) \qquad \rho(x,y) = \psi(x)\,\psi^*(y)\,, \qquad \text{p.p. } (x,y) \in \mathbb{R}^N \times \mathbb{R}^N\,.$$

Une propriété remarquable de cette transformée est la suivante: si $\psi = \psi(x,t)$ résout l'*équation de Schrödinger* (dans le cas d'une particule quantique libre)

$$(3) \qquad i\frac{\partial \psi}{\partial t} = -\frac{1}{2}\Delta\psi\,, \qquad \text{dans } \mathbb{R}^N_x \times \mathbb{R}_t\,,$$

alors $W(x,\xi,t)$ (donnée par (1) avec $\psi = \psi(t)$, pour tout $t \in \mathbb{R}$) résoud l'*équation de transport* libre (*i.e.*, l'*équation de Liouville* correspondant à une particule classique libre)

$$(4) \qquad \frac{\partial W}{\partial t} + \xi \cdot \nabla_x W = 0\,, \qquad \text{dans } \mathbb{R}^N_x \times \mathbb{R}^N_\xi \times \mathbb{R}_t\,.$$

Et W joue alors le rôle d'une densité de particules libres classiques du point de vue de la mécanique classique statistique. C'est pour cette

raison qu'on parle parfois de la densité de Wigner même si, et ce point qui gênait Wigner sera détaillé dans la suite, W n'est en général pas positive ou nulle.

Ce lien entre Mécanique Quantique (Statistique ou non) et Mécanique Classique est éclairé par les remarques suivantes classiques du point de vue physique: si $\hbar$ désigne la constante de Planck et si ψ $(= \psi_\hbar)$ résout

$$(5) \qquad i\hbar \frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2} \Delta \psi , \qquad \text{dans} \quad \mathbb{R}^N_x \times \mathbb{R}_t ,$$

alors $W_\hbar(x, \xi, t) = (1/\hbar^N)\, W(\psi)\left(x, \xi/\hbar, t\right)$ résout encore (4), comme on peut s'en convaincre aisément par un simple changement d'échelle en t. Ainsi, si $W_\hbar$ "converge" vers une fonction f quand $\hbar$ tend vers 0 en un sens convenable, on doit s'attendre à ce que f résolve également (4). Enfin, si l'on rajoute dans (5) un terme de force, *i.e.*

$$(6) \qquad i\hbar \frac{\partial \psi}{\partial t} = -\frac{\hbar}{2} \Delta \psi + V\psi , \qquad \text{dans} \quad \mathbb{R}^N_x \times \mathbb{R}_t ,$$

où V désigne un potentiel convenable (*i.e.*, une fonction sur $\mathbb{R}^N$), alors les mêmes considérations heuristiques indiquent que f, après passage à la limite quand $\hbar$ tend vers 0, devrait résoudre

$$(7) \qquad \frac{\partial f}{\partial t} + \xi \cdot \nabla_x f - \nabla V(x) \cdot \nabla_\xi f = 0 , \qquad \text{dans} \quad \mathbb{R}^N_x \times \mathbb{R}^N_\xi \times \mathbb{R}_t .$$

Et l'on reconnaît l'équation de Liouville (classique) correspondant au système (de Newton) Hamiltonien: $\dot{x} = \xi$, $\dot{\xi} = -\nabla V(x)$. Le formalisme introduit par Wigner permet donc d'établir la *limite semi-classique* et le passage de la Mécanique Quantique à la Mécanique Classique (de manière consistante du point de vue de la Physique Statistique) en faisant tendre $\hbar$ vers 0. Ce point de vue a été développé par de nombreux physiciens et nous nous contenterons de citer J. Yvon [46], [47], par des développements concernant la Physique Statistique.

La transformation de Wigner, la convergence de $W_\hbar$ quand $\hbar$ tend vers 0 et la limite semi-classique sont précisément les trois points essentiels que nous étudions rigoureusement dans cet article, et sont développées respectivement dans les trois sections qui le composent. A ce point, il convient d'ajouter que notre objectif n'est pas uniquement de rendre rigoureuses des considérations formelles classiques en

Physique. En effet, cette étude nécessite l'introduction de mesures obtenues comme limites de W_ε pour des suites arbitraires u_ε de $L^2(\mathbb{R}^N)$. Et ces mesures, que nous appelons *mesures de Wigner* (ou W-mesures), ont des propriétés mathématiques tout à fait remarquables, nous semble-t-il, qui seront certainement utiles pour des problèmes sans rapport avec la Mécanique Quantique. En fait, dans un travail indépendant du nôtre annoncé dans [19], P. Gérard a introduit ces mesures par une approche très différente (basée sur des éléménts du calcul pseudo-différentiel), a obtenu quelques-unes des propriétés que nous établissons et, surtout, a pu résoudre grâce à ces mesures des problèmes délicats d'homogénéisation. Nous décrivons brièvement dans la suite de cette Introduction les propriétés essentielles de ces mesures.

La troisième motivation de ce travail concerne la *limite semi-classique* et la possibilité de relier rigoureusement les équations de Schrödinger linéaires ou nonlinéaires (par exemple de type Hartree) aux équations dites *cinétiques* de la Mécanique Statistique et plus précisément aux équations de Liouville (dans le cas linéaire) ou aux équations de Vlasov (dans le cas nonlinéaire). Cela permet d'une part la déduction ab initio de tels systèmes cinétiques mais aussi dans le cas linéaire de contourner les difficultés des développements semi-classiques usuels liées aux caustiques. Enfin, il convient de noter que les équations obtenues par transformée de Wigner à partir d'équations de type Hartree sont utilisées en Physique Nucléaire ou dans la Physique des semi-conducteurs (l'interaction étant coulombienne comme dans le modèle original de Hartree en Physique Atomique, on parle alors de l'équation de Wigner-Poisson -voir par exemple G. Grimwall [23], U. Ravaioli et al. [36], W.R. Frensley [17], V.I. Tatarskii [42], P.A. Markowich [32].

Décrivons maintenant un peu plus précisément les résultats que nous avons établis. La Section II ci-dessous est consacrée à l'étude des transformées de Wigner non seulement pour des fonctions ψ dans L^2 mais également pour des opérateurs hermitiens de Hilbert-Schmidt sur L^2 c'est-à-dire pour des matrices-densité $\rho \in L^2(\mathbb{R}^N_x \times \mathbb{R}^N_y)$ vérifiant: $\rho(x,y) = \rho(x,y)^*$. Et nous supposerons le plus souvent que ces opérateurs sont positifs ou nuls.

Nous rappelons également dans cette section les liens entre équations de Schrödinger du type (3) ou (6) (et leurs formulations en terme d'équations de Liouville quantiques) et des équations de transport intégro-différentielles satisfaites par les transformées de Wigner (ces équa-

tions sont parfois appelées équations de Wigner). Signalons que cette réécriture de l'équation de Schrödinger est utilisée dans B. Perthame et P.L. Lions [30] pour établir ou retrouver des propriétés de dispersion et de régularité locale des solutions de l'équation de Schrödinger.

La Section III est la section centrale de notre article puisque:

1) nous introduisons les limites (au sens des distributions et en fait dans une dualité qui sera précisée) de

$$W_\varepsilon = \frac{1}{(2\pi\varepsilon)^N} \int_{\mathbb{R}^N} e^{-i(\xi/\varepsilon)\cdot y} \rho_\varepsilon\big(x + \frac{y}{2}, x - \frac{y}{2}\big)\, dy\,,$$

où $(\rho_\varepsilon)_{\varepsilon>0}$ est une suite bornée dans $L^2(\mathbb{R}^N_x \times \mathbb{R}^N_y)$ de noyaux définissant des opérateurs auto-adjoints, positifs ou nuls et de trace bornée,

2) nous montrons que ces limites sont des mesures *positives* bornées ou $\mathbb{R}^N_x \times \mathbb{R}^N_\xi$ *arbitraires*, la positivité étant une conséquence simple de l'observation suivante

$$(8) \qquad \tilde{W}_\varepsilon \;=\; W_\varepsilon * e^{-(|x|^2+|\xi|^2)/\varepsilon} \;\geq\; 0\,, \qquad \text{p.p. sur} \quad \mathbb{R}^N_x \times \mathbb{R}^N_\xi\,.$$

En fait, $\tilde{W}_\varepsilon$ est appelée transformée de Husimi de l'opérateur ρ_ε et est reliée aux états cohérents et aux paquets d'onde [8], [25] and [37],

3) nous donnons diverses autres constructions équivalentes de ces mesures notamment en faisant le lien avec la limite des quantités quadratiques construites à partir d'opérateurs pseudo-différentiels convenables. Il est à noter que les transformées de Wigner peuvent être considérées comme le symbole de Weyl des opérateurs considérés et que les limites semi-classiques que nous étudions correspondent aux règles de quantifications de Weyl (voir H. Weyl [44, page 274]) même si nous pouvons utiliser nos résultats dans le cadre du calcul symbolique usuel,

4) et enfin, nous montrons que ces mesures peuvent servir à mesurer les pertes de *compacité dans* L^2 (et dans L^2 uniquement) de $(\rho_\varepsilon)_{\varepsilon>0}$, enregistrant les concentrations et les oscillations éventuelles, en tous cas lorsque ces dernières ont une "longueur d'ordre $1/\varepsilon$" (si

$$\rho_\varepsilon = \psi_\varepsilon(x)\psi^*_\varepsilon(y)\,,$$

cela signifie que $\hat{\psi}_\varepsilon$ est essentiellement, au sens de la norme L^2, supportée dans des boules de rayon R/ε avec $R < \infty$). On voit donc un

lien méthodologique avec les H-mesures introduites indépendamment par L. Tartar [41] et P. Gérard [18] pour mesurer les défauts de compacité de suites bornées dans L^2, idée qui est à rapprocher de mesures de défaut de compacité introduites antérieurement dans d'autres situations par P.L. Lions [27], [28] (méthode de concentration-compacité) puis par R.J. DiPerna et A. Majda [15], [16]. Un résumé grossier des liens que nous établissons est donné par les assertions suivantes: i) la (ou les) H-mesures sont définies et ont un sens pour des suites arbitraires tandis que les mesures de Wigner nécessitent d'avoir une longueur caractéristique d'oscillations -comme c'est le cas dans les limites semi-classiques ou dans l'homogénéisation périodique, voir P. Gérard [19]-, ii) si la mesure de Wigner μ est "bien définie" (au sens précédent), la H-mesure σ est la mesure sur $\mathbb{R}^N_x \times \mathbb{R}^{N-1}_\omega$ définie par

$$d\sigma(x,\omega) = \int_0^\infty d\mu(x,t\omega)\,dt\,. \tag{9}$$

Cette observation permet en fait de donner des procédés nouveaux de construction des H-mesures.

Nous donnons également dans cette Section III quelques autres propriétés de ces mesures de Wigner ainsi que divers exemples représentatifs.

Enfin, la Section IV est consacrée à l'analyse de la *limite semi-classique* ($\hbar \to 0$) dans des équations de Schrödinger linéaires ((5), (6)) ou nonlinéaires. Ce type de limites est bien sûr classique en Physique et diverses présentations heuristiques se trouvent dans les traités classiques de Mécanique Quantique ou de Physique Nucléaire -une présentation rapide d'un cas nonlinéaire peut être trouvée dans V.P. Maslov [35]. En fait, nous partons de l'équation de Liouville où l'inconnue est la matrice densité $\rho(x,y)$

$$i\hbar\,\frac{\partial \rho}{\partial t} = [H,\rho]\,, \tag{10}$$

où $H = -\frac{\hbar^2}{2}\Delta + V$, $\hbar$ joue maintenant le rôle du paramètre ε précédent. Dans le cas linéaire, V est fixé tandis qu'un exemple canonique de problème nonlinéaire est donné par $V = V_0 * \rho$ et $\rho = \rho(x)$ est la densité ($= \rho(x,x)$). Bien sûr, si $\rho(x,y)$ vu comme un opérateur est un projecteur de rang 1, *i.e.*, $\rho(x,y) = \psi(x)\,\psi^*(y)$ avec $\|\psi\|_{L^2} = 1$, (10) est

une formulation équivalente de (6). Les hypothèses minimales sur ρ qui dépend de $\hbar$ bien évidemment sont alors: $\rho = \rho^*$, $\rho \geq 0$, $\mathrm{Tr}(\rho)$ bornée. La limite semi-classique que nous étudions consiste à introduire

$$f_\hbar = W_\hbar(\rho) = \frac{1}{(2\pi\hbar)^N} \int e^{-i(\xi/\hbar)\cdot y} \, \rho\big(x + \frac{y}{2}, x - \frac{y}{2}\big) \, dy$$
$$= \frac{1}{(2\pi)^N} \int e^{-i\xi\cdot y} \, \rho\big(x + \frac{\hbar y}{2}, x - \frac{\hbar y}{2}\big) \, dy$$

et à établir l'équation satisfaite par la (ou les) mesures de Wigner limites de $f_\hbar$ quand $\hbar$ tend vers 0. Si on note f une telle limite, on s'attend bien sûr à trouver les équations de *Liouville* (dans le cas linéaire) ou de *Vlasov* (dans le cas nonlinéaire) de la mécanique statistique classique à savoir

$$\frac{\partial f}{\partial t} + \xi \cdot \nabla_x f - \nabla_x V \cdot \nabla_\xi f = 0, \tag{11}$$

avec $V = V_0 * \rho$ et $\rho = \int_{\mathbb{R}^N} f(x,\xi)\, d\xi$ dans le cas non-linéaire.

Les problèmes mathématiques associés à ce passage à la limite sont nombreux: justification de la limite, type de convergence de $f_\hbar$ vers f et régularité du potentiel nécessaire à cela. Nous obtenons ici trois types de résultats:

i) si V (ou V_0) $\in C_0^1$ (*i.e.*, V, $\partial V/\partial x_i \in C_0(\mathbb{R}^N)$ pour $1 \leq i \leq N$), nous vérifions que $f_\hbar$ converge (par exemple au sens des distributions) vers f solution faible de (11). Même si la régularité nécessitée en V est importante, il est à noter qu'elle n'entraîne pas l'unicité des solutions de (11) ou de

$$\dot{x} = \xi, \qquad \dot{\xi} = -\nabla V(x), \tag{12}$$

ii) si V est régulier et, au temps $t = 0$, $f_\hbar$ converge "régulièrement" vers f_0 régulière alors $f_\hbar$ converge fortement dans $C([O,T]\,;\, L^2(\mathbb{R}^N_x \times \mathbb{R}^N_\xi))$ (pour tout $T < \infty$) vers la solution de (11) vérifiant $f|_{t=0} = f_0$ et de plus on peut écrire et justifier un développement asymptotique de $f_\hbar$ en puissances de $\hbar$ (valable par exemple au sens de $C([0,T]\,;\, L^2(\mathbb{R}^N_x \times \mathbb{R}^N_\xi))$ (pour tout $T < \infty$).

iii) si V ou V_0 sont peu réguliers et si, au temps $t = 0$, $f_\hbar$ est bornée dans $L^2(\mathbb{R}^N_x \times \mathbb{R}^N_\xi)$, on peut encore obtenir la convergence (faible dans

L^2) de $f_\hbar$ vers une solution f de (11). L'hypothèse essentielle que nous devons faire dans ce cas sur V ou V_0 est:

$$\nabla V \in L^2(\mathbb{R}^N) + L^q(\mathbb{R}^N), \qquad \text{où } 2 < q < \infty,$$
$$\nabla V_0 \in L^r(\mathbb{R}^N) + L^q(\mathbb{R}^N), \qquad \text{avec } \frac{2N+8}{N+8} < r < q < \infty.$$

Il est à noter que cette dernière hypothèse autorise le potentiel coulombien en dimension 3 (*i.e.*, $V_0(x) = 1/|x|$, $N = 3$) -puisque $3/2 > 14/11$. En d'autres termes, notre analyse donne en particulier la *convergence de Wigner-Poisson vers Vlasov-Poisson* en dimension 3. Rappelons que diverses études du système de Wigner-Poisson ont été réalisées ([34], [5], [39], [4], [6], [2], [40], [10], [11], [31], [32], [33]) tandis que le système de Vlasov-Poisson est maintenant assez bien compris (les résultats les plus généraux d'existence de solutions faibles ou de solutions régulières peuvent être trouvés respectivement dans R. J. DiPerna et P. L. Lions [12], [13], P. L. Lions et B. Perthame [30]).

Un des principaux ingrédients mathématiques de notre analyse dans le cas iii) est un résultat à la Lieb-Thirring [26] (voir aussi B. Simon [38] pour une présentation de ces résultats) détaillé dans l'appendice: si $(\psi_n)_{n\geq 1}$ est un système orthonormé dans $L^2(\mathbb{R}^N)$ et si $(\lambda_n)_{n\geq 1} \in l^p$ avec $\lambda_n \geq 0$ (pour tout $n \geq 1$), $1 < p \leq \infty$, alors on a dans le cas où $N = 3$ (par exemple)

$$\|\rho\|_{L^q(\mathbb{R}^3)} \leq C \, \|(\lambda_n)_n\|_{l^p}^{\theta} \left(\int_{\mathbb{R}^3} \sum_n \lambda_n |\nabla \psi_n|^2 \, dx \right)^{(1-\theta)/2}, \tag{13}$$

avec

$$\rho = \sum_n = \sum_n \lambda_n |\psi_n|^2, \quad q = \frac{2p'+3}{2p'+1}, \quad p' = \frac{p}{p-1}, \quad \theta = \frac{3}{2p'+3},$$

où $C > 0$ ne dépend que de p. Nous obtenons également des résultats, apparemment nouveaux également, du même type sur

$$j = \sum_n \lambda_n \, \mathrm{Im}\,(\nabla \psi_n \psi_n^*).$$

Enfin, signalons pour conclure cette longue introduction que l'étude du cas iii) sera prolongée par un travail en cours en collaboration avec

P. Gérard dans lequel nous étudions le cas de systèmes du type Vlasov-Maxwell, et nous revenons sur les résultats donnés en iii) en traitant d'autres conditions initiales (ce qui permet en fait d'affaiblir encore l'hypothèse faite sur V_0) et en précisant la convergence ainsi que la nature des solutions obtenues en faisant tendre $\hbar$ vers 0. Cela est rendu possible par la possibilité de "renormaliser" l'équation de Liouville quantique à la manière du cas classique (voir R. J. DiPerna et P. L. Lions [14]).

II. Transformées de Wigner.

Ainsi que nous l'avons indiqué dans l'Introduction, cette section est consacrée à diverses généralités sur la transformée de Wigner.

Pour commencer, soit $\psi \in L^2(\mathbb{R}^N)$, on note $\rho(x,y)$ la matrice densité associée à savoir l'expression donnée par (2). Bien sûr, $\rho(x,y) \in L^2(\mathbb{R}^N \times \mathbb{R}^N)$ et peut être considéré comme le noyau d'un opérateur borné sur $L^2(\mathbb{R}^N)$: cet opérateur n'est rien d'autre que l'opérateur de projection sur $\mathbb{R}\psi$ au facteur multiplicatif $\|\psi\|_{L^2}^2$ près. C'est donc un opérateur compact, positif, hermitien de trace égale à $\|\psi\|_{L^2}^2$. On voit en outre que $\rho(x,x) = |\psi(x)|^2$ a un sens et appartient à $L^1_+(\mathbb{R}^N)$. Une manière systématique de comprendre ce dernier point consiste à introduire $\tilde{\rho}(x,y)$

$$\tilde{\rho}(x,y) \;=\; \rho\big(x+\frac{y}{2}\,,\,x-\frac{y}{2}\big)\,, \qquad \text{p.p. } (x,y) \in \mathbb{R}^N \times \mathbb{R}^N\,. \tag{14}$$

Alors, bien sûr,

$$\tilde{\rho} \in L^2(\mathbb{R}^N \times \mathbb{R}^N) \cap C_0(\mathbb{R}^N_y; L^1(\mathbb{R}^N_x)) \cap C_0(\mathbb{R}^N_x; L^1(\mathbb{R}^N_y))$$

et ρ, vu comme opérateur, est hermitien si et seulement si

$$\tilde{\rho}(x,-y) \;=\; \tilde{\rho}(x,y)^*\,, \qquad \text{p.p. } (x,y) \in \mathbb{R}^N \times \mathbb{R}^N\,. \tag{15}$$

Enfin, $\rho(x) = \rho(x,x)$ (appelée densité) n'est rien d'autre que la restriction de $\tilde{\rho}$ au sous-espace $\{y=0\}$ de $\mathbb{R}^N \times \mathbb{R}^N$.

Ces considérations élémentaires s'étendent à des opérateurs plus généraux sur $L^2(\mathbb{R})$. En effet, soit K un opérateur hermitien de Hilbert-Schmidt sur $L^2(\mathbb{R}^N)$ alors -et il s'agit d'une caractérisation évidente- il existe un noyau $\rho(x,y) \in L^2(\mathbb{R}^N \times \mathbb{R}^N)$ vérifiant

$$\rho(x,y) \;=\; \rho(y,x)^*\,, \qquad \text{p.p. } (x,y) \in \mathbb{R}^N \times \mathbb{R}^N\,. \tag{16}$$

De plus, et c'est encore une fois un point de vue équivalent, on peut écrire

$$(17) \qquad \rho(x,y) = \sum_{i\in I} \lambda_i \psi_i(x) \psi_i^*(y), \qquad \text{p.p. } (x,y) \in \mathbb{R}^N \times \mathbb{R}^N,$$

où I est au plus dénombrable (vide si $\rho \equiv 0$), $\lambda_i \in \mathbb{R} \setminus \{0\}$ (pour tout $i \in I$) et

$$(18) \qquad \sum_i \lambda_i^2 < +\infty, \qquad \int_{\mathbb{R}^N} \psi_i \psi_j^* = \delta_{ij} \quad \text{(pour tout } i,j \in I).$$

En outre, K est positif si et seulement si

$$(19) \qquad \lambda_i > 0, \qquad \text{pour tout } i \in I.$$

Et, dans ce cas, K est de trace finie si et seulement si $\sum_i \lambda_i < +\infty$, ou également si et seulement si $\tilde{\rho}$ donné par (14) appartient à

$$C_0(\mathbb{R}^N_y ; L^1(\mathbb{R}^N_x)) \qquad (\cap C_0(\mathbb{R}^N_x ; L^1(\mathbb{R}^N_y))).$$

On note alors

$$\rho(x) = \tilde{\rho}|_{y=0} = \sum_i \lambda_i |\psi_i(x)|^2 \in L^1_+(\mathbb{R}^N_x)$$

de sorte que

$$(20) \qquad \mathrm{Tr}(K) = \sum_i \lambda_i = \int_{\mathbb{R}^N} \rho(x)\, dx$$

tandis que

$$(21) \qquad \mathrm{Tr}(K^2) = \sum_i \lambda_i^2 = \|\rho\|^2_{l^2(\mathbb{R}^N \times \mathbb{R}^N)}.$$

Dans tout ce qui suit, nous identifierons complètement K et son noyau $\rho(x,y)$. Nous utiliserons également la notation $\rho = \rho(x)$ pour la densité $\sum_i \lambda_i |\psi_i(x)|^2$ et nous parlerons simplement d'un opérateur ρ de Hilbert-Schmidt, hermitien, positif, éventuellement de trace finie.

On définit alors la transformée de Wigner de ρ

$$W(x,\xi) = \frac{1}{(2\pi)^N} \int_{\mathbb{R}^N} e^{-i\xi\cdot y}\, \rho(x+\frac{y}{2}, x-\frac{y}{2})\, dy\,, \tag{22}$$

pour $(x,\xi) \in \mathbb{R}^N \times \mathbb{R}^N$, *i.e.*, la transformée de Fourier de $\tilde{\rho}(x,\cdot)$. Au vu de (15), ρ est un opérateur hermitien si et seulement si W est réelle, de Hilbert-Schmidt si et seulement si $W \in L^2(\mathbb{R}^N_x \times \mathbb{R}^N_\xi)$ et

$$\|W\|^2_{L^2} = \frac{1}{(2\pi)^N} \|\tilde{\rho}\|^2_{L^2} = \frac{1}{(4\pi)^N} \|\rho\|^2_{L^2}\,. \tag{23}$$

La traduction en terme de W du fait que ρ soit de trace finie est plus délicate. Formellement, on observe que

$$\int_{\mathbb{R}^N} W(x,\xi)\, d\xi = \rho(x) \in L^1_+(\mathbb{R}^N)$$

en tout cas si ρ est positif. Cette intégration en ξ peut être en fait justifiée, toujours dans le cas où ρ est positif, en considérant

$$\int_{\mathbb{R}^N} W(x,\xi)\, e^{\varepsilon|\xi|^2/2}\, d\xi = \int_{\mathbb{R}^N_y} \tilde{\rho}(x,y)\, \frac{e^{-|y|^2/2\varepsilon}}{(2\pi\varepsilon)^{N/2}}\, dy$$

qui converge dans $L^1(\mathbb{R}^N)$ vers $\rho(x) \in L^1_+(\mathbb{R}^N)$ si et seulement si ρ est de trace finie. Nous verrons plus loin une caractérisation très simple de la propriété de trace finie.

La positivité de ρ semble se traduire difficilement en terme des transformées de Wigner. Une caractérisation implicite consiste à procéder par équivalences: $\rho \geq 0$ équivaut à

$$\begin{aligned} 0 &\leq \iint_{\mathbb{R}^N \times \mathbb{R}^N} \rho(x,y)\, \psi(y)\, \psi^*(x)\, dx\, dy \\ &= \iint_{\mathbb{R}^N \times \mathbb{R}^N} \tilde{\rho}(x,y)\, \psi\big(x-\frac{y}{2}\big)\, \psi^*\big(x+\frac{y}{2}\big)\, dx\, dy\,, \end{aligned} \tag{23}$$

d'où

$$\iint_{\mathbb{R}^N_x \times \mathbb{R}^N_\xi} W\, W_\psi\, dx\, d\xi \geq 0\,, \tag{24}$$

pour tout $\psi \in L^2(\mathbb{R}^N)$, où W_ψ est la transformée de Wigner de ψ ou plus précisément de $\psi(x)\,\psi^*(y)$. La condition nécessaire et suffisante (24) n'entraîne pas que $W \geq 0$. De plus, si on choisit

$$\psi(x) \;=\; \pi^{-N/4}\, e^{|x-x_0|^2/2}\, e^{i\xi_0\cdot x}\,, \qquad \text{où} \;\; (x_0,\xi_0) \in \mathbb{R}^{2N}$$

on trouve $W_\psi(x,\xi) = \pi^{-N}\, \exp\left(-(|x-x_0|^2+|\xi-\xi_0|^2)\right)$ et (24) implique

$$\tilde{W} \;=\; W * G \;\geq\; 0\,, \qquad \text{sur} \;\; \mathbb{R}^{2N}\,. \tag{25}$$

On appelle cette dernière quantité la transformée de Husimi de ρ.

En fait, (24) implique également beaucoup d'autres positivités de convoleés de W. En effet, si $\tilde{\rho}'$ est un noyau dans $L^2(\mathbb{R}^{2N})$ définissant un opérateur hermitien positif et si W' est sa transformée de Wigner (qui est donc réelle et appartient à $L^2(\mathbb{R}^{2N})$ d'après ce qui précède), on a donc d'après (23)-(24)

$$\iint_{\mathbb{R}^{2N}} W\,W'\,dx\,d\xi \;\geq\; 0\,. \tag{24'}$$

Or si on considère $\tilde{\rho}(x-x_0, y-y_0)\, e^{i\xi_0\cdot(x-y)}$, avec (x_0,ξ_0) quelconque dans $\mathbb{R}^{2N}$, ce nouveau noyau a les mêmes propriétés que $\tilde{\rho}$ et sa transformée de Wigner n'est rien d'autre que $W'(x-x_0,\xi-\xi_0)$ de sorte que (24′) implique

$$W * Z \;\geq\; 0\,, \qquad \text{sur} \;\; \mathbb{R}^{2N}\,, \tag{26}$$

avec $Z(x,\xi) \;=\; W'(-x,-\xi)$ sur $\mathbb{R}^{2N}$.

Enfin, si on revient sur l'hypothèse de trace finie pour ρ, on peut maintenant facilement la traduire en terme de W en écrivant que $\tilde{W} \in L^1(\mathbb{R}^{2N})$. On a alors

$$\begin{aligned} &\iint_{\mathbb{R}^{2N}} \tilde{W}\,dx\,d\xi \;=\; \sum_i \lambda_i \;=\; \int_{\mathbb{R}^N} \rho(x)\,dx\,, \\ &\int_{\mathbb{R}^N} \tilde{W}\,d\xi \;=\; \rho(x)\,, \qquad \text{p.p.} \;\; x \in \mathbb{R}^N\,. \end{aligned} \tag{27}$$

En *conclusion*, la transformée de Wigner W d'un noyau $\rho \in L^2(\mathbb{R}^{2N})$ hermitien est caractérisé par: W est réelle et $W \in L^2(\mathbb{R}^{2N})$. La po-

sitivité de ρ (en tant qu'opérateur sur $L^2(\mathbb{R}^N)$) est caractérisée par (24) (ou (26)). Et, dans ce cas, ρ est de trace finie si et seulement si $\tilde{W} = W * G$ (par exemple) $\in L^1(\mathbb{R}^{2N})$ -et cette dernière quantité est positive ou nulle sur $\mathbb{R}^{2N}$. Enfin, les identités (23) et (27) ont lieu. De façon à ne pas toujours rappeler les hypothèses faites sur ρ, de tels noyaux ρ vérifiant *toutes* les propriétés précédentes seront simplement appelées des *matrices densité*.

Signalons bien sûr que, puisque

$$\tilde{\rho} \in C_0(\mathbb{R}^N_x; L^1(\mathbb{R}^N_y)), \qquad W \in C_0(\mathbb{R}^N_x; \mathcal{F}L^1(\mathbb{R}^N_\xi))$$

et en remarquant que l'on a

$$(28) \qquad \begin{aligned} (\mathcal{F}_x^{-1} W)(\eta, \xi) &= (2\pi)^{-2N} \lambda_i \hat{\psi}_i \left(\frac{\xi - \eta}{2}\right) \hat{\psi}_i \left(\frac{\xi + \eta}{2}\right)^* \\ &\in C_0(\mathbb{R}^N_\xi; L^1(\mathbb{R}^N_\eta)), \end{aligned}$$

on en déduit également que $W \in C_0(\mathbb{R}^N_\xi; \mathcal{F}L^1(\mathbb{R}^N_x))$. Enfin,

$$\rho \in \mathcal{S}(\mathbb{R}^N_x \times \mathbb{R}^N_y) \quad \text{si et seulement si} \quad W \in \mathcal{S}(\mathbb{R}^N_x \times \mathbb{R}^N_\xi)$$

ce qui permet les arguments de densité habituels.

Nous allons maintenant conclure cette section par une brève présentation des équations satisfaites par la transformée de Wigner de la solution d'une équation de Schrödinger ou de Liouville. Nous commencerons par le cas de particules libres où donc $\psi \in C(\mathbb{R}_t; L^2(\mathbb{R}^N_x))$ est solution de (3). Plus généralement, si ρ_0 est une matrice densité, il existe une unique solution $\rho \in C(\mathbb{R}_t; L^2(\mathbb{R}^N_x \times \mathbb{R}^N_y))$ de l'équation de Liouville

$$(29) \qquad i \frac{\partial \rho}{\partial t} = [H_0, \rho], \qquad \text{pour } t \in \mathbb{R}, \quad \rho|_{t=0} = \rho_0,$$

avec H_0 donné par l'opérateur hermitien $\left(-\frac{1}{2}\Delta\right)$ non borné sur $L^2(\mathbb{R}^N)$ (de domaine $H^2(\mathbb{R}^N)$). Pour tout $t \in \mathbb{R}$, $\rho(t)$ est une matrice densité et si

$$\rho_0 = \sum_i \lambda_i \psi_i(x) \psi_i^*(y),$$

avec

$$\lambda_i > 0, \quad \sum_i \lambda_i < \infty, \quad \int_{\mathbb{R}^N} \psi_i \psi_j^* \, dx = \delta_{ij} \quad (\text{pour tout } i, j),$$

alors

$$\rho(t) = \sum_i \lambda_i \, \psi_i(t,x) \, \psi_i^*(t,y)$$

et $\psi_i(t) = e^{-itH_0}\psi_i$ est la solution de (3) dans $C(\mathbb{R}_t; L^2(\mathbb{R}^N_x)$ vérifiant $\psi_i(0) = \psi_i$. On peut également écrire $\rho(t) = e^{-itH_0} \, \rho_0 \, e^{itH_0}$.

On introduit alors $W(t,x,\xi)$ qui est, pour chaque $t \in \mathbb{R}$, la transformée de Wigner de $\rho(t)$. Un calcul élémentaire de transformée de Fourier donne que W vérifie (4) à savoir l'équation de transport (classique) libre. Ce passage de Schrödinger au transport libre est d'ailleurs systématiquement utilisé dans P.L. Lions et B. Perthame [30] pour préciser les lemmes de régularité locale de l'équation de Schrödinger (de type dispersion) grâce à l'équation (4) et les relier à des lemmes de gains locaux de moments pour des équations du type (4).

Il est d'ailleurs utile de préciser les liens entre moments de W et énergie cinétique de ρ. En effet, si $\psi_i \in H^1(\mathbb{R}^N)$ (pour tout i) et si

$$\sum_i \lambda_i \int_{\mathbb{R}^N} |\nabla \psi_i|^2 \, dx < \infty$$

ce que l'on peut écrire de manière synthétique $\mathrm{Tr}\,(H_0\rho_0) < +\infty$, cette propriété est bien sûr conservée (indépendamment de t) pour la solution de (29). Cela, en terme de W, signifie uniquement que

$$\iint_{\mathbb{R}^{2N}} W|\xi|^2 \, dx \, d\xi$$

est indépendant de $t \in \mathbb{R}$. En effet, en tout cas si $\rho \in \mathcal{S}(\mathbb{R}^N_x \times \mathbb{R}^N_y)$ de façon à ce que $W \in \mathcal{S}(\mathbb{R}^N_x \times \mathbb{R}^N_\xi)$, on a

$$\iint_{\mathbb{R}^{2N}} W|\xi|^2 \, dx \, d\xi \;=\; \mathrm{Tr}\,(H_0\rho)\,. \tag{30}$$

De manière encore plus précise, on peut observer que $\mathrm{Tr}\,(H_0\rho) < +\infty$ équivaut à

$$|\xi|^2 \, \tilde{W} \in L^1(\mathbb{R}^N_x \times \mathbb{R}^N_\xi) \quad \text{et} \quad \iint_{\mathbb{R}^{2N}} |\xi|^2 \, \tilde{W} \, dx \, d\xi = \mathrm{Tr}\,(H_0\rho) + N\,,$$

où $\tilde{W} = W * G$ est la transformée de Husimi, et ceci ne nécessite plus de supposer que $\rho \in \mathcal{S}(\mathbb{R}^N_x \times \mathbb{R}^N_y)$. En fait, l'identité entre moments

en ξ de W et normes (ou semi-normes) de type Sobolev sur ρ ou les ψ_i peut se voir grâce à la remarque générale suivante basée sur (28):

$$\int_{\mathbb{R}^N} W\,dx = \frac{1}{(2\pi)^N}\sum_i \lambda_i\,|\hat{\psi}_i(\xi)|^2\,. \tag{31}$$

Considérons maintenant le cas d'une équation de Schrödinger avec un potentiel V (réel)

$$\begin{cases} i\dfrac{\partial\psi}{\partial t} = -\dfrac{1}{2}\Delta\psi + V\psi\,, & \text{dans } \mathbb{R}_t\times\mathbb{R}^N_x\,,\\ \psi|_{t=0} = \psi_0\,, & \text{dans } \mathbb{R}^N\,, \end{cases} \tag{32}$$

ou de l'équation de Liouville associée

$$i\frac{\partial\rho}{\partial t} = [H,\rho]\,, \qquad \text{pour tout } t\in\mathbb{R}\,, \quad \rho|_{t=0}=\rho_0\,, \tag{33}$$

où H est l'opérateur $(-\frac{1}{2}\Delta + V)$. Une façon de résoudre (32) ou (33) consiste à écrire $\psi(t) = e^{-itH}\psi_0$ ou $\rho(t) = e^{-itH}\,\rho_0 e^{itH}$ pour $\psi_0\in L^2(\mathbb{R}^N)$ et ρ_0 étant une matrice densité. Il faut alors des conditions sur V assurant que H, sur un domaine convenable est un opérateur autoadjoint. Par exemple, d'après T. Kato [24], M. Aizenman et B. Simon [1] -voir aussi R. Dautray [9], on sait que H est un opérateur autoadjoint minoré si V vérifie

$$V^+\in L^1_{\mathrm{loc}}(\mathbb{R}^N)\,, \qquad V^-\in K^N(\mathbb{R}^N)\,, \tag{34}$$

où la classe $K^N(\mathbb{R}^N)$ est définie par

$$K^N(\mathbb{R}^N) = \left\{ f\in L^1_{\mathrm{loc}}(\mathbb{R}^N):\ \lim_{\varepsilon\to 0}\sup_x \int_{|x-y|\le\varepsilon} |x-y|^{2-N}|f(y)|\,dy = 0\right\},$$

si $N\ge 3$,

$$K^N(\mathbb{R}^N) = \left\{ f\in L^1_{\mathrm{loc}}(\mathbb{R}^2):\ \lim_{\varepsilon\to 0}\sup_x \int_{|x-y|\le\varepsilon} (\log|x-y|^{-1})|f(y)|\,dy = 0\right\},$$

si $N\ge 2$, et

$$K^N(\mathbb{R}^N) = \left\{ f\in L^1_{\mathrm{loc}}(\mathbb{R}):\ \sup_x \int_{|x-y|\le 1} |f(y)|\,dy < \infty\right\},$$

si $N \geq 3$.

Le domaine de H est donné par

$$D(H) = \left\{ \psi \in H^1(\mathbb{R}^N) : \ |V|\psi \in L^1_{\rm loc}(\mathbb{R}^N), \ -\frac{1}{2}\Delta\psi + V\psi \in L^2(\mathbb{R}^N) \right\}.$$

Pourtant, si on veut écrire l'équation satisfaite par la transformée de Wigner, des hypothèses supplémentaires sur V semblent nécessaires. Les hypothèses que nous ferons sont faciles à comprendre si on veut écrire que $\psi = e^{-itH}\psi_0 \in C(\mathbb{R}_t; L^2(\mathbb{R}^N_x))$ vérifie (32) dans $\mathcal{S}'$ (dans $\mathcal{D}'$ il suffirait de supposer que $V \in L^2_{\rm loc}(\mathbb{R}^N)$) et nous supposerons donc qu'il existe $C \geq 0$, $m \geq 0$

$$(35) \qquad V \in L^2_{\rm loc}(\mathbb{R}^N), \qquad \int_{|x|\leq R} |V(x)|^2\, dx \ \leq \ C\,(1+R)^m\,,$$

pour $R \geq 0$.

On démontre alors (sans grande difficulté) la

Proposition II.1. *Soit V vérifiant (34)-(35) et soit ρ_0 une matrice densité. On pose $\rho(t) = e^{-itH}\,\rho_0\,e^{itH}$ et on note $W(t,x,\xi)$ la transformée de Wigner de $\rho(t)$ (pour tout $t \in \mathbb{R}$). Alors, W est une fonction dans*

$$C\big(\mathbb{R}_t; L^2(\mathbb{R}^N_x \times \mathbb{R}^N_\xi)\big) \cap C_b\big(\mathbb{R}_t \times \mathbb{R}^N_x; \mathcal{F}L^1(\mathbb{R}^N_\xi)\big) \cap C_b\big(\mathbb{R}_t \times \mathbb{R}^N_\xi; \mathcal{F}L^1(\mathbb{R}^N_x)\big)$$

et vérifie

$$(36) \qquad \frac{\partial W}{\partial t} + \xi \cdot \nabla_x W + K \underset{\xi}{*} W \ = \ 0 \qquad \textit{dans} \ \ \mathcal{D}'\,,$$

où

$$\begin{aligned} K \ &= \ \frac{i}{(2\pi)^N} \int_{\mathbb{R}^N} e^{-i\xi\cdot y} \left(V\big(x + \frac{y}{2}\big) - V\big(x - \frac{y}{2}\big) \right) dy \\ &= \ i\left(e^{i2\xi\cdot x}\,\hat{V}(2\xi) - e^{-i2\xi\cdot x}\,\hat{V}(2\xi)^* \right). \end{aligned}$$

REMARQUE II.1. Le produit de convolution $K \underset{\xi}{*} W$ a un sens (dans $\mathcal{D}'$ ou dans $\mathcal{S}'$) puisque d'après (35) $\int |V(x)|^2 (1+|x|^2)^{-p}\, dx < \infty$ pour $p > 0$ suffisamment grand $(p > m/2)$ et donc $\hat{V} \in H^{-p}(\mathbb{R}^N)$. Ceci entraîne que

$$K \in C(\mathbb{R}^N_x; H^{-p}(\mathbb{R}^N_\xi)) \qquad \text{et} \qquad \|K(x,\cdot)\|_{H^{-p}} \leq C\,(1+|x|)^p\,.$$

Comme $W \in L^2(\mathbb{R}^N_x \times \mathbb{R}^N_\xi)$, on peut alors définir aisément $K \underset{\xi}{*} W$.

REMARQUE II.2. Plusieurs choix de normalisation sont possibles pour les transformées de Wigner. Nous avons choisi la normalisation classique.

REMARQUE II.3. L'énergie totale au niveau de (32) ou (33) s'écrit

$$\text{(37)} \qquad \begin{aligned} \operatorname{Tr}(H_0\rho + V\rho) &= \frac{1}{2}\iint_{\mathbb{R}^{2N}} W|\xi|^2\,dx\,d\xi + \int_{\mathbb{R}^N} V\,\rho\,dx \\ &= \iint_{\mathbb{R}^{2N}} W\left(\frac{1}{2}|\xi|^2 + V(x)\right)dx\,d\xi\,, \end{aligned}$$

avec $\rho(x) = \int_{\mathbb{R}^N} W(x,\xi)\,d\xi$. Il est intéressant de retrouver sur (36) les invariances classiques de l'équation de Schrödinger à savoir que $\operatorname{Tr}(\rho)$, $\operatorname{Tr}(\rho^2)$ et $\operatorname{Tr}(H_0\rho + V\rho)$ sont indépendants de t. Ceci se fait aisément sur (36) -en supposant pour simplifier la présentation que V, W sont réguliers (dans $\mathcal{S}$ par exemple)- en remarquant que

$$\iint_{\mathbb{R}^{2N}} W\,dx\,d\xi\,, \qquad \iint_{\mathbb{R}^{2N}} W^2\,dx\,d\xi\,,$$

et

$$\frac{1}{2}\iint_{\mathbb{R}^{2N}} W|u|^2\,dx\,d\xi + \int_{\mathbb{R}^N} V\rho\,dx$$

sont indépendants de t. En effet, pour la première invariance, il suffit d'observer que $\int_{\mathbb{R}^N} K\,d\rho = 0$. Pour la deuxième invariance, on remarque que

$$\begin{aligned} &\int_{\mathbb{R}^N} (K * W)W\,d\xi \\ &= -2\,\operatorname{Im}\left(\iint_{\mathbb{R}^N\times\mathbb{R}^N} \hat{V}(2\xi)\,e^{2i(\xi-\eta)\cdot x}\,W(x,\xi)\,W(x,\eta)\,d\xi\,d\eta\right) = 0\,, \end{aligned}$$

en échangeant ξ et η puisque $\hat{V}(-\xi) = \hat{V}(\xi)^*$ (car V est réel). Enfin, la conservation de l'énergie totale découle des identités suivantes

$$\begin{aligned} &\frac{d}{dt}\left(\frac{1}{2}\iint_{\mathbb{R}^{2N}} W|\xi|^2\,d\xi\,dx\right) = -\frac{1}{2}\iint_{\mathbb{R}^{2N}} (K*W)|\xi|^2\,d\xi\,dx \\ &= -\frac{i}{2(2\pi)^N}\int_{\mathbb{R}^{3N}} dx\,dy\,d\eta\,W(x,\eta)\Big(V\big(x+\frac{y}{2}\big) - V\big(x-\frac{y}{2}\big)\Big) \end{aligned}$$

$$\cdot \int_{\mathbb{R}^N} |\xi|^2 \, e^{-i(\xi-\eta)\cdot y} \, d\xi$$

$$= -\frac{i}{2} \int_{\mathbb{R}^{2N}} dx \, d\eta W(x,y)(-\Delta_y) \left(\left(V\left(x+\frac{y}{2}\right) - V\left(x-\frac{y}{2}\right) \right) e^{i\eta\cdot y} \right)|_{y=0}$$

$$= -\int_{\mathbb{R}^N} dx \, \nabla V(x) \cdot \int_{\mathbb{R}^N} d\eta \, \eta \, W(x,\eta)$$

$$= \int_{\mathbb{R}^N} dx \, V(x) \, \operatorname{div} \int_{\mathbb{R}^N} W(x,\eta)\eta \, d\eta \, .$$

Or, toujours d'après (36),

$$\frac{\partial \rho}{\partial t} + \operatorname{div} \int_{\mathbb{R}^N} W(x,\xi)\xi \, d\xi \; = \; 0 \, ,$$

ce qui permet de conclure:

$$\frac{d}{dt} \left(\frac{1}{2} \iint_{\mathbb{R}^{2N}} W|\xi|^2 \, d\xi \, dx + \int_{\mathbb{R}^N} V\rho \, dx \right) \; = \; 0 \, .$$

Signalons une dernière identité remarquable satisfaite par les solutions de (32) ou (33)

$$\frac{d^2}{dt^2} \int_{\mathbb{R}^N} |x|^2 \, \rho(x) \, dx \; = \; 2 \operatorname{Tr}(H_0 \rho) - 2 \int_{\mathbb{R}^N} x \cdot \nabla V(x) \, \rho(x) \, dx \tag{38}$$

que l'on peut retrouver grâce à (36) par des calculs (fastidieux) semblables à ce qui précède. Cette identité en terme de W devient bien sûr

$$\begin{aligned} &\frac{d^2}{dt^2} \int_{\mathbb{R}^N} |x|^2 \, \rho(x) \, dx \\ &\qquad = \iint_{\mathbb{R}^{2N}} W|\xi|^2 \, dx \, d\xi - 4 \int_{\mathbb{R}^N} x \cdot \nabla V \rho \, dx \\ &\qquad = \iint_{\mathbb{R}^{2N}} W'(|\xi|^2 - 4\, x \cdot \nabla V) \, dx \, d\xi \, . \end{aligned} \tag{39}$$

III. Mesures de Wigner.

Cette section est consacrée à l'étude des limites de transformées de Wigner pour des suites bornées dans $L^2(\mathbb{R}^N)$ ou plus généralement pour des suites de noyaux hermitiens, positifs ou nuls, de trace bornée et bornées dans $L^2(\mathbb{R}^N_x \times \mathbb{R}^N_y)$. Nous conviendrons dans la suite d'appeler simplement suite bornée de matrices densité de telles suites de noyaux. Nous établissons les principales propriétés de ces limites en commençant par le cas de suites bornées u_ε dans $L^2(\mathbb{R}^N)$ (avec $\varepsilon \in]0,1]$ par exemple).

Ainsi que nous l'avons expliqué dans l'Introduction, nous introduisons

$$
\begin{aligned}
W_\varepsilon(x,\xi) &= \frac{1}{(2\pi\varepsilon)^N} \int_{\mathbb{R}^N} e^{-i(\xi/\varepsilon)\cdot y}\, u_\varepsilon\Big(x+\frac{y}{2}\Big)\, u_\varepsilon^*\Big(x-\frac{y}{2}\Big)\, dy \\
&= \frac{1}{(2\pi)^N} \int_{\mathbb{R}^N} e^{-i\xi\cdot z}\, u_\varepsilon\Big(x+\frac{\varepsilon z}{2}\Big)\, u_\varepsilon^*\Big(x-\frac{\varepsilon z}{2}\Big)\, dz
\end{aligned}
\tag{40}
$$

(transformée de Wigner avec changement d'échelle d'ordre ε en ξ) ainsi que les transformées de Husimi correspondantes

$$
\tilde{W}_\varepsilon(x,\xi) = W_\varepsilon * \left(e^{-|x|^2/\varepsilon}\, e^{-|\xi|^2/\varepsilon}\, (\pi\varepsilon)^{-N} \right). \tag{41}
$$

Nous avons vu dans la section précédente que $\tilde{W}_\varepsilon \geq 0$ sur $\mathbb{R}^N_x \times \mathbb{R}^N_\xi$ en fait, on a

$$
\begin{aligned}
\tilde{W}_\varepsilon(x,\xi) &= 2^{N/2}(2\pi\varepsilon)^{-N} \\
&\quad \cdot \left| \int_{\mathbb{R}^N} u_\varepsilon(z)\, e^{-|x-z|^2/(4\varepsilon)}\, (2\pi\varepsilon)^{-N/4}\, e^{-i\xi\cdot z/2\varepsilon}\, dz \right|^2
\end{aligned}
\tag{42}
$$

et on reconnaît là (à quelques constantes près) le module au carré des paquets d'ondes de A. Córdoba et C. Fefferman [8].

Il est clair que ε n'étant que le paramètre définissant la suite u_ε, la normalisation $\big(\xi/\varepsilon\,,\,(2\pi\varepsilon)^{-N}\big)$ dans (40) est arbitraire. Nous verrons plus loin que les limites de W_ε (quand $\varepsilon \to 0_+$) donneront des informations précises sur le comportement de u_ε lorsque ε est soigneusement choisi en fonction des $(u_\varepsilon)_\varepsilon$ et plus précisément lorsque ε est "la longueur caractéristique des oscillations de u_ε".

Bien sûr, si $(u_\varepsilon)_\varepsilon$ est borné dans $L^2(\mathbb{R}^N)$, on voit immédiatement que $\tilde{W}_\varepsilon$ est une suite bornée de fonctions positives ou nulles dans

$L^1(\mathbb{R}^N)$ (rappeler par exemple que $\iint_{\mathbb{R}^{2N}} \tilde{W}_\varepsilon \, dx \, d\xi = \int_{\mathbb{R}^N} |u_\varepsilon|^2 \, dx$). Par contre, les bornes sur W_ε sont probablement moins évidentes. Une manière d'en obtenir consiste à introduire l'espace suivant (il s'agit en fait d'une algèbre) de fonctions test

$$\mathcal{A} = \{\varphi \in C_0(\mathbb{R}^N_x \times \mathbb{R}^N_\xi) : (\mathcal{F}_\xi \varphi)(x,z) \in L^1(\mathbb{R}^N_z; C_0(\mathbb{R}^N_x))\}$$

muni de la norme

$$\|\mathcal{F}_\xi \varphi\|_{L^1_z(C_x)} = \int_{\mathbb{R}^N} \sup_x |\mathcal{F}_\xi \varphi|(x,z) \, dz \, .$$

On vérifie sans peine que $\mathcal{A}$ est un espace (et une algèbre) de Banach séparable contenant $\mathcal{S}(\mathbb{R}^N_x \times \mathbb{R}^N_\xi)$ et que $\mathcal{S}(\mathbb{R}^N_x \times \mathbb{R}^N_\xi)$, $C_0^\infty(\mathbb{R}^N_x \times \mathbb{R}^N_\xi)$, $\{\varphi \in \mathcal{S} : \mathcal{F}_\xi \varphi \in C_0^\infty(\mathbb{R}^N_x \times \mathbb{R}^N_z)\}$ ou même les combinaisons linéaires de produits $\psi_1(x)\psi_2(\xi)$ (avec $\psi_1, \psi_2 \in C_0^\infty$ ou $\mathcal{F}(C_0^\infty)$) sont denses dans $\mathcal{A}$.

Proposition III.1. *La suite W_ε est bornée dans $\mathcal{A}'$.*

DÉMONSTRATION (évidente). Pour tout $\varphi \in \mathcal{A}$,

$$\begin{aligned}
&\int_{\mathbb{R}^{2N}} W_\varepsilon \varphi \, dx \, d\xi \\
&= \int_{\mathbb{R}^{2N}} dx \, d\xi \, \frac{1}{(2\pi)^N} \int_{\mathbb{R}^N} e^{-i\xi\cdot z} \, u_\varepsilon\big(x + \frac{\varepsilon z}{2}\big) \, u_\varepsilon^*\big(x - \frac{\varepsilon z}{2}\big) \, \varphi(x,\xi) \, dx \, dz \\
&= \frac{1}{(2\pi)^N} \int_{\mathbb{R}^{2N}} dx \, dz \, \Big((\mathcal{F}_\xi \varphi)(x,z) \, u_\varepsilon\big(x + \frac{\varepsilon z}{2}\big) \, u_\varepsilon^*\big(x - \frac{\varepsilon z}{2}\big) \Big) \, ,
\end{aligned}$$

d'où

$$\begin{aligned}
\left| \int_{\mathbb{R}^{2N}} W_\varepsilon \varphi \, dx \, d\xi \right| &\leq \frac{1}{(2\pi)^N} \Big(\int_{\mathbb{R}^N} \sup_x |\mathcal{F}_\xi \varphi|(x,z) \, dz \Big) \\
&\quad \cdot \Big(\sup_x \Big| \int_{\mathbb{R}^N} u_\varepsilon\big(x + \frac{\varepsilon z}{2}\big) \, u_\varepsilon^*\big(x - \frac{\varepsilon z}{2}\big) \, dx \Big) \\
&\leq \frac{1}{(2\pi)^N} \, \|\varphi\|_{\mathcal{A}} \, \|u_\varepsilon\|_{L^2}^2 \, .
\end{aligned}$$

Quitte à extraire une sous-suite, on peut donc toujours supposer que $u_\varepsilon, W_\varepsilon, \tilde{W}_\varepsilon$ convergent faiblement vers $u, \mu, \tilde{\mu}$ respectivement dans $L^2(\mathbb{R}^N)$, $\mathcal{A}'$ (muni de la topologie faible $*$) et au sens des mesures.

Nous pratiquerons systématiquement dans ce qui suit l'abus de langage consistant à noter encore $u_\varepsilon, W_\varepsilon, \tilde{W}_\varepsilon$ les suites extraites. En fait, dans toute cette section, ε désignera une suite $\varepsilon_n > 0$ qui converge vers 0. Il convient de noter également que même si u_ε ou u_{ε_n} converge faiblement dans L^2 vers u, rien n'assure a priori que $W_\varepsilon, \tilde{W}_\varepsilon$ (ou $W_{\varepsilon_n}, \tilde{W}_{\varepsilon_n}$) convergent faiblement sans qu'il ne soit nécessaire d'extraire des sous-suites (supplémentaires). Enfin, nous noterons $\mathcal{M} = \mathcal{M}(\mathbb{R}^k)$ le cône des mesures positives ou nulles bornées sur $\mathbb{R}^k$.

Avec ces hypothèses et notations, on démontre le

Théorème III.1.

1) *On a* $\mu \equiv \tilde{\mu}$. *En particulier,* $\mu \in \mathcal{M}(\mathbb{R}^{2N})$.

2) *L'inégalité suivante a lieu*

$$\mu \geq |u(x)|^2 \, \delta_0(\xi)\,, \tag{43}$$

d'où en particulier

$$\int_{\mathbb{R}^N} |u(x)|^2 \, dx \leq \iint_{\mathbb{R}^{2N}} d\mu \leq \liminf_{\varepsilon \to 0} \int_{\mathbb{R}^N} |u_\varepsilon|^2 \, dx\,. \tag{44}$$

3) $|u_\varepsilon(x)|^2$ *converge faiblement au sens des mesures vers* $\int_{\mathbb{R}^N_\xi} d\mu(\cdot, \xi)$ *si et seulement si* $\left|\hat{u}_\varepsilon\left(\xi/\varepsilon\right)\right|^2/\varepsilon^N$ *est une suite étroitement relativement compacte dans* $\mathcal{M}(\mathbb{R}^N)$, *i.e.*

$$\sup_\varepsilon \left\{ \frac{1}{\varepsilon^N} \int_{|\xi| \geq R} \left|\hat{u}_\varepsilon\left(\frac{\xi}{\varepsilon}\right)\right|^2 d\xi = \int_{|\xi| \geq R/\varepsilon} |\hat{u}_\varepsilon(\xi)|^2 \, d\xi \right\} \to 0\,, \tag{45}$$

si $R \to +\infty$. *Si* $|u_\varepsilon(x)|^2$, $\left|\hat{u}_\varepsilon\left(\xi/2\right)\right|^2/(2\pi\varepsilon)^N$ *(ou des sous-suites) convergent faiblement au sens des mesures vers des mesures positives ou nulles bornées notées respectivement* μ_x, μ_ξ, *alors:* $\mu_x \geq \int_{\mathbb{R}^N} d\mu(\cdot, \xi)$, $\mu_\xi \geq \int_{\mathbb{R}^N} d\mu(x, \cdot)$.

4) *L'égalité*

$$\iint_{\mathbb{R}^{2N}} d\mu = \lim_{\varepsilon \to 0} \int_{\mathbb{R}^N} |u_\varepsilon|^2 \, dx$$

a lieu si et seulement si $|u_\varepsilon(x)|^2$ *et* $\varepsilon^{-N}\left|\hat{u}_\varepsilon\left(\xi/2\right)\right|^2$ *sont étroitement relativement compactes dans* $\mathcal{M}(\mathbb{R}^N)$. *En particulier, si cette hypothèse*

est vérifiée, u_ε converge dans $L^2(\mathbb{R}^N)$ vers u si et seulement si $\mu = |u(x)|^2\,\delta_0(\xi)$.

5) *Soit $\mu \in \mathcal{M}(\mathbb{R}^{2N})$, soit $u \in L^2(\mathbb{R}^N)$ telle que $\mu \geq |u(x)|^2\,\delta_0(\xi)$. Alors, il existe $u_\varepsilon \in L^2(\mathbb{R}^N)$ telle que u_ε converge faiblement dans $L^2(\mathbb{R}^N)$ vers u, W_ε converge faiblement dans $\mathcal{A}'_{w-*}$ vers μ et*

$$\int_{\mathbb{R}^N} |u_\varepsilon(x)|^2\,dx \to \iint_{\mathbb{R}^{2N}} d\mu\,, \qquad \textit{dans} \quad \varepsilon \to 0_+\,.$$

Avant de démontrer ce résultat, il convient de faire quelques remarques et de donner une série d'exemples où l'on peut "calculer" μ. Bien sûr, nous appelons μ la *mesure de Wigner* associée à u_ε (ou à la sous-suite convenable).

REMARQUE III.1. Rappelons qu'une suite bornée de mesures positives ou nulles bornées μ_n sur $\mathbb{R}^k$ est étroitement relativement compacte si et seulement si

$$\sup_n \mu_n\{|x| > R\} \to 0\,, \qquad \text{si} \quad R \to +\infty\,.$$

REMARQUE III.2. Ce théorème montre que les limites de transformées de Wigner sont des mesures bornées positives ou nulles "arbitraires" sur l'espace des phases et qu'elles contiennent toute l'obstruction à la compacité dans $L^2(\mathbb{R}^N)$ de suites bornées dans $L^2(\mathbb{R}^N)$, en tout cas si $|u_\varepsilon|^2$ est étroitement compacte (pour éviter le phénomène dit de bosse glissante, par exemple $u_\varepsilon(x) = u(x + e/\varepsilon)$ avec $|e| = 1$) et surtout si $\left|\hat{u}_\varepsilon(\xi/2)\right|^2/\varepsilon^N$ est étroitement compacte. Cette dernière hypothèse est fondamentale pour obtenir une mesure μ informative sur le comportement de la suite u_ε. Elle signifie en effet que la longueur "non initiale d'oscillation" de u_ε est d'ordre ε; ce point sera illustré par les exemples qui suivent et nous reviendrons plus loin sur cette question en indiquant également des hypothèses simples et réalistes qui permettent de vérifier cette hypothèse.

EXEMPLE III.1. (Suite compacte) $u_\varepsilon \underset{\varepsilon}{\rightharpoonup} u$ dans $L^2(\mathbb{R}^N)$. Alors

$$u_\varepsilon\Big(x + \frac{\varepsilon z}{2}\Big)\,u_\varepsilon^*\Big(x - \frac{\varepsilon z}{2}\Big) \underset{\varepsilon}{\rightharpoonup} |u(x)|^2$$

(faiblement) dans $\mathcal{S}'(\mathbb{R}^N_x \times \mathbb{R}^N_z)$, d'où

$$W_\varepsilon = (2\pi)^{-N} \mathcal{F}_z \left(u_\varepsilon(x + \frac{\varepsilon z}{2})\, u^*_\varepsilon(x - \frac{\varepsilon z}{2}) \right) \underset{\varepsilon}{\rightharpoonup} |u(x)|^2\, \delta_0(\xi)$$

et $\mu = |u(x)|^2\, \delta_0(\xi)$.

EXEMPLE III.2. (Suite oscillante) $u_\varepsilon = u\varphi(x/\varepsilon^\alpha)$ où $u \in L^2(\mathbb{R}^N)$, $\varphi \in L^\infty(\mathbb{R}^N)$ est périodique en $x_1, \ldots, x_N$ et $\alpha > 0$. Alors, si $\alpha < 1$, le même raisonnement que celui de l'Exemple III.1 s'applique et donne $\mu = \langle |\varphi|^2 \rangle\, |u(x)|^2\, \delta_0(\xi)$ où $\langle |\varphi|^2 \rangle$ désigne la moyenne de $|\varphi|^2$ sur sa période. Si $\alpha > 1$, on peut vérifier que $\mu = 0$ par le même argument que celui que nous allons maintenant donner dans le cas $\alpha = 1$. Si $\alpha = 1$ donc, par un simple argument de densité, il suffit de considérer le cas où $u \in C_0^\infty(\mathbb{R}^N)$ et d'étudier la limite de

$$(2\pi)^{-N} \mathcal{F}_z \left(|u(x)|^2\, \varphi(\frac{x}{\varepsilon} + \frac{z}{2})\, \varphi^*(\frac{x}{\varepsilon} - \frac{z}{\varepsilon}) \right).$$

Pour simplifier les notations, on peut supposer que φ est périodique en chacun des x_i de période 2π et on développe φ en série de Fourier

$$\varphi = \sum_{k \in \mathbb{Z}^N} \varphi_k\, e^{ik \cdot x}.$$

On obtient alors aisément que $\mu = |u(x)|^2 \sum_k \delta_k(\xi) |\varphi_k|^2$. On peut également écrire cette expression en introduisant

$$\Gamma(z) = \langle \varphi(\cdot + \frac{z}{2})\, \varphi(\cdot - \frac{z}{2}) \rangle$$

et on voit que $\mu = (2\pi)^{-N} |u(x)|^2\, \hat{\Gamma}(\xi)$.

EXEMPLE III.3. (Suite avec concentration ponctuelle)

$$u_\varepsilon = \frac{1}{\varepsilon^{N\alpha/2}}\, u(\frac{x}{\varepsilon^\alpha}), \qquad \text{où } u \in L^2(\mathbb{R}^N).$$

Si $\alpha < 1$, comme à l'Exemple III.1, on voit que

$$\mu = \left(\int_{\mathbb{R}^N} |u(x)|^2\, dx \right) \delta_0(x)\, \delta_0(\xi).$$

Si $\alpha > 1$, on vérifie facilement que $\mu = 0$. Enfin, si $\alpha = 1$, on observe que

$$W_\varepsilon = \frac{1}{\varepsilon^N} W\Big(\frac{x}{\varepsilon}, \xi\Big),$$

avec

$$W(x,\xi) = \frac{1}{(2\pi)^N} \int e^{-i\xi\cdot z}\, u\big(x + \frac{z}{2}\big)\, u^*\big(x - \frac{z}{2}\big)\, dz,$$

de sorte que (en raisonnant par densité par exemple)

$$\mu = \delta_0(x)\Big(\int_{\mathbb{R}^N} W(y,\xi)\, dy\Big).$$

Et d'après la section précédente (voir (31))

$$\int_{\mathbb{R}^N} W(y,\xi)\, dy = \frac{1}{(2\pi)^N} |\hat{u}(\xi)|^2.$$

D'où finalement $\mu = \delta_0(x)\,(4\pi)^{-n} |\hat{u}(\xi)|^2$.

EXEMPLE III.4. (État cohérent) On pose

$$u_\varepsilon = \varepsilon^{-N\alpha/\varepsilon}\, u\Big(\frac{x - x_0}{\varepsilon^\alpha}\Big)\, e^{i(\xi_0/\varepsilon)\cdot x}$$

où $u \in L^2(\mathbb{R}^N)$, $\alpha > 0$, $(x_0, \xi_0) \in \mathbb{R}^{2N}$. Alors, si $0 < \alpha < 1$

$$\begin{aligned} W_\varepsilon(x,\xi) = {} & (2\pi\varepsilon)^{-N}\, \varepsilon^{-N\alpha} \int_{\mathbb{R}^N} e^{-i(\xi/\varepsilon)\cdot z}\, u\left(\frac{x - x_0 + z/2}{\varepsilon^\alpha}\right) \\ & \cdot u^*\left(\frac{x - x_0 + z/2}{\varepsilon^\alpha}\right) e^{i(\xi_0/\varepsilon)\cdot z}\, dz \\ = {} & (2\pi\varepsilon)^{-N} \int_{\mathbb{R}^N} e^{-i(\xi/\varepsilon^{1-\alpha})\cdot y}\, u\left(\frac{x - x_0}{\varepsilon^\alpha} + \frac{y}{2}\right) \\ & \cdot u^*\left(\frac{x - x_0}{\varepsilon^\alpha} - \frac{y}{2}\right) e^{-i(\xi_0/\varepsilon^{1-\alpha})\cdot y}\, dy \\ = {} & \varepsilon^{-N}\, W\left(\frac{x - x_0}{\varepsilon^\alpha}, \frac{\xi - \xi_0}{\varepsilon^{1-\alpha}}\right) \\ \underset{\varepsilon}{\rightharpoonup} {} & \left(\iint_{\mathbb{R}^{2N}} W\, dx\, d\xi\right) \delta_{x_0}(x)\, \delta_{\xi_0}(\xi), \end{aligned}$$

où

$$W(x,\xi) = (2\pi)^{-N} \int_{\mathbb{R}^N} e^{-i\xi\cdot y}\, u\big(x+\frac{y}{2}\big)\, u^*\big(x-\frac{y}{2}\big)\, dy\,,$$

de sorte que d'après la section précédente

$$\iint_{\mathbb{R}^{2N}} W\, dx\, d\xi = \int_{\mathbb{R}^N} |u(x)|^2\, dx\,.$$

D'où,

$$\mu = \Big(\int_{\mathbb{R}^N} |u(x)|^2\, dx\Big)\, \delta_{x_0}(x)\, \delta_{\xi_0}(\xi)\,.$$

Si $\alpha > 1$, on vérifie sans peine que $\mu = 0$. Enfin, si $\alpha = 1$, on obtient de la même manière que dans l'Exemple III.4, $\mu = (2\pi)^{-N}\, \delta_0(x)\, |\hat{u}(\xi - \xi_0)|^2$.

EXEMPLE III.5. (État WKB) On pose $u_\varepsilon = u\, e^{ia(x)/\varepsilon^\alpha}$ où $u \in L^2(\mathbb{R}^N)$, $a \in W^{1,1}_{\mathrm{loc}}(\mathbb{R}^N)$ et a est réelle, $\alpha \in]0,1]$.

Alors, $u_\varepsilon\big(x+\varepsilon z/2\big)\, u^*_\varepsilon\big(x-\varepsilon z/2\big)$ converge dans $\mathcal{S}'(\mathbb{R}^{2N})$ vers $|u(x)|^2$ si $\alpha < 1$, $|u(x)|^2\, e^{i\nabla a(x)\cdot z}$ si $\alpha = 1$. D'où

$$\mu = |u(x)|^2\, \delta_0(\xi) \quad \text{si } \alpha > 1\,, \qquad \mu = |u(x)|^2\, \delta_{\nabla a(x)}(\xi) \quad \text{si } \alpha = 1\,.$$

Le cas $\alpha > 1$ est plus délicat sauf si on suppose $\nabla a(x) \neq 0$ p.p. sur $\mathbb{R}^N$ auquel cas on voit facilement que $\mu = 0$. Si par contre $\nabla a(x) = 0$ sur un ensemble de mesure positive (et $\alpha > 1$), l'identification de μ requiert des informations supplémentaires sur a en ses points critiques.

EXEMPLE III.6. (États liés de l'oscillateur harmonique) Pour $A_1, \ldots A_N > 0$ rationnellement dépendants, on pose

$$u_\varepsilon = c_\varepsilon\, H_{A_1/\varepsilon}\big(\frac{x_1}{\sqrt{\varepsilon}}\big) \ldots H_{A_N/\varepsilon}\big(\frac{x_N}{\sqrt{\varepsilon}}\big)\, e^{-|x|^2/(2\varepsilon)}\,,$$

avec c_ε constante de normalisation L^2 et $H_k(x)$ polynome d'Hermite, alors, pour $\varepsilon_k \to 0$ avec

$$\varepsilon_k = \frac{A_1}{n_k^1} = \cdots = \frac{A_N}{n_k^N}\,,$$

n_k^i entiers,

$$\mu = \delta(x_1^2 + \xi_1^2 - A_1) \ldots \delta(x_N^2 + \xi_N^2 - A_N)\,.$$

Cet exemple montre un exemple concret où il faut prendre une sous-suite en ε, situation générique lorsque μ a un support compact (règles de quantification).

REMARQUE III.3. On voit que dans les exemples III.2, III.3, III.4 et III.5, $\mu \equiv 0$. Ceci est bien sûr à rapprocher de la Remarque III.2 puisque $\alpha > 1$ signifie que la longueur caractéristique des "oscillations" de u_ε est d'ordre ε^α et donc plus petite que ε. Ceci explique le fait que μ se trivialise et ne donne donc plus aucune information sur le comportement de u_ε.

REMARQUE III.4. On observe que si μ est la mesure de Wigner associée à une suite $(u_\varepsilon)_\varepsilon$ alors $\mu(\cdot - x_0, \cdot - \xi_0)$ est la mesure de Wigner associée à

$$u_\varepsilon(x - x_0)\, e^{i(\xi_0/\varepsilon)\cdot x}\,, \qquad \text{pour tout } \ (x_0, \xi) \in \mathbb{R}^{2N}\,.$$

REMARQUE III.5. La mesure de Wigner μ est aussi la limite (quand ε tend vers 0) de

$$(2\pi)^{-N} \int e^{-i\xi\cdot z}\, u_\varepsilon\big(x + \alpha\frac{\varepsilon z}{2}\big)\, u_\varepsilon^*\big(x - \beta\frac{\varepsilon z}{2}\big)\, dz$$

pour tous $\alpha, \beta \in \mathbb{R}$ tels que $\alpha + \beta = 1$.

REMARQUE III.6. La mesure de Wigner associée à une suite $(u_\varepsilon + v_\varepsilon)_\varepsilon$ où $u_\varepsilon, v_\varepsilon$ génèrent des mesures de Wigner μ, ν n'est en général pas $\mu + \nu$ (prendre $v_\varepsilon = u_\varepsilon$ et remarquer que la mesure de Wigner associée à $2u_\varepsilon$ est 4μ !). Par contre, si μ et ν sont étrangères (ou mutuellement singulières), on peut démontrer que cette additivité est alors vraie -par exemple en utilisant le fait que les mesures de Wigner sont les limites des transformées de Husimi $\tilde{W}_\varepsilon$ (*cf.* Théorème III.1). En effet, grâce au caractère quadratique de W_ε et donc de $\tilde{W}_\varepsilon$, on voit que (avec des notations évidentes)

$$\tilde{W}_\varepsilon(u_\varepsilon + u_\varepsilon) = \tilde{W}_\varepsilon(u_\varepsilon) + \tilde{W}_\varepsilon(u_\varepsilon) + 2\, R_\varepsilon\,,$$

avec

$$R_\varepsilon(x, \xi) \in L^1 \quad \text{et} \quad |R_\varepsilon(x, \xi)| \leq \tilde{W}_\varepsilon(u_\varepsilon)^{1/2}\, \tilde{W}_\varepsilon(u_\varepsilon)^{1/2}\,.$$

Pour montrer que R_ε converge faiblement vers 0, il suffit alors de prendre φ arbitraire dans $C_0^\infty(\mathbb{R}^{2N})$ et de construire, pour tout $\alpha > 0$,

$\psi_\alpha, \chi_\alpha \in C_0^\infty(\mathbb{R}^{2N})$ tels que $0 \leq \psi_\alpha, \chi_\alpha \leq 1$, $\psi_\alpha + \chi_\alpha \equiv 1$ sur le support de φ et $\int_{\mathbb{R}^{2N}} \psi_\alpha \, d\nu$, $\int_{\mathbb{R}^{2N}} \chi_\alpha \, d\mu \leq \alpha$. On voit alors que

$$\int_{\mathbb{R}^{2N}} R_\varepsilon \, \varphi \, dx \, d\xi = \int_{\mathbb{R}^{2N}} \varphi \, \psi_\alpha \, R_\varepsilon \, dx \, d\xi + \int_{\mathbb{R}^{2N}} \varphi \, \chi_\alpha \, R_\varepsilon \, dx \, d\xi \,,$$

$$\left| \int_{\mathbb{R}^{2N}} \varphi \psi_\alpha R_\varepsilon \, dx \, d\xi \right| \leq \left(\sup_{\mathbb{R}^{2N}} |\varphi| \right) \int_{\mathbb{R}^{2N}} \psi_\alpha \tilde{W}_\varepsilon(u_\varepsilon)^{1/2} \, \tilde{W}_\varepsilon(u_\varepsilon)^{1/2} \, dx \, d\xi \,,$$

$$\left| \int_{\mathbb{R}^{2N}} \varphi \chi_\alpha R_\varepsilon \, dx \, d\xi \right| \leq \left(\sup_{\mathbb{R}^{2N}} |\varphi| \right) \int_{\mathbb{R}^{2N}} \chi_\alpha \tilde{W}_\varepsilon(u_\varepsilon)^{1/2} \, \tilde{W}_\varepsilon(u_\varepsilon)^{1/2} \, dx \, d\xi \,.$$

De plus,

$$\int_{\mathbb{R}^{2N}} \psi_\alpha \, \tilde{W}_\varepsilon(u_\varepsilon)^{1/2} \, \tilde{W}_\varepsilon(u_\varepsilon)^{1/2} \, dx \, d\xi \leq \int_{\mathbb{R}^{2N}} \psi_\alpha \left(\frac{\sqrt{\alpha}}{2} \tilde{W}_\varepsilon(u_\varepsilon) + \frac{1}{2\sqrt{\alpha}} \tilde{W}_\varepsilon(u_\varepsilon) \right) dx \, d\xi \,,$$

d'où

$$\overline{\lim_\varepsilon} \left| \int_{\mathbb{R}^{2N}} \varphi \psi_\alpha \, R_\varepsilon \, dx \, d\xi \right| \leq \frac{\sqrt{\alpha}}{2} \int_{\mathbb{R}^{2N}} \psi_\alpha \, d\mu + \frac{1}{2\sqrt{\alpha}} \int_{\mathbb{R}^{2N}} \psi_\alpha \, d\nu \leq C_1 \sqrt{\alpha} \,.$$

De même, on obtient:

$$\overline{\lim_{\varepsilon \to 0}} \left| \int_{\mathbb{R}^{2N}} \varphi \chi_\alpha \, R_\varepsilon \, dx \, d\xi \right| \leq C_2 \sqrt{\alpha} \,,$$

où C_1 et C_2 sont des constantes indépendantes de α. On en déduit aisément que

$$\int_{\mathbb{R}^{2N}} \varphi R_\varepsilon \, dx \, d\xi \underset{\varepsilon}{\to} 0 \,,$$

ce qui prouve l'additivité de la mesure de Wigner dans ce cas.

Signalons une autre situation où $W_\varepsilon(u_\varepsilon * v_\varepsilon) \underset{\varepsilon}{\rightharpoonup} \mu + \nu$: on suppose que u_ε converge (fortement) dans $L^2(\mathbb{R}^N)$ vers u et donc $\mu = |u(x)|^2 \, \delta_0(\xi)$ (*cf.* Exemple III.1) et que v_ε converge faiblement dans

$L^2(\mathbb{R}^N)$ vers 0. Un exemple de ce phénomène est donné par la superposition d'un nombre fini d'états cohérents

$$u_\varepsilon = \sum_{j=1}^{N} \alpha_j u^j\left(\frac{x-x_j}{\sqrt{\varepsilon}}\right) e^{-N/4}\, e^{i(\xi_j\cdot x)/\varepsilon},$$

où $(x_j,\xi_j)_{1\le j\le N}$ est un ensemble fini de points distincts de $\mathbb{R}^{2N}$, $\alpha_j\in\mathbb{C}$, $u^j\in L^2(\mathbb{R}^N)$ et $\int_{\mathbb{R}^N}|u^j|^2\,dx=1$. On vérifie facilement que la mesure de Wigner associée à une telle suite est donnée par

$$\mu = \sum_{j=1}^{N} |\alpha_j|^2\,\delta_{x_j}(x)\,\delta_{\xi_j}(\xi).$$

DÉMONSTRATION DU THÉORÈME III.1.

DÉMONSTRATION DU POINT 1). Il suffit bien sûr de prouver que $\mu=\tilde{\mu}$. Or,

$$\tilde{W}_\varepsilon = W_\varepsilon * G_\varepsilon, \qquad \text{où } G_\varepsilon = (\pi\varepsilon)^{-N} e^{-(|x|^2+|\xi|^2)/\varepsilon},$$

et il suffit donc de prouver que si $\varphi\in\mathcal{A}$ (ou une partie dense de $\mathcal{A}$) alors $\varphi * G_\varepsilon$ converge dans $\mathcal{A}$ vers φ. Comme

$$\mathcal{F}_\xi(\varphi * G_\varepsilon)(x,z) = \left[(\mathcal{F}_\xi\varphi)(x,z) \underset{x}{*} (\pi\varepsilon)^{-N/2} e^{-|x|^2/\varepsilon}\right] e^{-\varepsilon|z|^2/4},$$

on voit que

$$|\varphi * G_\varepsilon - \varphi|_{\mathcal{A}} \le \int_{\mathbb{R}^N} dz \sup_x \left|(\mathcal{F}_\xi\varphi) - (\mathcal{F}_\xi\varphi) * (\pi\varepsilon)^{-N/2} e^{-|x|^2/\varepsilon}\right| + \int_{\mathbb{R}^N} \left(1-e^{-\varepsilon|z|^2/4}\right) \sup_x |\mathcal{F}_\xi\varphi|\,dz.$$

Le deuxième terme tend vers 0 quand ε tend vers 0_+. Il en va de même pour le premier terme si φ et donc $\mathcal{F}_\xi\varphi\in\mathcal{S}(\mathbb{R}^N\times\mathbb{R}^N)$ ce qui suffit pour établir 1). Mais en fait, par un argument de densité standard, ce premier terme converge vers 0 pour tout $\varphi\in\mathcal{A}$.

DÉMONSTRATION DU POINT 2). La première inégalité de (44) se déduit bien sûr de (43) et la deuxième inégalité se déduit de l'observation suivante

$$\iint_{\mathbb{R}^{2N}} d\mu = \iint_{\mathbb{R}^{2N}} d\tilde{\mu} \le \varliminf_{\varepsilon} \iint_{\mathbb{R}^{2N}} \tilde{W}_\varepsilon\,dx\,d\xi = \varliminf_{\varepsilon} \int_{\mathbb{R}^N} |u_\varepsilon|^2\,dx,$$

d'après (27). Il suffit donc d'établir (43). Il est particulièrement commode d'utiliser pour ce faire les transformées de Husimi en observant que l'on a (avec des notations évidentes)

$$\begin{aligned}\tilde{W}_\varepsilon(u_\varepsilon) &= \tilde{W}_\varepsilon(u) + \tilde{W}_\varepsilon(u_\varepsilon - u) + 2\,\tilde{W}_\varepsilon(u, u_\varepsilon - u)\\ &\geq \tilde{W}_\varepsilon(u) + 2\,\tilde{W}_\varepsilon(u, u_\varepsilon - u)\,.\end{aligned}$$

D'après l'Exemple III.1, $\tilde{W}_\varepsilon(u)$ (comme $W_\varepsilon(u)$) converge vers $|u(x)|^2 \cdot \delta_0(\xi)$. Il suffit donc de vérifier que $\tilde{W}_\varepsilon(u, u_\varepsilon - u)$ converge faiblement (au sens des mesures) vers 0. Bien sûr $\tilde{W}_\varepsilon(v_1, v_2)$ est bilinéaire symétrique en (v_1, v_2).

Mais, on a bien sûr

$$\|\tilde{W}_\varepsilon(u,v)\|_{L^1} \leq \|u\|_{L^2}\,\|v\|_{L^2}$$

(en fait, $|\tilde{W}_\varepsilon(u,v)| \leq \tilde{W}_\varepsilon(u)^{1/2}\,\tilde{W}_\varepsilon(v)^{1/2}$ p.p. $(x,\xi) \in \mathbb{R}^{2N}$). Par densité, il suffit alors de montrer que $\tilde{W}_\varepsilon(u, v_\varepsilon)$ converge faiblement (au sens des mesures) vers 0 si $u \in C_0^\infty(\mathbb{R}^N)$ et si v_ε converge faiblement vers 0 dans $L^2(\mathbb{R}^N)$.

Or, $\tilde{W}_\varepsilon(u, v_\varepsilon) = W_\varepsilon(u, v_\varepsilon) * G_\varepsilon$ et

$$\begin{aligned}W_\varepsilon(u, v_\varepsilon) &= (2\pi)^{-N} \int_{\mathbb{R}^N} e^{-i\xi\cdot z}\,\frac{1}{2}\left\{u\big(x + \frac{\varepsilon z}{2}\big)\,v_\varepsilon^*\big(x - \frac{\varepsilon z}{2}\big)\right.\\ &\qquad \left. + v_\varepsilon\big(x + \frac{\varepsilon z}{2}\big)\,u^*\big(x - \frac{\varepsilon z}{2}\big)\right\} dz\\ &= (2\pi)^{-N}\,\mathrm{Re}\int_{\mathbb{R}^N} e^{-i\xi\cdot z}\,u\big(x + \frac{\varepsilon z}{2}\big)\,v_\varepsilon^*\big(x - \frac{\varepsilon z}{2}\big)\,dz\,.\end{aligned}$$

Donc, pour tout $\varphi \in \mathcal{S}(\mathbb{R}^N \times \mathbb{R}^N)$, on a

$$\begin{aligned}&\langle W_\varepsilon(u, v_\varepsilon), \varphi\rangle_{\mathcal{A}'\times\mathcal{A}}\\ &\quad = (2\pi)^{-N}\,\mathrm{Re}\iint_{\mathbb{R}^{2N}} dx\,dz\,u\big(x + \frac{\varepsilon z}{2}\big)\,v_\varepsilon^*\big(x - \frac{\varepsilon z}{2}\big)\,(\mathcal{F}_\xi\varphi)(x,z)\\ &\quad = (2\pi)^{-N}\,\mathrm{Re}\iint_{\mathbb{R}^{2N}} dy\,dz\,v_\varepsilon^*(y)\,u(y + \varepsilon z)(\mathcal{F}_\xi\varphi)\big(y + \frac{\varepsilon z}{2}, z\big)\,.\end{aligned}$$

Or,

$$u(y + \varepsilon z)(\mathcal{F}_\xi\varphi)\big(y + \frac{\varepsilon z}{2}, z\big) \underset{\varepsilon}{\to} u(y)(\mathcal{F}_\xi\varphi)(y,z)$$

dans $L^1(\mathbb{R}^N_z, L^2(\mathbb{R}^N_x))$ puisque $\varphi \in \mathcal{S}$, $u \in C_0^\infty(\mathbb{R}^N)$. Comme v_ε converge faiblement vers 0 dans $L^2(\mathbb{R}^N)$, on obtient donc que $W_\varepsilon(u, v_\varepsilon)$ converge faiblement dans $\mathcal{A}'(w-*)$ vers 0. Par le même raisonnement que celui utilisé pour le point 1), on en déduit que $\tilde{W}_\varepsilon(u, v_\varepsilon)$ converge faiblement au sens des mesures vers 0.

REMARQUE III.7. La démonstration précédente montre que

$$\begin{aligned} W_\varepsilon(u_\varepsilon) &= W_\varepsilon(u) + W_\varepsilon(u-u_\varepsilon) + 2\,W_\varepsilon(u, u-u_\varepsilon)\,, \\ \tilde{W}_\varepsilon(u_\varepsilon) &= \tilde{W}_\varepsilon(u) + \tilde{W}_\varepsilon(u-u_\varepsilon) + 2\,\tilde{W}_\varepsilon(u, u-u_\varepsilon) \end{aligned}$$

et $W_\varepsilon(u)$ ou $\tilde{W}_\varepsilon(u)$, $W_\varepsilon(u-u_\varepsilon)$ ou $\tilde{W}_\varepsilon(u-u_\varepsilon)$, $W_\varepsilon(u, u-u_\varepsilon)$ ou $\tilde{W}_\varepsilon(u, u-u_\varepsilon)$ convergent faiblement (dans $\mathcal{A}'_{w-*}$ ou au sens des mesures) respectivement vers $|u(x)|^2\,\delta_0(\xi)$, $\mu - |u(x)|^2\,\delta_0(\xi)$, 0.

DÉMONSTRATION DU POINT 3): Nous commençons par la deuxième partie du point 3) et si μ_x, μ_ξ sont respectivement les limites faibles de $|u_\varepsilon(x)|^2$, $(\pi\varepsilon)^{-N}\left|\hat{u}_\varepsilon(\xi/\varepsilon)\right|^2$, remarquons qu'elles sont encore les limites faibles de

$$\begin{aligned} |u_\varepsilon(x)|^2 * \frac{e^{-|x|^2/\varepsilon}}{(\pi\varepsilon)^{N/2}} &= \int_{\mathbb{R}^N} W_\varepsilon\, d\xi\,, \\ \frac{1}{(2\pi\varepsilon)^N}\left|\hat{u}_\varepsilon\Big(\frac{\xi}{\varepsilon}\Big)\right|^2 * \frac{e^{-|\xi|^2/\varepsilon}}{(\pi\varepsilon)^{N/2}} &= \int_{\mathbb{R}^N} W_\varepsilon\, dx\,. \end{aligned}$$

Il suffit alors d'introduire $\varphi_R(\cdot) = \varphi\big(\cdot/R\big)$ où $\varphi \in C_0^\infty(\mathbb{R}^N)$, $0 \le \varphi \le 1$, $\varphi = 1$ sur $B(0,1)$ et $R > 0$ et d'observer que $\tilde{W}_\varepsilon$ étant positive ou nulle, nous avons pour tout $\psi \in C_0^\infty(\mathbb{R}^N)$, $\psi \ge 0$:

$$\begin{aligned} \int_{\mathbb{R}^N} \psi(x)\Big(\int_{\mathbb{R}^N} \tilde{W}_\varepsilon\, d\xi\Big)\, dx &\ge \iint_{\mathbb{R}^{2N}} \psi(x)\,\varphi_R(\xi)\,\tilde{W}_\varepsilon\, d\xi\, dx\,, \\ \int_{\mathbb{R}^N} \psi(\xi)\Big(\int_{\mathbb{R}^N} \tilde{W}_\varepsilon\, dx\Big)\, d\xi &\ge \iint_{\mathbb{R}^{2N}} \psi(\xi)\,\varphi_R(x)\,\tilde{W}_\varepsilon\, d\xi\, dx\,; \end{aligned}$$

d'où en passant à la limite quand ε tend vers 0_+

$$\begin{aligned} \int_{\mathbb{R}^N} \psi\, d\mu_x \ge \iint_{\mathbb{R}^{2N}} \psi(x)\,\varphi_R(\xi)\, d\mu &\underset{R\to\infty}{\to} \int_{\mathbb{R}^N} \psi(x)\Big(\int_{\mathbb{R}^N} d\mu(\cdot,\xi)\Big)\,, \\ \int_{\mathbb{R}^N} \psi\, d\mu_\xi \ge \iint_{\mathbb{R}^{2N}} \psi(\xi)\,\varphi_R(x)\, d\mu &\underset{R\to\infty}{\to} \int_{\mathbb{R}^N} \psi(\xi)\Big(\int_{\mathbb{R}^N} d\mu(x,\cdot)\Big)\,. \end{aligned}$$

Si on suppose que $(4\pi\varepsilon)^{-N}\left|\hat{u}_\varepsilon(\xi/(2\varepsilon))\right|^2$ est étroitement relativement compacte, comme cette quantité n'est rien d'autre que $\int_{\mathbb{R}^N} W_\varepsilon\,dx$, on en déduit aisément que

$$\int_{\mathbb{R}^N} \tilde{W}_\varepsilon\,dx = \Big(\int_{\mathbb{R}^N} W_\varepsilon\,dx\Big) * \frac{e^{-|\xi|^2/\varepsilon}}{(\pi\varepsilon)^{N/2}}$$

est également étroitement compacte. Et donc

$$\sup_\varepsilon \iint_{\mathbb{R}^{2N}} 1_{\{|\xi|>R\}}\,\tilde{W}_\varepsilon\,dx\,d\xi \;\to\; 0\,, \qquad \text{si } R\to+\infty\,.$$

Cela suffit bien sûr à assurer que pour tout $\varphi\in C_0^\infty(\mathbb{R}^N)$

$$\iint_{\mathbb{R}^{2N}} \tilde{W}_\varepsilon\,\varphi(x)\,dx\,d\xi \;\to\; \iint_{\mathbb{R}^{2N}} \varphi(x)\,d\mu(x,\xi)$$

et le point 3) est démontré.

REMARQUE III.8. La condition donnée au point 3) est suffisante pour assurer que $\mu_x = \int_{\mathbb{R}^N} d\mu(\cdot,\xi)$ mais n'est en général pas nécessaire. Il suffit pour s'en convaincre de considérer l'exemple suivant. On choisit $u_\varepsilon(x) = u\big(x+e_1/\varepsilon\big)\,e^{ix_1/\varepsilon^2}$ où $u\in C_0^\infty(\mathbb{R}^N)$ et on vérifie sans peine que $\mu_x\equiv 0$, $\mu\equiv 0$. Pourtant,

$$\frac{1}{\varepsilon^N}\Big|\hat{u}_\varepsilon\big(\frac{\xi}{\varepsilon}\big)\Big|^2 = \frac{1}{\varepsilon^N}\Big|\hat{u}\big(\frac{\xi}{\varepsilon}-\frac{e_1}{\varepsilon^3}\big)\Big|^2$$

n'est pas étroitement relativement compacte.

DÉMONSTRATION DU POINT 4): La deuxième partie du point 4) est une conséquence immédiate de la première partie et de l'Exemple III.1. En ce qui concerne le premier point, on remarque tout d'abord que $\iint_{\mathbb{R}^{2N}} \tilde{W}_\varepsilon\,dx\,d\xi = \int_{\mathbb{R}^N} |u_\varepsilon|^2\,dx$ converge vers $\iint_{\mathbb{R}^{2N}} d\mu$ si et seulement si $\tilde{W}_\varepsilon$ est étroitement relativement compacte. De plus, $\tilde{W}_\varepsilon$ est étroitement relativement compacte si et seulement si $\int_{\mathbb{R}^N} \tilde{W}_\varepsilon\,dx$, $\int_{\mathbb{R}^N} \tilde{W}_\varepsilon\,d\xi$ sont étroitement relativement compactes. Or, on a

$$\begin{aligned}
\int_{\mathbb{R}^N} \tilde{W}_\varepsilon\,dx &= (2\pi\varepsilon)^{-N}\Big|\hat{u}\big(\frac{\xi}{2}\big)\Big|^2 * \frac{e^{-|\xi|^2/\varepsilon}}{(\pi\varepsilon)^{N/2}}\,,\\
\int_{\mathbb{R}^N} \tilde{W}_\varepsilon\,d\xi &= (2\pi\varepsilon)^{-N}\,|u_\varepsilon|^2 * \frac{e^{-|x|^2/\varepsilon}}{(\pi\varepsilon)^{N/2}}\,.
\end{aligned} \tag{46}$$

Il ne nous reste donc plus qu'à montrer que si μ_ε est une suite bornée de mesures positives ou nulles bornées sur $\mathbb{R}^N$, μ_ε est étroitement relativement compacte si et seulement si $u_\varepsilon * \left(e^{-|y|^2/\varepsilon}/(\pi\varepsilon)^{N/2}\right)$ est étroitement relativement compacte. Or, on a d'une part

$$\begin{aligned}
&\int_{|x|>R}\int_{\mathbb{R}^N} \frac{e^{-|x-y|^2/\varepsilon}}{(\pi\varepsilon)^{N/2}}\,dx\,d\mu_\varepsilon(y)\\
&\quad=\int_{\mathbb{R}^N}\left(\int_{|x|>R}\frac{e^{-|x-y|^2/\varepsilon}}{(\pi\varepsilon)^{N/2}}\,dx\right)d\mu_\varepsilon(y)\\
&\quad\le\int_{|y|>(R-1)} d\mu_\varepsilon(y)+\int_{\mathbb{R}^N}\left(\int_{|x-y|\ge 1}\frac{e^{-|x-y|^2/\varepsilon}}{(\pi\varepsilon)^{N/2}}\,dx\right)d\mu_\varepsilon(y)\\
&\quad\le\int_{|y|>(R-1)} d\mu_\varepsilon(y)+C\left(\int_{|z|\ge\varepsilon^{-1/2}}\frac{e^{-|z|^2/\varepsilon}}{(\pi)^{N/2}}\,dz\right).
\end{aligned}$$

D'autre part, on a également

$$\begin{aligned}
&\int_{|x|>R}\int_{\mathbb{R}^N} \frac{e^{-|x-y|^2/\varepsilon}}{(\pi\varepsilon)^{N/2}}\,dx\,d\mu_\varepsilon(y)\\
&\quad\ge\int_{|y|>(R+1)}\left(\int_{|x-y\le 1}\frac{e^{-|x-y|^2/\varepsilon}}{(\pi\varepsilon)^{N/2}}\,dx\right)d\mu_\varepsilon(y)\\
&\quad\ge\left(\int_{|z|\le\varepsilon^{-1/2}}\frac{e^{-|z|^2}}{(\pi)^{N/2}}\,dz\right)\int_{|y|>(R+1)} d\mu_\varepsilon(y)\,;
\end{aligned}$$

et ces inégalités suffisent pour conclure.

DÉMONSTRATION DU POINT 5). Au vu de la Remarque III.6, il suffit de traiter le cas où $u\equiv 0$. D'après le point 4), il nous faut construire une suite (et une véritable famille u_ε paramétrée par $\varepsilon\in]0,1]$ peut également être construite avec la même démonstration et un peu plus de soin) u_ε convergeant faiblement vers 0 dans $L^2(\mathbb{R}^N)$ telle que $W_\varepsilon, \tilde{W}_\varepsilon$ convergent faiblement vers μ et $\int_{\mathbb{R}^N}|u_\varepsilon|^2\,dx$ converge vers $\iint_{\mathbb{R}^{2N}} d\mu$ (quand $\varepsilon\to 0_+$). Toujours d'après la Remarque III.6, la construction est explicite dans le cas où μ est purement atomique avec un nombre fini d'atomes *i.e.* $\mu=\sum_{i=1}^n \mu_i\,\delta_{x_i}(x)\,\delta_{\xi_i}(\xi)$ avec $1\le n$, $\mu_i\ge 0$ $(1\le i\le n)$, (x_i,ξ_i) sont des points distincts de $\mathbb{R}^{2N}$ $(1\le i\le u)$. Dans le cas général, il suffit d'approcher une mesure μ positive ou nulle bornée quelconque par une suite μ_n de telles mesures atomiques de façon à ce que μ_n converge étroitement vers μ. On conclut alors par un procédé habituel de suite diagonale.

REMARQUE III.9. (Condition suffisante pour (45)). Une condition simple assurant que (45) a lieu est de supposer

$$\varepsilon^s \|D^s u_\varepsilon\|_{L^2} \leq C, \tag{47}$$

pour un $s > 0$, où C désigne une constante positive indépendante de ε. En effet, ceci signifie que

$$\int_{\mathbb{R}^N} \varepsilon^{2s} |\xi|^{2s} |\hat{u}_\varepsilon(\xi)|^2 \, d\xi \leq C,$$

d'où bien sûr

$$\int_{|\xi|>R/\varepsilon} |\hat{u}_\varepsilon(\xi)|^2 \, d\xi \leq C R^{-2s}.$$

REMARQUE III.10. (Liens avec le calcul pseudo-différentiel). Une autre manière de construire les mesures de Wigner (il s'agit en fait de la construction privilégiée par P. Gérard [18]) est de considérer des opérateurs pseudo-différentiels de symbole $a(x, \varepsilon\xi)$ où $a \in \mathcal{S}(\mathbb{R}^N \times \mathbb{R}^N)$ (par exemple) et d'étudier la limite des quantités quadratiques $\int_{\mathbb{R}^N} (a(x, \varepsilon D) \cdot u_\varepsilon) u_\varepsilon^* \, dx$. En effet, si on note $\hat{u}(x, z) = (\mathcal{F}_\xi a)(x, z)$, on voit que

$$\begin{aligned}
&\int_{\mathbb{R}^N} (a(x, \varepsilon D) \cdot u_\varepsilon) u_\varepsilon^* \, dx \\
&\quad = \iint_{\mathbb{R}^{2N}} u_\varepsilon^*(x) \, (2\pi\varepsilon)^{-N} \, \hat{a}\big(x, \frac{y-x}{\varepsilon}\big) \, u_\varepsilon(y) \, dy \, dx \\
&\quad = (2\pi)^{-N} \iint_{\mathbb{R}^{2N}} \hat{a}(x, z) \, u_\varepsilon(x+z) \, u_\varepsilon^*(x) \, dz \, dx
\end{aligned}$$

et ceci converge quand ε tend vers 0 vers $\iint_{\mathbb{R}^{2N}} a(x, \xi) \, d\mu$ -voir la Remarque III.5.

REMARQUE III.11. (Liens avec les H-mesures) L. Tartar [41] et P. Gérard [18] ont introduit indépendamment une mesure positive ou nulle, bornée σ sur $\mathbb{R}^N_x \times S^{N-1}_\xi$ mesurant les défauts de compacité forte d'une suite u_ε convergeant faiblement dans $L^2(\mathbb{R}^N)$ vers 0. On vérifie aisément que si la mesure de Wigner μ associée à u_ε (ou une sous-suite) ne change pas $\{\xi = 0\}$ et si $\varepsilon^{-N} \left|\hat{u}_\varepsilon(\xi/2)\right|^2$ est étroitement relativement compacte (voir Remarque III.9) alors σ est "la moyenne de μ sur les rayons passant par ξ". En d'autres termes, on a pour tout $\varphi \in C_b(\mathbb{R}^N \times S^{N-1})$

$$\iint_{\mathbb{R}^N \times S^{N-1}} \varphi \, d\sigma = \iint_{\mathbb{R}^{2N}} \tilde{\varphi} \, d\mu,$$

où $\tilde{\varphi}(x,\xi) \in C_b(\mathbb{R}^N \times (\mathbb{R}^N - \{0\}))$ est définie par $\tilde{\varphi}(x,\xi) = \varphi(x, \xi/|\xi|)$. On voit donc que la H-mesure est moins précise que la mesure de Wigner mais que par contre elle ne nécessite pas de connaître la taille des oscillations de la suite u_ε.

D'autre part, les constructions (variées) que nous avons établies pour la mesure de Wigner permettent de donner des procédés apparemment nouveaux de construction des H-mesures. En effet, pour une suite $(u_\varepsilon)_\varepsilon$ convergeant faiblement vers 0 dans $L^2(\mathbb{R}^N)$, on peut former

$$H_\varepsilon(x,\xi) = \int_0^\infty t^{n-1}\, W(u^\varepsilon)(x,t,\xi)\, dt\,, \tag{48}$$

pour tout $(x,\xi) \in \mathbb{R}^N \times S^{N-1}$, où $W(u^\varepsilon)$ est la transformée de Wigner associée à u^ε, *i.e.*

$$W(u^\varepsilon) \;=\; (2\pi)^{-N}\, \mathcal{F}_y \left(u^\varepsilon\big(x + \frac{y}{2}\big)\, u^\varepsilon\big(x - \frac{y}{2}\big)\right).$$

On démontre alors aisément que $W(u_\varepsilon)$ converge faiblement dans $\mathcal{S}'(\mathbb{R}^N \times S^{N-1})$ vers la H-mesure σ de u_ε (après extraction éventuelle d'une sous-suite). On peut d'ailleurs retrouver la positivité de σ en montrant que σ est la limite faible (au sens des mesures sur $\mathbb{R}^N \times S^{N-1}$) de quantités faisant intervenir des transformées de Husimi convenables. Nous ne développerons pas plus ici ce type d'approche pour les H-mesures.

Signalons maintenant une propriété de localisation des mesures de Wigner lorsque u_ε satisfait une équation aux dérivées partielles du type:

$$P_\varepsilon\big(x\,,\, \varepsilon D u_\varepsilon\big) \;=\; f_\varepsilon\,, \tag{49}$$

où $f_\varepsilon \underset{\varepsilon}{\to} 0$ dans $L^2(\mathbb{R}^N)$ et u_ε est bornée dans $L^2(\mathbb{R}^N)$. On suppose que $P_\varepsilon(x,\xi) \in C^\infty(\mathbb{R}^N_x \times \mathbb{R}^N_\xi)$ est polynomiale en ξ de degré borné, *i.e.*

$$P_\varepsilon(x,\xi) \;=\; \sum_{|\alpha| \le M} a^\varepsilon_\alpha(x)\, \xi^\alpha\,, \qquad a^\varepsilon_\alpha \in C^\infty(\mathbb{R}^N_x)\,. \tag{50}$$

On suppose que a^ε_α converge vers a^0_α (par exemple) dans $C^\infty_{\mathrm{loc}}(\mathbb{R}^N_x)$ et que $W_\varepsilon(u_\varepsilon)$ converge (faiblement dans $\mathcal{A}'_{w-*}$) vers μ. Alors, on démontre facilement que $P_0\mu = 0$ sur $\mathbb{R}^{2N}$, en notant bien sûr

$$P_0(x,\xi) = \sum_{|\alpha| \le M} a^0_\alpha(x)\, \xi^\alpha\,.$$

Cette observation est bien sûr à rapprocher des propriétés analogues connues pour les H-mesures (voir L. Tartar [41], P. Gérard [18]). On peut bien sûr notablement affaiblir les hypothèses de régularité sur les coefficients -comme nous le ferons dans la section suivante pour un problème relié- mais nous n'aborderons pas ce point technique (mais important) ici. On peut aussi considérer des équations pseudo-différentielles à la place de (49).

Egalement, comme pour les H-mesures, il est possible d'obtenir une équation de transport sur μ si $f_\varepsilon/\varepsilon \underset{\varepsilon}{\to} 0$ dans $L^2(\mathbb{R}^N)$ et si P_0 est réel -il faut en outre préciser le comportement de P_ε quand ε tend vers 0-, mais nous nous contenterons d'aborder ce thème sur l'exemple de la limite semi-classique dans des équations de Schrödinger. C'est l'objet de la section suivante. De façon générale, indiquons que les résultats établis dans L. Tartar [41] et P. Gérard [18] pour les H-mesures s'adaptent aux mesures de Wigner.

Nous voulons maintenant conclure cette section en généralisant l'étude précédente à des suites d'objets plus "complexes" que des suites de fonctions scalaires dans $L^2(\mathbb{R}^N)$. Une première généralisation consiste à considérer des suites u_ε bornées dans $L^2(\mathbb{R}^N;\mathbb{R}^m)$ (on pourrait également considérer le cas de fonctions prenant leurs valeurs dans un Hilbert séparable). Tout ce que nous avons fait précédemment s'adapte aisément en considérant les limites faibles μ_{ij} de $W_\varepsilon((u_\varepsilon)_i,(u_\varepsilon)_j)$ qui sont des mesures bornées sur $\mathbb{R}^N$ et la matrice symétrique $(\mu_{ij})_{ij}$ est positive.

La généralisation que nous voulons développer consiste à introduire comme dans la section précédente des matrices densité ρ_ε et nous considérons dans tout le reste de cette section une suite bornée de matrices densité $(\rho_\varepsilon)_\varepsilon$ (avec le même abus de langage sur la signification réelle de ε). Rappelons que ceci veut dire que ρ_ε est bornée dans $L^2(\mathbb{R}^N\times\mathbb{R}^N)$, définit une suite d'opérateurs hermitiens $\big(\rho_\varepsilon(x,y)=\rho_\varepsilon(y,x)^*\big)$ positifs ou nuls et que leurs traces sont uniformément majorées $(\rho_\varepsilon(x)=\rho_\varepsilon(x,x)\geq 0$ est bornée dans $L^1(\mathbb{R}^N))$. On introduit alors les transformées de Wigner sur $\mathbb{R}^N_x\times\mathbb{R}^N_\xi$

$$
\begin{aligned}
W_\varepsilon(x;\xi) &= \frac{1}{(2\pi\varepsilon)^N}\int_{\mathbb{R}^N} e^{-i(\xi/\varepsilon)\cdot y}\,\rho_\varepsilon\big(x+\frac{y}{2},x-\frac{y}{2}\big)\,dy \\
&= \frac{1}{(2\pi)^N}\int_{\mathbb{R}^N} e^{-i\xi\cdot y}\,\rho_\varepsilon\big(x+\frac{\varepsilon y}{2},x-\frac{\varepsilon y}{2}\big)\,dy
\end{aligned}
\tag{51}
$$

ainsi que les transformées de Husimi également définies sur $\mathbb{R}^N_x \times \mathbb{R}^N_\xi$

$$\tilde{W}_\varepsilon(x,\xi) = \frac{1}{(\pi\varepsilon)^N} W_\varepsilon * e^{-(|x|^2+|\xi|^2)/\varepsilon} . \tag{52}$$

On rappelle (*cf.* Section II) que $\tilde{W}_\varepsilon \geq 0$ est bornée dans $L^1(\mathbb{R}^N_x \times \mathbb{R}^N_\xi)$ car

$$\iint_{\mathbb{R}^{2N}} \tilde{W}_\varepsilon \, dx \, d\xi = \mathrm{Tr}\,(\rho_\varepsilon) = \iint_{\mathbb{R}^N} \rho_\varepsilon(x) \, dx .$$

De plus, la Proposition III.1 s'adapte facilement et montre que W_ε est bornée dans $\mathcal{A}'$.

On peut donc extraire une sous-suite (encore notée ε) telle que $\rho_\varepsilon(x,y)$ converge faiblement dans $L^2(\mathbb{R}^N \times \mathbb{R}^N)$ vers ρ (qui définit bien sûr un opérateur hermitien, positif ou nul, de trace finie), $\rho_\varepsilon(x)$ converge faiblement au sens des mesures (sur $\mathbb{R}^N$) vers μ_x, W_ε converge faiblement dans $\mathcal{A}'_{w-*}$ vers μ et $\tilde{W}_\varepsilon$ converge faiblement au sens des mesures (sur $\mathbb{R}^{2N}$) vers $\tilde{\mu}$ (qui est donc une mesure bornée positive ou nulle sur $\mathbb{R}^{2N}$). On démontre alors, de la même manière que le Théorème III.1, le

Théorème III.2.

1) *On a:* $\mu \equiv \tilde{\mu}$. *En particulier,* $\mu \in \mathcal{M}(\mathbb{R}^{2N})$.

2) *L'inégalité suivante a lieu, en notant* $\rho(x) = \rho(x,x)$*:*

$$\mu \geq \rho(x)\,\delta_0(\xi) \tag{53}$$

d'où en particulier

$$\int_{\mathbb{R}^N} \rho(x)\,dx \leq \int_{\mathbb{R}^{2N}} d\mu \leq \liminf_{\varepsilon\to 0} \int_{\mathbb{R}^N} \rho_\varepsilon(x)\,dx . \tag{54}$$

3) *L'égalité* $\mu_x = \int_{\mathbb{R}^N_\xi} d\mu(\cdot,\xi)$ *a lieu si* $\varepsilon^{-N}\,\hat{\rho}_\varepsilon\big(\xi/\varepsilon,\xi/\varepsilon\big)$ *est une suite étroitement relativement compacte dans* $\mathcal{M}(\mathbb{R}^N)$*, où*

$$\hat{\rho}_\varepsilon(\xi,\eta) = (\mathcal{F}_x\,\overline{\mathcal{F}}_y)\,\rho_\varepsilon(x,y) .$$

Si $(2\pi\varepsilon)^{-N}\,\hat{\rho}\big(\xi/\varepsilon,\xi/\varepsilon\big)$ *(ou une sous-suite) converge faiblement au sens des mesures vers une mesure* μ_ε *alors*

$$\mu_x \geq \int_{\mathbb{R}^N_\xi} d\mu(\cdot,\xi) \qquad et \qquad \mu_x \geq \int_{\mathbb{R}^N_x} d\mu(x,\cdot) .$$

4) *L'égalité*

$$\iint_{\mathbb{R}^{2N}} d\mu = \lim_{\varepsilon \to 0} \int_{\mathbb{R}^N} \rho_\varepsilon(x)\, dx$$

a lieu si et seulement si $\rho_\varepsilon(x)$ *et* $\varepsilon^{-N}\, \hat{\rho}(\xi/\varepsilon, \xi/\varepsilon)$ *sont étroitement relativement compactes. En particulier, si cette hypothèse est vérifiée,* ρ_ε *converge dans* $L^2(\mathbb{R}^{2N})$ *vers* ρ *si et seulement si* $\mu = \rho(x)\, \delta_0(\xi)$.

REMARQUE III.12. Rappelons que $\int_{\mathbb{R}^N} \rho_\varepsilon(x)\, dx = \operatorname{Tr}(\rho_\varepsilon)$. D'autre part, il est clair que $\hat{\rho}_\varepsilon \in L^2(\mathbb{R}^N \times \mathbb{R}^N)$ définit un opérateur hermitien, positif ou nul, de trace finie ($\operatorname{Tr}(\hat{\rho}_\varepsilon) = (2\pi)^N \operatorname{Tr}(\rho_\varepsilon)$) et donc $\hat{\rho}_\varepsilon(\xi, \xi) \geq 0$, $(2\pi\varepsilon)^{-N}\, \hat{\rho}(\xi/\varepsilon, \xi/\varepsilon)$ est bornée dans $L^1_+(\mathbb{R}^N)$. En particulier, μ_ξ est donc une mesure bornée sur $\mathbb{R}^N$.

REMARQUE III.13. On pourrait également donner l'analogue du point 5) du Théorème III.1 ou des Remarques III.6, III.7, III.8 ou III.10. Nous ne le ferons évidemment pas mais nous nous contenterons de signaler que l'analogue de (47) est

$$\varepsilon^s \operatorname{Tr}\left(H_0^{s/2}\, \rho_\varepsilon\right) \leq C\,, \tag{55}$$

où $H_0 = -\frac{1}{2}\Delta$. Ceci peut se réécrire plus simplement en diagonalisant ρ_ε :

$$\rho_\varepsilon = \sum_i \lambda_i^\varepsilon\, \psi_i^\varepsilon \quad \text{avec} \quad \lambda_i^\varepsilon \geq 0\,, \quad \int_{\mathbb{R}^N} \psi_i^\varepsilon\, \psi_j^{\varepsilon *}\, dx = \delta_{ij}\,.$$

Auquel cas, (55) devient

$$\sum_i \lambda_i^\varepsilon\, \varepsilon^s \int_{\mathbb{R}^N} |D^s \psi_i^\varepsilon|^2\, dx \leq C\,. \tag{55'}$$

Nous allons maintenant conclure cette section en donnant quelques exemples de noyaux ρ_ε qui approchent des mesures arbitraires sur $\mathbb{R}^{2N}$.

EXEMPLE III.6. (Mélange d'états WKB). On considère

$$\rho_\varepsilon = \int_{\mathbb{R}^N} u_\varepsilon(x)\, u_\varepsilon^*(y)\, e^{-i\eta\cdot(x-y)}\, dm(\eta)\,, \tag{56}$$

où $u_\varepsilon(x) = u(x)\, e^{i\, a(x)/\varepsilon}$, $u \in L^2(\mathbb{R}^N)$, $a \in W^{1,1}_{\mathrm{loc}}(\mathbb{R}^N)$, $m \in \mathcal{M}(\mathbb{R}^N)$. Alors, on vérifie que $\mu = |u(x)|^2\, dm\,(\nabla a(x) - \xi)$.

EXEMPLE III.7 (Approximation d'une mesure $\mu \in \mathcal{M}(\mathbb{R}^{2N})$ par un mélange d'états cohérents). Soit $\mu \in \mathcal{M}(\mathbb{R}^{2N})$ et soit $u_0 \in Lr(\mathbb{R}^N)$ (ou $C_0^\infty(\mathbb{R}^N)$) tel que $\|u\|_{L^2} = 1$. On pose

$$\rho_\varepsilon = \iint_{\mathbb{R}^{2N}} \varepsilon^{-N/2}\, u_0\Big(\frac{x-x_0}{\varepsilon^{1/2}}\Big)\, u_0^*\Big(\frac{y-x_0}{\varepsilon^{1/2}}\Big) \cdot e^{i(\xi_0/\varepsilon)\cdot(x-y)}\, d\mu(x_0,\xi_0)\,. \tag{57}$$

On vérifie que

$$W_\varepsilon = u * W_\varepsilon^0\,, \qquad \text{où } W_\varepsilon^0 = \frac{1}{\varepsilon^N}\, W^0\Big(\frac{x-x_0}{\varepsilon^{1/2}}, \frac{\xi-\xi_0}{\varepsilon^{1/2}}\Big)\,, \tag{58}$$

et

$$W^0(x,\xi) = (2\pi)^{-N} \int_{\mathbb{R}^N} e^{-i\xi\cdot y}\, u_0(x+y/2)\, u_0^*(x-y/2)\, dy\,.$$

Et on déduit aisément de (58) que W_ε converge vers μ quand ε tend vers 0 dans $\mathcal{A}'_{w-*}$ et même dans $\mathcal{M}(\mathbb{R}^{2N}$ si $W^0 \in L^1(\mathbb{R}^{2N})$ (ce qui est le cas bien sûr si $u_0 \in \mathcal{S}(\mathbb{R}^N)$).

Dans le cas où μ admet une densité par rapport à la mesure de Lebesgue (par exemple) encore notée μ, on peut d'une part établir des "inégalités de convexité" liant ρ_ε à μ et d'autre part préciser la convergence de Wigner μ. En effet, voir W. Thirring [43] par exemple, on a

$$\mathrm{Tr}\Big((2\pi\varepsilon)^N\, F\big(\frac{\rho_\varepsilon}{(2\pi\varepsilon)^N}\big)\Big) \geq \iint_{\mathbb{R}^{2N}} F(\mu)\, dx\, d\xi\,, \tag{59}$$

pour toute fonction F convexe, positive ou nulle (par exemple), nulle en 0. On admet bien sûr dans (59) la possibilité que le membre de gauche ou que les deux membres valent $+\infty$. Cette inégalité se démontre aisément en diagonalisant ρ_ε

$$\rho_\varepsilon = \sum_j \lambda_j\, \psi_j(x)\, \psi_j^*(y) \quad \text{avec } \lambda_j \geq 0\,, \qquad \int_{\mathbb{R}^N} \psi_j \psi_j^* = \delta_{jk}\,,$$

(et on peut toujours supposer que la famille $(\psi_j)_j$ définit une base orthonormée de $L^2(\mathbb{R}^N)$). En effet, on a alors

$$\mathrm{Tr}\Big(F\big((2\pi\varepsilon)^{-N}\rho_\varepsilon\big)\Big) = \sum_j F\big((2\pi\varepsilon)^{-N}\lambda_j\big)$$

et

$$\lambda_j = \iint_{\mathbb{R}^{2N}} \mu(x_0,\xi_0) \left|\left(u_\varepsilon^{x_0,\xi_0},\psi_j\right)_{L^2}\right|^2 dx_0\, d\xi_0$$

en notant

$$u_\varepsilon^{x_0,\xi_0} = \varepsilon^{-N/4} u_0\left(\frac{x-x_0}{\varepsilon^{1/2}}\right) e^{i(\xi_0/\varepsilon)\cdot x}\,.$$

Il suffit alors d'observer que pour tout $\psi \in L^2(\mathbb{R}^N)$

$$\iint_{\mathbb{R}^{2N}} \left|\left(u_\varepsilon^{x_0,\xi_0},\psi\right)_{L^2}\right|^2 dx_0\, d\xi_0 = (2\pi\varepsilon)^N \|\psi\|_{L^2}^2\,, \tag{60}$$

car

$$\int_{\mathbb{R}^N} \left|\left(u_\varepsilon^{x_0,\xi_0},\psi\right)_{L^2}\right|^2 d\xi_0 = (2\pi\varepsilon)^N \int_{\mathbb{R}^N} \varepsilon^{-N/2} \left|u\left(\frac{x-x_0}{\varepsilon^{1/2}}\right)\right|^2 |\psi(x)|^2\, dx\,.$$

On voit donc que

$$\sum_j F((2\pi\varepsilon)^{-N}\lambda_j) \geq \sum_j \iint_{\mathbb{R}^{2N}} F(\mu)(x_0,\xi_0) \left|\left(u_\varepsilon^{x_0,\xi_0},\psi_j\right)_{L^2}\right|^2 (2\pi\varepsilon)^N\, dx_0\, d\xi_0$$

et

$$\sum_j \left|\left(u_\varepsilon^{x_0,\xi_0},\psi_j\right)_{L^2}\right|^2 = \left\|u_\varepsilon^{x_0,\xi_0}\right\|_{L^2}^2 = 1\,.$$

D'autre part, il est clair que si $W_0 \in L^1(\mathbb{R}^{2N})$ et si $\mu \in L^p(\mathbb{R}^{2N})$ avec $1 \leq p \leq \infty$ alors W_ε est bornée dans $L^p(\mathbb{R}^{2N})$ et converge dans $L^p(\mathbb{R}^{2N})$ vers μ (si $p < \infty$). De plus, si W_0 est positive ou nulle -c'est le cas si $u_0 = (\pi)^{-N/4} e^{-|x|^2/2}$ puisqu'on a alors $W_0(x,\xi) = \pi^{-N} e^{-(|x|^2+|\xi|^2)}$-, on voit que l'on a aussi

$$\iint_{\mathbb{R}^{2N}} F(W_\varepsilon)\, dx\, d\xi \leq \iint_{\mathbb{R}^{2N}} F(\mu)\, dx\, d\xi\,, \tag{61}$$

pour toute fonction F convexe, positive ou nulle sur $[0,+\infty[$, nulle en 0.

Signalons pour conclure des "inégalités de convexité" liant une suite bornée de matrices-densité $(\rho_\varepsilon)_\varepsilon$ à leurs transformées de Husimi.

Nous allons les énoncer pour des transformées en fait plus générales. En effet, à partir de W_ε, on peut considérer

$$\tilde{W}_\varepsilon^\mu = W_\varepsilon * W_\varepsilon^0, \qquad \text{où } W_\varepsilon^0 = \frac{1}{\varepsilon^N} W^0\left(\frac{\cdot}{\varepsilon^{1/2}}, \frac{\cdot}{\varepsilon^{1/2}}\right) \tag{62}$$

et

$$W^0 = (2\pi)^{-N} \int_{\mathbb{R}^N} e^{i\xi\cdot y}\, u(-x+y/2)\, u^*(-x-y/2)\, dy,$$

où $u \in L^2(\mathbb{R}^N)$ (ou $\mathcal{S}, C_0^\infty$) et $\|u\|_{L^2} = 1$. Noter que

$$W_\varepsilon^0(-x,-\xi) = (2\pi\varepsilon)^{-N} \int_{\mathbb{R}^N} e^{-i(\xi/\varepsilon)\cdot y} \left(\varepsilon^{-N/4}\, u\Big(\frac{x+y/2}{\varepsilon^{1/2}}\Big)\right) \cdot \left(\varepsilon^{-N/4}\, u\Big(\frac{x-y/2}{\varepsilon^{1/2}}\Big)\right)^* dy$$

et que

$$\tilde{W}_\varepsilon^u(x,\xi) = (2\pi\varepsilon)^{-N} \big(\rho_\varepsilon \cdot u_\varepsilon^{x,\xi}, u_\varepsilon^{x,\xi}\big)_{L^2},$$

avec

$$u_\varepsilon^{x,\xi}(z) = \varepsilon^{-N/4}\, u\Big(\frac{z-x}{\varepsilon^{1/2}}\Big)\, e^{i(\xi/\varepsilon)\cdot z}.$$

En particulier (voir également la Section II), $\tilde{W}_\varepsilon^u \geq 0$ sur $\mathbb{R}^{2N}$ et $\tilde{W}_\varepsilon^u \in L^1(\mathbb{R}^{2N})$. Enfin, $\tilde{W}_\varepsilon$ correspond au choix particulier de

$$u = (\pi)^{-N/4}\, e^{-|x|^2/2}.$$

On démontre alors (voir W. Thirring [43] pour l'inégalité (64)) que pour toute fonction convexe F, positive ou nulle sur $[0, +\infty[$, nulle en 0,

$$\big(F((2\pi\varepsilon)^{-N}\rho_\varepsilon)\cdot u_\varepsilon^{x,\xi},\, u_\varepsilon^{x,\xi}\big)_{L^2} \geq F(\tilde{W}_\varepsilon^u)(x,\xi), \tag{63}$$

pour tout (x,ξ). En effet, en diagonalisant ρ_ε (comme précédemment), on a

$$\begin{aligned}\big(F((2\pi\varepsilon)^{-N}\rho_\varepsilon)u_\varepsilon^{x,\xi},\, u_\varepsilon^{x,\xi}\big)_{L^2} &= \sum_j F\Big(\frac{\lambda_j}{(2\pi\varepsilon)^N}\Big)\left|\big(u_\varepsilon^{x,\xi}, \psi_j\big)_{L^2}\right|^2 \\ &\geq F\left(\sum_j \frac{\lambda_j}{(2\pi\varepsilon)^N}\left|\big(u_\varepsilon^{x,\xi}, \psi_j\big)_{L^2}\right|^2\right)\end{aligned}$$

$$= F\Big((2\pi\varepsilon)^N \big(\rho_\varepsilon \cdot u_\varepsilon^{x,\xi}, u_\varepsilon^{x,\xi}\big)_{L^2}\Big)$$
$$= F(\tilde{W}_\varepsilon^u)(x,\xi)\,,$$

l'inégalité étant due au fait que $\sum_j \big|(u_\varepsilon^{x,\xi}, \psi_j)_{L^2}\big|^2 = \|u_\varepsilon^{x,\xi}\|_{L^2}^2 = 1$. Et, d'après (60), on déduit de (63) en intégrant par rapport à $(x,\xi) \in \mathbb{R}^{2N}$

$$\text{(64)} \qquad \mathrm{Tr}\left((2\pi\varepsilon)^N F((2\pi\varepsilon)^{-N}\rho_\varepsilon)\right) \geq \iint_{\mathbb{R}^{2N}} F(\tilde{W}_\varepsilon^u)\,dx\,d\xi\,,$$

inégalité valable pour tout $u \in L^2(\mathbb{R}^N)$ et donc en particulier pour $\tilde{W}_\varepsilon$. Enfin les inégalités suivantes ont lieu:

$$\begin{aligned} \tilde{F(\rho)} &\leq F(\tilde{\rho})\,, \qquad \text{si } F \text{ est concave,} \\ &\geq F(\tilde{\rho})\,, \qquad \text{si } F \text{ est convexe.} \end{aligned}$$

En effet, décomposant ρ sur une base orthogonale,

$$\rho = \sum_i \lambda_i |\psi_i >< \psi_i|\,,$$

ce qui signifie

$$\rho(x,y) = \sum_i \lambda_i \psi_i(x)\psi_i^*(y)\,,$$

on a $F(\rho) = \sum_i F(\lambda_i)|\psi_i >< \psi_i|$ et donc $\tilde{F}(\rho) = \sum_i F(\lambda_i)|(u_{\varepsilon=1}^{x,\xi}, \psi_i)|^2$ avec $\sum_i |(u_{\varepsilon=1}^{x,\xi}, \psi_i)|^2 = 1$.

IV. Limite semi-classique.

Nous allons étudier dans cette section la limite quand $\hbar$ tend vers 0 des transformées de Wigner des solutions de l'équation de Schrödinger (6) ou de l'équation de Liouville associée (10). Et nous traiterons aussi bien le cas linéaire (où le potentiel V est donné) que le cas nonlinéaire (où le potentiel V dépend de la densité $\rho(x)$: $V = V_0 * \rho$ et V_0 est alors un potentiel donné).

Comme nous l'avons indiqué dans l'Introduction, les limites "doivent" vérifier l'équation de Liouville classique (11) qui dans le cas nonlinéaire où $V = V_0 * \rho$ est en fait l'équation de Vlasov.

Avant d'indiquer l'organisation de cette section, il est utile, d'expliquer au moins formellement l'obtention de ces équations de transport classiques (au sens de la Mécanique Classique). On considère donc une solution $\psi^\hbar$ dans $C(\mathbb{R}_t; L^2(\mathbb{R}^N))$ de (6) ou une matrice-densité $\rho^\hbar$ solution de (10) qui est donc de la forme

$$\rho^\hbar(t) = \sum_j \lambda_j \, \psi^\hbar(t,x) \, \psi_j^\hbar(t,y)^* \, ,$$

avec $\lambda_j \geq 0$, $\int_{\mathbb{R}^N} \psi_j^\hbar \, \psi_k^{\hbar *} \, dx = \delta_{jk}$, pour tout j,k, et pour tout $t \in \mathbb{R}$, et $\psi_j^\hbar \in C(\mathbb{R}_t; L^2(\mathbb{R}^N))$. Comme dans la Section III et en choisissant $\varepsilon = \hbar$, on introduit $f^\hbar = W_\hbar(\psi^\hbar)$ ou $W_\hbar(\rho^\hbar)$

$$\begin{aligned} & f^\hbar(t,x,\xi) \\ & = (2\pi\hbar)^{-N} \int_{\mathbb{R}^N} e^{-i(\xi/\hbar)\cdot y} \, \psi_\hbar\big(t, x + \frac{y}{2}\big) \, \psi_\hbar\big(t, x - \frac{y}{2}\big)^* \, dy \\ & = (2\pi)^{-N} \int_{\mathbb{R}^N} e^{-i\xi\cdot y} \, \psi_\hbar\big(t, x + \frac{\hbar y}{2}\big) \, \psi_\hbar\big(t, x - \frac{\hbar y}{2}\big)^* \, dy \, , \end{aligned} \tag{65}$$

ou

$$\begin{aligned} f^\hbar(t,x,\xi) & = (2\pi\hbar)^{-N} \int_{\mathbb{R}^N} e^{-i(\xi/\hbar)\cdot y} \, \rho_\hbar\big(t \, , \, x + \frac{y}{2} \, , \, x - \frac{y}{2}\big) \, dy \\ & = (2\pi)^{-N} \int_{\mathbb{R}^N} e^{-i\xi\cdot y} \, \rho_\hbar\big(t \, , \, x + \frac{\hbar y}{2} \, , \, x - \frac{\hbar y}{2}\big) \, dy \, , \end{aligned} \tag{66}$$

pour $t \in \mathbb{R}$, $(x,\xi) \in \mathbb{R}^{2N}$.

L'analogue de la Proposition III.1 a lieu et implique que $f^\hbar$ résout dans $\mathbb{R}_t \times \mathbb{R}^{2N}_{x,\xi}$

$$\frac{\partial f^\hbar}{\partial t} + \xi \cdot \nabla_x f^\hbar + K_\hbar \underset{\xi}{*} f^\hbar \; = \; 0 \, , \tag{67}$$

où

$$K_\hbar \; = \; \frac{i}{(2\pi)^N} \int_{\mathbb{R}^N} e^{-i\xi\cdot y} \, \hbar^{-1} \Big(V\big(x + \frac{\hbar y}{2}\big) - V\big(x - \frac{\hbar y}{2}\big) \Big) \, dy \, .$$

On voit donc que, au moins formellement, quand $\hbar$ tend vers 0_+, $K_\hbar$ "converge" vers

$$K_0 = \frac{i}{(2\pi)^N} \nabla V(x) \cdot \mathcal{F}(y) = -\nabla V(x) \, \nabla \delta_0(\xi)$$

et que l'on peut s'attendre à ce que la limite de $f_\hbar$ résolve (11).

Les résultats que nous démontrons dans cette section concernent précisément la justification de cette limite formelle avec, en outre, une analyse de la convergence de $f^\hbar$ vers sa ou ses limites f. Bien sûr, $\psi^\hbar$ ou $\rho^\hbar$ seront les solutions de (6) et de (10) correspondant à des conditions initiales

$$\psi^\hbar|_{t=0} = \psi_0^\hbar, \qquad \text{dans } \mathbb{R}^N, \tag{68}$$

ou

$$\rho^\hbar|_{t=0} = \rho_0^\hbar, \qquad \text{dans } \mathbb{R}^{2N}, \tag{69}$$

et $\psi_0^\hbar, \rho_0^\hbar$ sont des suites bornées de $L^2(\mathbb{R}^N)$ ou de matrices-densité respectivement. On supposera toujours que $f_0^\hbar = W_\hbar(\psi_0^\hbar)$ ou $W_\hbar(\rho^\hbar)$ converge dans $\mathcal{A}'_{w-*}$, après extraction éventuelle d'une sous-suite, vers une mesure positive ou nulle, bornée notée f_0.

Nos objectifs seront d'établir que $f^\hbar$ converge (en un sens à préciser) vers f solution de (11), vérifiant la condition initiale

$$f|_{t=0} = f_0, \qquad \text{dans } \mathbb{R}^{2N}. \tag{70}$$

D'après la forme même de l'équation limite (11) où apparaît le terme $\nabla V(x)\nabla_\xi f$, il n'est pas surprenant d'obtenir des résultats de nature un peu différente suivante que l'on suppose que V est "régulier" ou non, et dans ce dernier cas il est alors naturel de requérir plus de "régularité" sur f c'est-à-dire plus de "régularité" sur $\rho^\hbar$ et donc sur $\rho_0^\hbar$.

Les premiers résultats que nous donnerons exigerons "un peu de régularité" sur V. Nous démontrerons ensuite que si toutes les données sont régulières, un développement asymptotique en puissances de $\hbar$ est possible. Enfin, nous aborderons le cas de potentiels moins réguliers. Dans tous les résultats concernant la limite $\hbar \to 0_+$, nous présenterons tout d'abord le cas linéaire puis le cas nonlinéaire, et sauf mention explicite nous traiterons toujours le cas de l'équation de Liouville (10), plus général que celui de (6).

Théorème IV.1. (Cas linéaire). *Nous supposons que V vérifie* (34) *et* (35).

1) *Si $V \in C^1(\mathbb{R}^N$, alors $f^\hbar$, après extraction éventuelle d'une sous-suite, converge uniformément sur tout compact de $\mathbb{R}_t$ dans $\mathcal{A}'w-*$ vers $f \in C_b(\mathbb{R}_t, \mathcal{M}w-*)$ qui vérifie* (11) *au sens des distributions et* (70).

2) *Si de plus $V \in C^{1,1}(\mathbb{R}^N)$ et si V vérifie*

$$(71) \qquad \begin{cases} \text{il existe } C \geq 0\,, \text{ pour tout } x \in \mathbb{R}^N\,, \\ V(x) \geq -C(1+|x|^2)\,, \end{cases}$$

alors f est l'unique solution de (11) *et* (70) *dans $C_b(\mathbb{R}_t, \mathcal{M}w-*)$ et f est donnée par le transport de f_0 par le flot Hamiltonien ($\dot{x} = \xi$, $\dot{\xi} = -\nabla V(x)$).*

REMARQUE IV.1. La convergence uniforme sur $[-T,+T]$ dans $\mathcal{A}_{w-*}$ signifie que, pour tout ψ dans $\mathcal{A}'$, $\iint_{\mathbb{R}^{2N}} f^\hbar\,\psi\,dx\,d\xi$ converge uniformément sur $[-T,+T]$ vers $\iint_{\mathbb{R}^{2N}} \psi\,df(t)$. De même, $f \in C_b(\mathbb{R}_t\,,\,\mathcal{M}w-*)$ signifie que $\iint_{\mathbb{R}^{2N}} \psi\,df(t) \in C(\mathbb{R}_t)$ pour tout $\psi \in C_0(\mathbb{R}^{2N})$ et que $f(t)$ est bornée dans $\mathcal{M}$. On peut également considérer que $f(t)$ appartient pour tout $t \in \mathbb{R}$ à une boule de $\mathcal{M}$ dans laquelle la topologie faible $*$ est métrisable et que f est continue en t à valeurs dans cet espace métrique.

L'écriture de (11) (au sens des distributions) consiste à écrire les termes $\xi \cdot \nabla_x f$ et $\nabla V(x) \cdot \nabla_\xi f$ sous forme conservative, *i.e.* $\mathrm{div}_x(\xi f)$ et $\mathrm{div}_\xi(\nabla V(x) f)$ respectivement.

En ce qui concerne 2), l'hypothèse (71) garantit que le flot Hamiltonien H_t est bien défini pour tout $t \in \mathbb{R}$ sur $\mathbb{R}^{2N}$ et que pour tout $\psi \in C_b(\mathbb{R}^N)$

$$(72) \qquad \iint_{\mathbb{R}^{2N}} \psi\,df(t) = \iint_{\mathbb{R}^{2N}} \psi \circ H_t\,df_0\,.$$

En particulier, $\iint_{\mathbb{R}^{2N}} df(t) = \iint_{\mathbb{R}^{2N}} df_0$, de sorte que si

$$\iint_{\mathbb{R}^{2N}} f_0^\hbar\,dx\,d\xi = \iint_{\mathbb{R}^{2N}} \tilde{f}_0^\hbar\,dx\,d\xi = \mathrm{Tr}\,(\rho_0^\hbar) \underset{\hbar \to 0_+}{\to} \iint_{\mathbb{R}^{2N}} df_0$$

(en d'autres termes, $\tilde{f}_0^\hbar$ converge étroitement vers f_0) alors

$$\iint_{\mathbb{R}^{2N}} f^\hbar(t)\,dx\,d\xi = \iint_{\mathbb{R}^{2N}} \tilde{f}^\hbar(t)\,dx\,d\xi = \mathrm{Tr}\,(\varphi^\hbar(t)) = \mathrm{Tr}\,(\rho_0^\hbar)$$

converge donc (uniformément sur $\mathbb{R}_t$) vers $\int\int_{\mathbb{R}^{2N}} df(t)$ (et donc, $\tilde{f}^\hbar$ converge étroitement vers $f(t)$ uniformément sur tout compact de $\mathbb{R}_t$).

Enfin, l'unicité de $f(t)$ montre également que si toute la suite $f_0^\hbar$ converge vers f_0 dans $\mathcal{A}'_{w-*}$ alors toute la suite $f^\hbar(t)$ converge vers $f(t)$ dans $\mathcal{A}'_{w-*}$ uniformément sur tout compact de $\mathbb{R}_t$.

REMARQUE IV.2. Les hypothèses (34) et (35) ne sont pas fondamentales pour le résultat précédent. L'hypothèse (34) ne sert qu'à assurer l'existence de $\rho_\hbar(t)$ et peut donc être supprimée si on suppose que $\rho_\hbar(t)$ existe. Enfin, (35) ne sert qu'à formuler directement et simplement l'équation (de Wigner) (36) et peut donc être supprimée à condition d'interpréter (36) de manière un peu différente.

REMARQUE IV.3. Il est important de noter que l'hypothèse faite sur V à savoir $V \in C^1$ ne garantit pas l'unicité du flot Hamiltonien $\dot{x} = \xi$, $\dot{\xi} = -\nabla V(x)$. En fait, on peut donner des exemples de non-unicité de $f(t)$ (*i.e.* des limites de $f^\hbar(t)$) même si toute la suite $f_0^\hbar$ converge vers f_0. Par exemple, on peut choisir $N = 1$ et

$$V(x) = -|x|^{\theta+1} + \beta(x),$$

où $\theta \in]0,1[$, $\beta \equiv 0$ sur $[1,+1]$, $\beta(x) \equiv |x|^2$ si $|x| \geq 2$, $\beta \in C^\infty(\mathbb{R} - \{0\})$, β paire, $\beta' \geq 0$ pour $x \geq 0$. Alors, en choisissant $f_0 = \delta_{(0,0)}$ $(= \delta_0(x)\delta_0(\xi))$, plusieurs solutions de (11) et (70) existent comme par exemple

$$f_\pm(t) = \delta_{x_\pm(t)}(x)\, \delta_{\dot{x}_\pm(t)}(\xi),$$

avec pour $|t| \leq c_0^{-1/\mu}$,

$$\dot{x}_\pm(t) = \pm\, c_0\, |t|^{\mu-1}\, t,$$

avec

$$c_0 = \left(\frac{2(1+\theta)}{(1-\theta)^2}\right)^{-1/(1-\theta)}, \qquad \mu = \frac{2}{1-\theta}.$$

Ces deux solutions peuvent être approchées par une suite $f^\hbar(t)$ convenable. Il suffit pour cela de montrer l'existence de suites $f_0^\hbar$ convergeant vers f_0 telles que $f^\hbar(t)$ converge vers $f_+(t)$ ou vers $f_-(t)$. On obtient la non-unicité annoncée en alternant les deux suites ainsi construites. Dans le cas de $f_+(t)$ (par exemple), on observe que si on choisit x_ε de l'ordre de ε (petit) et ξ_ε de l'ordre de ε^α avec $0 < \alpha < \theta$

alors localement autour de $(x_\varepsilon, \xi_\varepsilon)$ il existe une solution de (11) avec $f_\varepsilon|_{t=0} = \delta_{x_\varepsilon}(x)\,\delta_{\xi_\varepsilon}(\xi)$. De plus, $(x_\varepsilon(t), \xi_\varepsilon(t)) \underset{\varepsilon}{\to} (x_+(t), \dot{x}_+(t))$ uniformément sur tout compact. On considère alors

$$f_{0,\varepsilon}^\hbar = W_\hbar\Big(\hbar^{-N/4}\, u\big(\frac{x - x_\varepsilon}{\hbar^{1/2}}\big)\, e^{i(\xi_\varepsilon/\hbar)\cdot x}\Big)\,,$$

avec $\|u\|_{L^2} = 1$ de sorte que

$$f_{0,\varepsilon}^\hbar \underset{\hbar \to 0_+}{\rightharpoonup} f_{0,\varepsilon} = \delta_{x_\varepsilon}(x)\,\delta_{\xi_\varepsilon}(\xi)$$

et donc

$$f_\varepsilon^\hbar(t) \underset{\hbar \to 0_+}{\rightharpoonup} \delta_{x_\varepsilon(t)}(x)\,\delta_{\xi_\varepsilon(t)}(\xi)\,.$$

Comme

$$\delta_{x_\varepsilon(t)}\,\delta_{\xi_\varepsilon(t)}(\xi) \underset{\varepsilon \to 0_+}{\rightharpoonup} \delta_{x_+(t)}(x)\delta_{\dot{x}_+(\xi)}\,,$$

on conclut aisément par une construction de suite diagonale.

Théorème IV.2. (Cas nonlinéaire: $V = V_0 * \rho$). *On suppose que V_0 est minoré, $V_0 \in C^1(\mathbb{R}^N)$, $\nabla V_0 \in C_b(\mathbb{R}^N)$, $\mathrm{Tr}\,(H_0 \rho_0^\hbar)$ et*

$$\iint_{\mathbb{R}^{2N}} V_0^+(x - y)\,\rho_0^\hbar(x)\,\rho_0^\hbar(y)\,dx\,dy$$

sont bornés indépendamment de $\hbar$ où $H_0 = -\frac{\hbar^2}{2}\Delta$. On suppose enfin que $\nabla V_0 \in C_0(\mathbb{R}^N)$ ou que $\int_{\mathbb{R}^N} |x|^2\,\rho_0^\hbar(x)\,dx$ est borné indépendamment de $\hbar$.

1) *Alors, $f^\hbar$, après extraction éventuelle d'une sous-suite, converge uniformément sur tout compact de $\mathbb{R}_+$ dans $\mathcal{A}'_{w-*}$ vers $f \in C_b(\mathbb{R}_t, \mathcal{M}w - *)$ vérifiant (70) et l'équation de Vlasov*

$$\begin{aligned} &\frac{\partial f}{\partial t} + \mathrm{div}_x(\xi f) - \mathrm{div}_\xi(\nabla V(x) f) = 0\,, \qquad \text{dans } \mathcal{D}'\,, \\ &V = V_0 * \rho \quad , \quad \rho = \int_{\mathbb{R}^N} f\,d\xi\,. \end{aligned} \tag{73}$$

2) *D'autre part, si $V_0 \in C^{1,1}(\mathbb{R}^N)$, f est l'unique solution de (70) et (73) dans $C_b(\mathbb{R}_t, \mathcal{M}w - *)$ et f est donnée par le transport de f_0 par $(x, \xi) \mapsto (x(t), \xi(t))$ où $(x(t), \xi(t))$ est la solution de*

$$\dot{x}(t) = \xi(t)\,, \quad \dot{\xi}(t) = -\nabla V(t, x(t))\,, \quad x(0) = x\,, \quad \xi(0) = \xi\,.$$

REMARQUE IV.4. Rappelons que

$$\mathrm{Tr}\,(H_0\rho_0^\hbar) = \sum_j \frac{\hbar^2}{2}\lambda_j^\hbar \int_{\mathbb{R}^N} |\nabla\psi_{0,j}^\hbar|^2\,dx\,,$$

où

$$\rho_0^\hbar = \sum_j \lambda_j^\hbar\,\psi_{0,j}(x)\psi_{0,j}(y)^*\,, \quad \lambda_j^\hbar \geq 0\,, \quad \int_{\mathbb{R}^N} \psi_{0,j}\,\psi_{0,k}^*\,dx = \delta_{jk}\,.$$

REMARQUE IV.5. En fait, la résolution de l'équation de Liouville non-linéaire à $\hbar > 0$ fixé n'est pas tout à fait évidente. Une manière simple de s'en convaincre consiste à réécrire le problème en diagonalisant $\rho_0 = \sum_{j\geq 1}^\infty \lambda_j\,\varphi_j^0(x)\,\varphi_j^{0*}(y)$ avec $\lambda_j \geq 0$, $\sum_{j\geq 1}^\infty \lambda_j < \infty$, $\int_{\mathbb{R}^N} \varphi_j^0\varphi_k^{0*}\,dx = \delta_{jk}$. Il faut alors résoudre le système suivant d'équations de Schrödinger pour tout $j \geq 1$

$$\begin{cases} i\hbar\,\frac{\partial\varphi_j}{\partial t} &= -\frac{\hbar^2}{2}\Delta\varphi_j + V\varphi_j\,, & \text{dans } \mathbb{R}_t \times \mathbb{R}_x^N\,, \\ \varphi_j|_{t=0} &= \varphi_j^0\,, & \text{dans } \mathbb{R}^N\,, \end{cases}$$

et $V = V_0 * \rho$, $\rho(t,x) = \sum_{j\geq 1}^\infty \lambda_j|\varphi_j(t,x)|^2$. Ce système, bien qu'infini, est faiblement couplé et s'analyse simplement comme une équation de Schrödinger nonlinéaire (voir par exemple J. Ginibre et G. Velo [21], [22], Th. Cazenave et A. Haraux [7]). Le cas où $\lambda_j = 0$ si j est grand se réduit à un système fini standard et on peut construire une solution dans le cas général par un simple passage à la limite. C'est ainsi que sous les hypothèses faites sur V_0 dans le Théorème IV.2, on obtient l'existence de $\varphi_j \in C(\mathbb{R}_t; L^2(\mathbb{R}_x^N))$. De plus, si

$$\mathrm{Tr}\,(H_0\rho_0) + \iint_{\mathbb{R}^N\times\mathbb{R}^N} V_0^+(x-y)\,\rho^0(x)\,\rho^0(y)\,dx\,dy \leq C_0\,,$$

$$\left(\mathrm{Tr}\,(H_0\rho_0) = \sum_{j\geq 1}^\infty \lambda_j \frac{\hbar}{2}^2 \int_{\mathbb{R}^N} |\nabla\varphi_j^0|^2\,dx\right),$$

on démontre que

$$\mathrm{Tr}\,((H_0+V)\rho)(t) \leq \mathrm{Tr}\,((H_0+V)\rho^0)\,, \qquad (\text{pour tout } t\in\mathbb{R})\,,$$

d'où

$$\sum_{j\geq 1}^{\infty} \lambda_j \frac{\hbar^2}{2} \int_{\mathbb{R}^N} |\nabla \varphi_j(t,x)|^2 \, dx + \iint_{\mathbb{R}^N \times \mathbb{R}^N} V_0^+(x-y)\, \rho(t,x)\, dx\, dy$$
$$\leq C_0 + \left(\sup_{\mathbb{R}^N} V^-\right) \sum_{j\geq 1}^{\infty} \lambda_j \, ,$$

pour tout $t \in \mathbb{R}$.

REMARQUE IV.6. Le système Hamiltonien dépendant du temps

$$\dot{x}(t) = \xi(t)\,, \qquad \dot{\xi}(t) = -[\nabla V_0 * \rho(t)](x(t))\,, \tag{74}$$
$$x(0) = x\,, \qquad \xi(0) = \xi\,,$$

peut être réécrit comme une "évolution newtonienne"

$$\dot{x}(t) = \xi(t)\,, \quad \dot{\xi}(t) = -\iint_{\mathbb{R}^{2N}} \nabla V_0(x(t) - x(t,y,\eta))\, df_0(y,\eta)\,, \tag{75}$$
$$x(0) = x\,, \qquad \xi(0) = \xi\,,$$

où on note $x(t,y,\eta)$ la solution correspondant à des conditions initiales (y,η).

REMARQUE IV.7. Le résultat précédent est encore valable si on suppose que V_0 est minoré, $V_0 \in C^1(\mathbb{R}^N)$, $\nabla V_0 \in C_b(\mathbb{R}^N)$ et que $\mathrm{Tr}\,(\rho_0^\hbar)$ converge vers $\iint_{\mathbb{R}^{2N}} df_0$ (ou en d'autres termes que $f_0^\hbar$ converge étroitement vers f_0). Il faut alors légèrement modifier la démonstration du Théorème IV.2 que nous donnons ci-dessous. On introduit $\varphi \in C_0^\infty(\mathbb{R}^N)$, $\varphi \equiv 1$ sur $B(0,1)$, $\varphi \equiv 0$ sur $B(0,2)^c$ et on observe que, grâce à (67), on a pour tout $n \geq 1$

$$\frac{d}{dt} \iint_{\mathbb{R}^{2N}} f^\hbar\, \varphi\Big(\frac{x}{n}\Big)\, \varphi\Big(\frac{\xi}{n}\Big)\, dx\, d\xi$$
$$= \iint_{\mathbb{R}^{2N}} f^\hbar\, \frac{\xi}{n} \cdot \nabla\varphi\Big(\frac{x}{n}\Big)\, \varphi\Big(\frac{\xi}{n}\Big)\, dx\, d\xi$$
$$+ \iint_{\mathbb{R}^{2N}} \rho^\hbar\Big(x + \frac{\hbar}{2}y\Big)\Big(\frac{V\big(x+\frac{\hbar}{2}y\big) - V\big(x-\frac{\hbar}{2}y\big)}{\hbar}\Big)$$
$$\cdot \varphi\Big(\frac{x}{n}\Big)\hat{\varphi}(ny)\, n^{-N}\, dx\, dy\,.$$

Or le deuxième terme peut se majorer par C/n avec $C \geq 0$ indépendant de $\hbar$. En considérant comme dans la preuve du Théorème IV.2 la limite f de $f^\hbar$, on obtient $f \in C_b(\mathbb{R}_+, \mathcal{M}_{w-*})$ qui vérifie (au sens des distributions)

$$\left| \frac{d}{dt} \iint_{\mathbb{R}^{2N}} \varphi(\frac{x}{n})\psi(\frac{\xi}{n})\, df(t) - \iint_{\mathbb{R}^{2N}} \frac{\xi}{n}\nabla\varphi(\frac{x}{n})\varphi(\frac{\xi}{n})\, df(t) \right| \leq \frac{C}{n}\,.$$

Et on voit que le deuxième terme tend vers 0 quand n tend vers $+\infty$ (car $\nabla\varphi \equiv 0$ pour $|x| \leq 1$). On déduit donc de cette inégalité que $\iint_{\mathbb{R}^{2N}} df(t) = \iint_{\mathbb{R}^{2N}} df_0$ pour tout $t \in \mathbb{R}$. D'autre part, on a déjà:

$$\mathrm{Tr}\,(\rho^\hbar(t)) = \mathrm{Tr}\,(\rho_0^\hbar) \underset{\hbar}{\to} \iint_{\mathbb{R}^{2N}} df_0$$

et on obtient donc que

$$\mathrm{Tr}\,(\rho^\hbar(t)) \underset{\hbar}{\to} \iint_{\mathbb{R}^{2N}} df(t)\,, \qquad \text{(pour tout } t \in \mathbb{R})\,.$$

D'après le Théorème III.2 cela entraîne que $\int_{\mathbb{R}^N} f^\hbar(t)\, d\xi$ converge étroitement vers $\rho = \int_{\mathbb{R}^N} df(t,\cdot,\xi)$. Et on peut alors aisément adapter la démonstration du Théorème IV.2.

DÉMONSTRATION DU THÉORÈME IV.1. Pour démontrer le point 1, il nous suffit, au vu des remarques faites au début de la Section III, de montrer que pour tout

$$\varphi \in \{\psi \in \mathcal{S} :\ \mathcal{F}_\xi\psi \in C_0^\infty(\mathbb{R}^N_x \times \mathbb{R}^N_z)\}\,,$$

$\langle K_\hbar \underset{\xi}{*} f^\hbar\,,\, \varphi\rangle$ est borné indépendamment de $t \in \mathbb{R}$ et converge quand $\hbar$ tend vers 0_+ vers $\iint_{\mathbb{R}^{2N}} \nabla V(x)\cdot\nabla_\xi\varphi(x,\xi)\, df(t)$. Or, nous avons

$$\begin{aligned}\langle K_\hbar \underset{\xi}{*} f^\hbar\,,\, \varphi\rangle &= \frac{i}{(2\pi)^N} \iint_{\mathbb{R}^{2N}} f^\hbar(x,\eta) \int_{\mathbb{R}^N} (\mathcal{F}_\xi\varphi)(x,y) \\ &\qquad \cdot e^{i\eta\cdot y}\frac{1}{\hbar}\left(V\big(x+\frac{\hbar}{2}y\big) - V\big(x-\frac{\hbar}{2}y\big)\right) dy\, dx\, d\eta \\ &= \frac{i}{(2\pi)^N}\,\langle f^\hbar, \psi^\hbar\rangle_{\mathcal{A}'\times\mathcal{A}}\,,\end{aligned}$$

où

$$\psi^\hbar(x,\eta) = \int_{\mathbb{R}^N} (\mathcal{F}_\xi\varphi)(x,y)\, e^{i\eta\cdot y}\frac{1}{\hbar}\left(V\left(x+\frac{\hbar}{2}y\right)-V\left(x-\frac{\hbar}{2}y\right)\right)dy \in \mathcal{A},$$

puisque

$$(\mathcal{F}_\eta\psi^\hbar)(x,z) = (2\pi)^N\frac{1}{\hbar}(F_\xi\varphi)(x,z)\left(V\left(x+\frac{\hbar}{2}z\right)-V\left(x-\frac{\hbar}{2}z\right)\right)$$

et $\mathcal{F}_\xi\varphi \in C_0^\infty$, $V \in C^1$. On voit en outre que $\psi^\hbar$ converge dans $\mathcal{A}$ vers $\int_{\mathbb{R}^N}(\mathcal{F}_\xi\varphi)(x,y)\, y\cdot\nabla V(x)\, e^{i\eta\cdot y}\, dy$. En effet, $\mathcal{F}_\eta\psi^\hbar$ converge vers $(2\pi)^N(\mathcal{F}_\xi\varphi)z\cdot\nabla V(x)$ dans $L^1(\mathbb{R}^N_z, C^0(\mathbb{R}^N_x))$. Et on conclut aisément puisque

$$\begin{aligned}\frac{i}{(2\pi)^N}\int_{\mathbb{R}^N}(\mathcal{F}_\xi\varphi)(x,y)y\cdot\nabla V(x)\, e^{i\eta\cdot y}\, dy &= \nabla V(x)\cdot\nabla_\eta\left[\frac{1}{(2\pi)^N}\int(\mathcal{F}_\xi\varphi)(x,y)\, e^{i\eta\cdot y}\, dy\right]\\ &= \nabla V(x)\cdot\nabla_\eta\varphi(x,\eta)\,.\end{aligned}$$

Le point 2 est en fait une simple remarque sur (11). En effet, les hypothèses faites sur V assurent que le flot Hamiltonien H_t

$$\dot{x} = \xi, \qquad \dot{\xi} = -\nabla V(x)$$

est bien défini sur $\mathbb{R}^{2N}_{x,\xi}$ pour tout $t \in \mathbb{R}$ et si $\psi_0 \in C_0^\infty(\mathbb{R}^{2N})$ alors $\psi(t,x,\xi) = \psi_0 \circ H_t(x,\xi)$ est à support compact en (x,ξ) pour tout $t \in \mathbb{R}$ uniforme pour t borné et est lipschitzienne sur $[-T,+T]\times\mathbb{R}^{2N}$ (pour tout $T \in]0,+\infty[$). De plus, on a (p.p.)

$$\frac{\partial\psi}{\partial t} = \xi\cdot\nabla_x\psi - \nabla V(x)\cdot\nabla_\xi\psi\,.$$

Cela permet de déduire pour toute solution f de (11), (70) dans $C_b(\mathbb{R}_t;\mathcal{M}w-*)$ l'égalité (72) qui démontre le point 2). Néanmoins, cette déduction nécessite une intégration par parties qui doit être justifiée. Une démonstration possible consiste à régulariser f: on pose alors $f_\delta = f * \rho_\delta$ où $\rho_\delta = \delta^{-2N}\,\rho(x/\delta)\rho(\xi/\delta)$ avec $\rho \in C_0^\infty(\mathbb{R}^N)$, $\rho \geq 0$,

$\int_{\mathbb{R}^N} \rho(x)\,dx = 1$ et $\mathrm{Supp}\,(\rho) \subset B(0,1)$ (par exemple). Il suffit alors de montrer

$$\frac{\partial f_\delta}{\partial t} + \xi \cdot \nabla_x f_\delta - \nabla V(x) \cdot \nabla_\xi f_\delta \;=\; r_\delta\,, \tag{76}$$

où r_δ est borné dans $C_b(\mathbb{R}_t; L^1_{\mathrm{loc}}(\mathbb{R}^{2N}))$, $r_\delta \underset{\delta}{\rightharpoonup} 0$ faiblement (par exemple au sens des mesures sur $[-T,+T[\times\mathbb{R}^{2N}$ pour tout $T \in]0,+\infty[$). Or, on a

$$\begin{aligned} r_\delta = &\iint \Big(\frac{\xi-\eta}{\delta}\Big)\cdot\frac{1}{\delta^N}\nabla\rho\Big(\frac{x-y}{\delta}\Big)\,\frac{1}{\delta^N}\rho\Big(\frac{\xi-\eta}{\delta}\Big)\,df_\delta(y,\eta) \\ &- \iint \Big(\frac{\nabla V(x)-\nabla V(y)}{\delta}\Big)\frac{1}{\delta^N}\nabla\rho\Big(\frac{\xi-\eta}{\delta}\Big)\,\frac{1}{\delta^N}\rho\Big(\frac{x-y}{\delta}\Big)\,df_\delta(y,\eta)\,, \end{aligned}$$

formule qui permet de vérifier les assertions énoncées ci-dessus pour r_δ.

DÉMONSTRATION DU THÉORÈME IV.2. La démonstration du point 2) du Théorème IV.2 est très semblable à celle du point 2) du Théorème IV.1. Aussi nous contenterons-nous de démontrer le point 1. Grâce à la démonstration du point 1) du Théorème IV.1, on peut supposer, après extractwion éventuelle d'une sous-suite, que $f^\hbar$ converge uniformément sur tout borné de $\mathbb{R}_+$ dans $\mathcal{M}w-*$ vers $f \in C_b(\mathbb{R}_t; \mathcal{M}w-*)$. De plus, on déduit de la Remarque IV.5 les bornes suivantes

$$\begin{aligned} \sup_{t\in\mathbb{R}}\Big(&\iint_{\mathbb{R}^{2N}} f^\hbar |\xi|^2\,dx\,d\xi \\ &+ \iint_{\mathbb{R}^{2N}} V_0^+(x-y)\rho^\hbar(x)\rho^\hbar(y)\,dx\,dy\Big) \;\le\; C \end{aligned} \tag{77}$$

et donc

$$\sup_{t\in\mathbb{R}} \iint_{\mathbb{R}^{2N}} \tilde{f}^\hbar\,|\xi|^2\,dx\,d\xi \;\le\; C\,, \tag{78}$$

où C désigne diverses constantes indépendantes de $\hbar$ (convention adoptée dans tout ce qui suit). Cela permet de démontrer grâce au Théorème III.1 (et à sa démonstration) que $\rho^\hbar$ converge uniformément sur les bornées de $\mathbb{R}_t$ dans $\mathcal{M}w-*$ vers $\rho \in C_b(\mathbb{R}_t; \mathcal{M}w-*)$ et $d\rho = \int_{\mathbb{R}^N} df(\cdot,\xi)$. Bien sûr, (77) implique

$$\begin{aligned} \sup_{t\in\mathbb{R}}\Big(&\iint_{\mathbb{R}^{2N}} |u|^2\,df(t) \\ &+ \iint_{\mathbb{R}^{2N}} V_0^+(x-y)\,d\rho^\hbar(t,x)\,d\rho^\hbar(t,y)\Big) < \infty\,. \end{aligned} \tag{79}$$

L'étape suivante consiste à étudier la limite de $V^\hbar = V_0 * \rho^\hbar$: bien sûr $\nabla V^\hbar$ est borné dans $C_b(\mathbb{R}^N)^N$. De plus, si $\nabla V_0 \in C_0(\mathbb{R}^N)^N$, on voit aisément que $\nabla V_0 * \rho^\hbar$ converge, uniformément sur les bornées de $\mathbb{R}_t$, uniformément sur tout compact de $\mathbb{R}^N$ vers $\nabla V_0 * f$.

Dans le cas où $\int_{\mathbb{R}^N} |x|^2 \rho_0^\hbar \, dx$ est borné indépendamment de $\hbar$, on obtient grâce à (38) et au fait que ∇V_0 est borné:

$$\left| \frac{d^2}{dt^2} \int_{\mathbb{R}^N} |x|^2 \rho^\hbar \, dx \right| \leq C\left(1 + \int_{\mathbb{R}^N} |x|^2 \rho^\hbar \, dx\right). \tag{80}$$

De plus,

$$\frac{d}{dt}\left(\int_{\mathbb{R}^N} |x|^2 \rho^\hbar \, dx\right)|_{t=0} = \sum_{j\geq 1} \lambda_j \, \mathrm{Im} \int_{\mathbb{R}^N} x(\hbar \nabla \varphi_j^0)\overline{\varphi_j^0} \, dx \,,$$

d'où, d'après la borne sur $\mathrm{Tr}\,(H_0\rho_0)$, une borne indépendante de $\hbar$ pour

$$\frac{d}{dt} \int_{\mathbb{R}^N} |x|^2 \rho^\hbar \, dx|_{t=0} \,.$$

Cette borne associée à (80) implique la borne suivante

$$\int_{\mathbb{R}^N} |x|^2 \rho^\hbar \, dx \leq C(1+t^2), \qquad \text{pour tout } \; t \in \mathbb{R} \,. \tag{81}$$

On déduit de cette estimation uniforme, la convergence étroite dans $\mathbb{R}^N$, uniforme pour t borné, de $\rho^\hbar$ vers ρ qui assure dans ce cas également la convergence uniforme pour t, x bornés de $\nabla V_0 * \rho^\hbar$ vers $\nabla V_0 * \rho$.

On peut alors facilement adapter la démonstration correspondante du Théorème IV.1 pour conclure la preuve du point 1) du Théorème IV.2.

Le deuxième type de résultats que nous présenterons concerne des situations très régulières où f_0, V sont réguliers et où la convergence de $f_0^\hbar$ vers f_0 est "régulière" au sens où on peut donner un développement asymptotique en puissances de $\hbar$ de $f_0^\hbar$. Pour simplifier (et alléger) la présentation, nous supposerons qu'il existe $N_0 \geq 1$ tel que

$$V \in C_b^\infty(\mathbb{R}^N), \qquad i.e. \quad D^\alpha V \in C_b(\mathbb{R}^N), \quad \text{pour tout } \; \alpha \,, \tag{82}$$

$$(83)\qquad \begin{cases} f_0^\hbar \;=\; f_0 + \hbar^{1/2} f_0^1 + \hbar f_0^2 + \cdots + \hbar^{N_0/2} f_0^{N_0} + \hbar g_0^\hbar \,, \\ \|g_0^\hbar\|_{L^2} \;\leq\; C\,\hbar^{(N_0-1)/2}; \qquad f_0, f_0^1, \ldots, f_0^{N_0} \in \mathcal{S}(\mathbb{R}^{2N};\mathbb{R}) \,. \end{cases}$$

Nous donnerons des exemples plus bas montrant que le développement en puissances de $\hbar^{1/2}$ est naturel pour $f_0^\hbar$. Il sera également clair que l'on peut économiser de la régularité pour V et les f_0^i suivant l'ordre d'approximation (en puissance de $\hbar$) souhaité. Nous nous contenterons aussi d'analyser le cas linéaire *i.e.* l'équation de Liouville (10). Il est par contre important de noter que (83) implique bien sûr que $f_0^\hbar$ est bornée dans $L^2(\mathbb{R}^{2N})$ et que cette seule hypothèse exclut essentiellement le cas de l'équation de Schrödinger (6). En effet, si $f_0^\hbar = W_\hbar(\psi_0^\hbar)$, alors $\|f_0^\hbar\|_{L^2} = (2\pi\hbar)^{-N/2}\,\|\psi_0^\hbar\|_{L^2}^2$. Donc, la borne dans L^2 de $f_0^\hbar$ implique que $\psi_0^\hbar$ converge fortement dans L^2 vers 0, cas qui n'est pas intéressant.

Théorème IV.3. *Sous les hypothèses* (82) *et* (83), *il existe* $f, f^1, f^2, \ldots, f^{N_0}$ *fonctions réelles régulières en* (t,x) *et à décroissance rapide* (*ainsi que toutes leurs dérivées*) *en* x *uniformément en* t *borné tels que*

$$\begin{aligned} f \;&=\; f + \hbar^{1/2} f^1 + \cdots + \hbar^{\frac{N_0}{2}} f^{N_0} + g^\hbar \,, \\ \|g^\hbar\|_{C([-T,+T];L^2(\mathbb{R}^{2N}))} \;&\leq\; C_T\,\hbar^{(N_0+1/2)} \,, \end{aligned}$$

où T *est arbitraire dans* $]0,+\infty[$ *et* C_T *est une constante positive indépendante de* $\hbar$.

REMARQUE IV.8. La démonstration donne bien sûr des équations d'évolutions pour les f^i $(i \geq 1)$ qui sont toutes du type

$$\frac{\partial f^i}{\partial t} + \xi \cdot \nabla_x f^i - \nabla V(x) \cdot \nabla_\xi f^i \;=\; \mathcal{A}_i(f^{i-4}, f^{i-8}, \ldots)\,, \quad f^i|_{t=0} = f_0^i \,,$$

(en convenant que $f = f^0$) où $\mathcal{A}_i$ est un opérateur linéaire différentiel. En particulier, pour $i \leq 3$, on a comme pour f

$$\frac{\partial f^i}{\partial t} + \xi \cdot \nabla_x f^i - \nabla V(x) \cdot \nabla_\xi f^i \;=\; 0\,, \qquad f^i|_{t=0} = f_0^i \,.$$

REMARQUE IV.9. Donnons tout d'abord un exemple permettant d'illustrer la condition (83). On reprend la construction de la fin de la Section III : soit $f_0 \in \mathcal{S}(\mathbb{R}^{2N})$, $u_0 \in \mathcal{S}(\mathbb{R}^N)$, on introduit

$$\rho_0^\hbar \;=\; \iint_{\mathbb{R}^{2N}} \hbar^{-N/2} u_0\left(\frac{x-x_0}{\hbar^{1/2}}\right) u_0^*\left(\frac{y-x_0}{\hbar^{1/2}}\right)$$

$$\cdot\, e^{i(\xi_0/\hbar)\cdot(x-y)}\, f_0(x_0,\xi_0)\, dx_0\, d\xi_0\,.$$

Alors, on a

$$f_0^\hbar = f_0 * W_0^\hbar\,, \qquad W_0^\hbar(x,\xi) = \hbar^{-N}\, W_0\Big(\frac{x}{\hbar^{1/2}}, \frac{\xi}{\hbar^{1/2}}\Big)\,,$$
$$W_0 = W(u_0) \in \mathcal{S}(\mathbb{R}^{2N})\,.$$

On voit bien qu'un tel choix de $\rho_0^\hbar$ conduit à un développement en puissances de $\hbar^{1/2}$ comme cela est supposé dans (83) et que ce développement n'est en puissances entières de $\hbar$ que si u_0 est paire (ou plus généralement si les moments d'ordre impair s'annulent jusqu'à un ordre convenable).

En fait, on peut également traiter des cas plus complexes où le développement se fait suivant $\hbar^{k\alpha+l(1-\alpha)}$ ($h, l \geq 0$ entiers) -la démonstration qui suit restant essentiellement inchangée . Un tel développement apparaît naturellement si on considère

$$\rho_0^\hbar = \iint_{\mathbb{R}^{2N}} \hbar^{-N\alpha}\, u_0\Big(\frac{x-x_0}{\hbar^\alpha}\Big)\, u_0^*\Big(\frac{y-x_0}{\hbar^\alpha}\Big)$$
$$\cdot\, e^{i(\rho_0/\hbar)(x-y)}\, f_0(x_0,\xi_0)\, dx_0\, d\rho_0\,.$$

On trouve alors que

$$f_0^\hbar = f_0 * W_0^\hbar\,, \qquad W_0^\hbar(x,\xi) = \hbar^{-N}\, W_0\Big(\frac{x}{\hbar^\alpha}, \frac{\xi}{\hbar^{1-\alpha}}\Big)\,,$$
$$W_0 = W(u_0) \in \mathcal{S}(\mathbb{R}^{2N})\,.$$

DÉMONSTRATION DU THÉORÈME IV.3. En remarquant que l'on peut développer à tout ordre

$$\frac{1}{\hbar}\Big(V_0\big(x+\frac{\hbar y}{2}\big) - V_0\big(x-\frac{\hbar y}{2}\big)\Big)$$
$$= y\cdot\nabla_V(x) + \sum_{n\geq 1}\hbar^{2n} \sum_{|\alpha|=2n+1} (\alpha!)^{-1} y^\alpha \cdot D^\alpha V(x)\,,$$

on obtient aisément grâce à (67) un développement formel de $f^\hbar$ qui conduit à l'identification de $f^1, f^2, \ldots$ (voir Remarque IV.8). De plus,

si on pose $\overline{f}^\hbar = f + \hbar^{1/2} f^1 + \cdots + \hbar^{N_0/2} f^{N_0}$, on vérifie par des calculs fastidieux que $\overline{f}^\hbar$ vérifie

$$\frac{\partial \overline{f}^\hbar}{\partial t} + \xi \cdot \nabla_x \overline{f}^\hbar + K_\hbar * \overline{f}^\hbar = r^\hbar \in C(\mathbb{R}_t; L^2(\mathbb{R}^{2N})), \tag{85}$$

$$\overline{f}^\hbar|_{t=0} = f_0^\hbar - \hbar g_0^\hbar ,$$

et pour tout $T \in]0, +\infty[$

$$\sup_{t\in[-T,+T]} \|r^\hbar(t)\|_{L^2(\mathbb{R}^{2N})} \leq C_T \hbar^{(N_0+1)/2} . \tag{86}$$

D'où en posant $g^\hbar = f^\hbar - \overline{f}^\hbar$,

$$\frac{\partial g^\hbar}{\partial t} + \xi \cdot \nabla_x g^\hbar + K_\hbar * g^\hbar = r^\hbar , \qquad g^\hbar|_{t=0} = g_0^\hbar .$$

D'après la Remarque II.3, on déduit alors

$$\frac{d}{dt} \iint_{\mathbb{R}^{2N}} |g^\hbar|^2 \, dx \, d\xi \leq \iint_{\mathbb{R}^{2N}} r^\hbar g^\hbar \, dx \, d\xi$$

et on conclut aisément au vu de (86) et de (83).

Nous allons maintenant passer au troisième type de résultats énoncé dans l'Introduction où nous montrerons comment la régularité C^1 supposée pour V et V_0 peut être affaiblie de façon à atteindre des potentiels physiquement plus réalistes. Le prix à payer sera de faire des hypothèses supplémentaires sur $f_0^\hbar$. Nous traiterons tout d'abord le cas linéaire puis le cas nonlinéaire. Dans les deux cas, nous n'essaierons pas ici d'obtenir les résultats les plus généraux et nous nous contenterons d'illustrer une méthode par quelques exemples représentatifs. Une étude plus systématique reposant notamment sur des idées de "renormalisation de $\rho^\hbar$" permettant d'affaiblir encore plus les hypothèses de régularité sur V sera menée dans P. Gérard, P. L. Lions et T. Paul [20].

Nous supposerons que $f_0^\hbar$ et V vérifient

$$f_0^\hbar \quad \text{est bornée dans} \quad L^2(\mathbb{R}^{2N}) , \tag{87}$$

$$V \in H^1_{\text{loc}}(\mathbb{R}^N) . \tag{88}$$

Pour les raisons données plus haut, (87) n'a vraiment de sens que si l'on travaille sur l'équation de Liouville (10). Enfin, pour assurer que (6) ou (10) admettent des solutions, nous supposerons également

$$V^- \in K^N(\mathbb{R}^N) \tag{89}$$

de sorte que (34) est vérifiée. L'hypothèse (35) servant à assurer que (67) est faiblement interprétable au sens des distributions sera également faite mais, comme nous l'avons indiqué précédemment, elle peut être supprimée en interprétant (67) convenablement.

Théorème IV.4. *Sous les hypothèses* (35) *et* (87)-(89), $f^\hbar$, *après extraction éventuelle d'une sous-suite, converge faiblement dans* $L^\infty([-T,+T];L^2(\mathbb{R}^{2N}))_{w-*}$ (*et dans* $C([-T,+T];\mathcal{A}'_{w-*})$) *pour tout* $T \in]0,+\infty[$ *vers* $f \in C_b(\mathbb{R}_t;\mathcal{M}_{w-*}) \cap L^\infty(\mathbb{R}_t;L^1 \cap L^2(\mathbb{R}^{2N}))$ *qui vérifie* (11) *au sens des distributions et* (70).

DÉMONSTRATION. D'après la Remarque II.3, on voit que $f^\hbar$ est bornée dans $L^\infty(\mathbb{R}_t;L^2(\mathbb{R}^{2N}))$. En utilisant cette borne et en adaptant la démonstration du Théorème IV.1, on peut supposer que $f^\hbar$, quitte à extraire une sous-suite, converge au sens précisé ci-dessus vers $f \in C_b(\mathbb{R}_t;\mathcal{M}_{w-*}) \cap L^\infty(\mathbb{R}_t;L^1 \cap L^2(\mathbb{R}^{2N}))$ qui vérifie bien sûr (70). Le seul point réellement nouveau est la vérification de (11) pour laquelle nous utiliserons (88).

En effet, avec les notations de la démonstration du Théorème IV.1, il nous suffit de démontrer que si $\psi = (\mathcal{F}_\xi\varphi)(x,z) \in C_0^\infty(\mathbb{R}^{2N})$

$$\psi(x,z)\frac{1}{\hbar}\left(V\left(x+\frac{\hbar}{2}z\right)-V\left(x-\frac{\hbar}{2}z\right)\right) \underset{\hbar\to 0}{\to} \psi(x,z)z\cdot\nabla V(x)$$

dans $L^2(\mathbb{R}^{2N})$. Et cette convergence est immédiate au vu de (88).

Dans le cas nonlinéaire (où $V = V_0 * \rho$), nous conservons bien sûr l'hypothèse (87) et nous ferons sur V_0 les hypothèses suivantes

$$\begin{aligned} &V_0^- \in L^{(N+4)/4,\infty}(\mathbb{R}^N) + L^\infty(\mathbb{R}^N), \qquad \text{si} \quad N \le 3, \\ &V_0^- \in L^r(\mathbb{R}^N) + L^\infty(\mathbb{R}^N), \qquad \text{avec} \quad r > \frac{N}{2} \quad \text{si} \quad N \ge 4, \end{aligned} \tag{90}$$

$$\begin{aligned} &\nabla V_0 \in L^{(2N+8)/(N+8)}(\mathbb{R}^N) + L^q(\mathbb{R}^N), \\ &\text{avec} \quad \frac{2N+8}{N+8} < q < \infty. \end{aligned} \tag{91}$$

Bien sûr, nous pouvons également considérer des situations où $V = V_0 * \rho + V_1$ auquel cas il faudrait supposer que V_1 vérifie (88) et (89).

Le cas de l'équation de *Wigner-Poisson* qui intervient dans les semi-conducteurs (voir les références données dans l'Introduction) correspond à $N = 3$, $V_0 = 1/|x|$ et (90) et (91) ont lieu puisque

$$V_0^- \equiv 0\,, \quad \nabla V_0 \in L^{3/2,\infty}(\mathbb{R}^3)\,, \quad \text{et} \quad \frac{2N+8}{N+8} = \frac{14}{11} < \frac{3}{2}\,,$$

si $N = 3$.

En fait, sous ces seules hypothèses, la résolution de l'équation de Liouville nonlinéaire n'est pas évidente et découle d'une part de la Remarque IV.5 et des estimations a priori que nous démontrons ci-dessous. Nous ne voulons pas développer ce point assez technique ici.

Théorème IV.5. *Sous les hypothèses* (87), (90), (91) *et si*

$$\mathrm{Tr}\,(H_0\rho_o^\hbar)\,, \qquad \iint_{\mathbb{R}^{2N}} V_0^+(x-y)\rho_0^\hbar(x)\rho_0^\hbar(y)\,dx\,dy$$

sont bornés indépendamment de $\hbar$, alors

$$\|f^\hbar\|_{L^2(\mathbb{R}^{2N})}\,, \qquad \mathrm{Tr}\,(H_0\rho^\hbar(t))\,,$$

$$\iint_{\mathbb{R}^{2N}} V_0^\pm(x-y)\,\rho^\hbar(t,x)\rho^\hbar(t,y)\,dx\,dy \quad \textit{et} \quad \|\rho^\hbar(t)\|_{L^{(N+4)/(N+2)}(\mathbb{R}^N)}$$

sont bornés indépendamment de $\hbar$ et de $t \in \mathbb{R}$. De plus, $f^\hbar$, après extraction éventuelle d'une sous-suite, converge faiblement dans $L^\infty([-T,+T];L^2(\mathbb{R}^{2N})w-*$ (et dans $C([-T,+T];\mathcal{M}_{w-*})$) pour tout $T \in]0,+\infty]$ vers $f \in C_b(\mathbb{R}_t;\mathcal{M}_{w-*}) \cap L^\infty(\mathbb{R}_t, L^1 \cap L^2(\mathbb{R}^N))$ qui vérifie l'équation de Vlasov (73) *et* (70).

REMARQUE IV.10. Exactement comme dans le Théorème IV.2, la solution trouvée de (73) vérifie

$$\sup_{t\in\mathbb{R}}\Big(\iint_{\mathbb{R}^{2N}} f|\xi|^2\,dx\,d\xi + \iint_{\mathbb{R}^{2N}} V_0^\pm(x-y)\rho(x)\rho(y)\,dx\,dy\Big) < +\infty\,. \tag{92}$$

DÉMONSTRATION DE LA REMARQUE. En admettant provisoirement les bornes énoncées dans le Théorème IV.5 il est facile de conclure en adaptant les démonstrations des Théorèmes IV.2 et IV.4. En effet, la borne d'énergie cinétique $\mathrm{Tr}\,(H_0\rho^\hbar)$ assure que $\rho^\hbar$ est relativement compacte dans $C([-T,+T];\mathcal{M}_{w-*})$ (pour tout $T\in]0,+\infty[$). De plus $\rho^\hbar(t)$ étant borné uniformément en $\hbar$ et t dans $L^1\cap L^{(N+4)/(N+2)}(\mathbb{R}^N)$, on déduit de (91) que $\nabla V(t,x)=\nabla V_0*\rho^\hbar(t)$ est borné dans $C_b(\mathbb{R}_t;L^2(\mathbb{R}^N))$ et converge dans $C([-T,+T];L^2_{\mathrm{loc}}(\mathbb{R}^N))$ (pour tout $T\in]0,+\infty[$) vers $\nabla V_0*\rho(t)$. Il est alors facile de conclure.

DÉMONSTRATION DU THÉORÈME. La borne L^2 sur $f^\hbar$ provient de la Remarque II.3. Les autres bornes sont des conséquences des résultats énoncés dans l'appendice: en effet, il y est démontré qu'il existe une constante $C_0\geq 0$ telle que pour toute matrice densité ρ

$$\|\rho\|_{L^{(N+4)/(N+2)}(\mathbb{R}^N)}\ \leq\ C_0\left(\mathrm{Tr}\,(\rho^2)\right)^{\theta/2}\left(\mathrm{Tr}\,(\overline{H}\rho)\right)^{1-\theta}$$

où $\theta=4/(N+4)$, $\overline{H}=-\Delta$. Or $\mathrm{Tr}\,(\overline{H}\rho^\hbar)=(2/\hbar^2)\,\mathrm{Tr}\,(H_0\rho^\hbar)$ et

$$\begin{aligned}\mathrm{Tr}\,(\rho^\hbar)^{1/2}\ &=\ \left(\iint_{\mathbb{R}^{2N}}|\rho^\hbar(x,y)|^2\,dx\,dy\right)^{1/2}\\ &=\ (2\pi\hbar)^{N/2}\,\|f^\hbar\|_{L^2(\mathbb{R}^{2N})}\ \leq\ C\,\hbar^{N/2}\,.\end{aligned}$$

Or $N\theta/2=2N/(N+4)=2(1-\theta)$ d'où

$$\|\rho^\hbar\|_{L^{(N+4)/(N+2)}(\mathbb{R}^N)}\ \leq\ C\left(\mathrm{Tr}\,(H_0\rho^\hbar)\right)^{N/(N+4)}\,. \tag{93}$$

De plus, les hypothèses faites sur V_0^-, à savoir (90), permettent alors de majorer

$$\iint_{\mathbb{R}^{2N}}V_0^-(x-y)\,\rho^\hbar(x)\,\rho^\hbar(y)\,dx\,dy\ \leq\ C\left(1+(\mathrm{Tr}\,(H_0\rho^\hbar))^\alpha\right), \tag{94}$$

pour une puissance $\alpha\in]0,1[$. Cela suffit à assurer que

$$\iint_{\mathbb{R}^{2N}}V_0^-(x-y)\,\rho_0^\hbar(x)\,\rho_0^\hbar(y)\,dx\,dy$$

est borné indépendamment de $\hbar$. En particulier, l'énergie totale

$$\mathrm{Tr}\,(H_0\rho_0^\hbar)+\iint_{\mathbb{R}^{2N}}V_0(x-y)\,\varphi_0^\hbar\,\varphi_0^\hbar(y)\,dx\,dy$$

est bornée. On déduit alors de la conservation de l'énergie totale et de (93)-(94) les bornes énoncées dans le Théorème IV.5.

REMARQUE IV.11. Nous avons donné dans l'Introduction diverses références concernant Wigner-Poisson. Signalons en outre l'étude récente de P. Markowich et N. Mauser [32] qui établit la limite semi-classique vers Vlasov-Poisson en partant d'un modèle de type Wigner-Poisson où le potentiel coulombien V_0 est un peu régularisé.

Signalons pour conclure que nous donnerons dans [20] de multiples extensions des résultats énoncés ci-dessus (Théorèmes IV.4-IV.5) portant sur l'affaiblissement des hypothèses sur V et sur V_0, mais aussi sur le type de solutions obtenues (renormalisées, régulières) et enfin nous étudierons le cas d'équations de Schrödinger avec potentiel-vecteur (correspondant par exemple à un champ magnétique).

APPENDICE. Amélioration des bornes semi-classiques pour des systèmes orthonormés.

Soit ρ un noyau définissant un opérateur positif ou nul, compact, hermitien sur $L^2(\mathbb{R}^N)$. On peut bien sûr diagonaliser l'opérateur et ainsi trouver une base orthonormée $\{\psi_j\}_{j\geq 1}$ dans $L^2(\mathbb{R}^N)$, des réels positifs ou nuls $\{\lambda_j\}_{j\geq 1}$ tels que

$$\rho(x,y) = \sum_{j\geq 1} \lambda_j \psi_j(x) \psi_j^*(y) , \tag{A.1}$$

$$\lambda_1 \geq \lambda_2 \geq \dots , \quad \lambda_j \to 0 \quad \text{si} \quad j \to +\infty . \tag{A.2}$$

Bien sûr, si $p \geq 1$, $\mathrm{Tr}(\rho^p) = \sum_{j\geq 1} \lambda_j^p$ et nous noterons

$$\|\rho\|_p = \Big(\sum_{j\geq 1} \lambda_j^p\Big)^{1/p} \quad \text{et} \quad \|\rho\|_\infty = \max_{j\geq 1} \lambda_j .$$

Nous noterons également

$$\rho(x) = \sum_{j\geq 1} \lambda_j |\psi_j(x)|^2 , \qquad \text{p.p.} \quad x \in \mathbb{R}^N \tag{A.3}$$

et

$$(A.4) \quad j(x) = \sum_{j\geq 1} \lambda_j \, \mathrm{Im}(\nabla\psi_j(x)\psi_j^*(x)), \qquad \text{p.p.} \quad x \in \mathbb{R}^N .$$

Cette dernière quantité n'est définie que sous certaines hypothèses sur ψ_j et λ_j, de même que l'intégrabilité de ρ (ou de puissances de ρ) n'est pas assurée automatiquement. En fait, les bornes que nous allons établir permettront aussi de définir ces quantités dans des cadres généraux.

Nous introduisons enfin "l'énergie cinétique" de ρ

$$(A.5) \qquad \mathrm{Tr}(\overline{H}\rho) = \sum_{j\geq 1} \lambda_j \int_{\mathbb{R}^N} |\nabla\psi_j(x)|^2 \, dx .$$

Notre résultat principal est une variante d'un résultat classique de E. H. Lieb et W. Thirring [26] (traitant le cas de ρ pour $\lambda_j = 1$ pour $1 \leq j \leq M$, $\lambda_j = 0$ pour $j > M$).

Théorème. *Soit $p \in [1, +\infty]$. On suppose que $\mathrm{Tr}(\overline{H}\rho)$ et $\|\rho\|_p$ sont finis. Alors, les séries définissant ρ et j sont convergentes dans respectivement $L^q(\mathbb{R}^N)$, $L^r(\mathbb{R}^N)$ avec*

$$q = \frac{(N+2)p - N}{Np - (N-2)}, \qquad r = \frac{(N+2)p - N}{(N+1)p - (N-1)} .$$

De plus, il existe des constantes positives C_0, C_1 indépendantes de ρ telles que

$$(A.6) \qquad \|\rho\|_{L^q(\mathbb{R}^N)} \leq C_0 \, \|\rho\|_p^\theta \, (\mathrm{Tr}(\overline{H}\rho))^{1-\theta},$$

avec $\theta = \dfrac{2p}{(N+2)p - N}$,

$$(A.7) \qquad \|j\|_{L^r(\mathbb{R}^N)} \leq C_1 \, \|\rho\|_p^\theta \, (\mathrm{Tr}(\overline{H}\rho))^{1-\theta},$$

avec $\theta = \dfrac{p}{(N+2)p - N}$.

REMARQUE. Le cas $p = 1$ est trivial. Aussi, nous supposerons $p > 1$ dans tout ce qui suit.

REMARQUE. Par transformée de Wigner (et de Husimi), on peut déduire de ces inégalités grâce aux inégalités prouvées en fin de Section III les inégalités suivantes d'interpolation (classiques dans l'analyse des équations de Vlasov). Si $f \geq 0 \in L^1(|\xi|^2\,dx\,d\xi) \cap L^p(\mathbb{R}^{2N})$ alors $\rho \in L^q(\mathbb{R}^N)$, $j = \int_{\mathbb{R}^N} \xi\, f(x,\xi)\,d\xi \in L^r(\mathbb{R}^N)$ et

$$\text{(A.6')} \qquad \|\rho\|_{L^q(\mathbb{R}^N)} \leq C_0' \|f\|^{\theta}_{L^p(\mathbb{R}^N)} \left(\iint_{\mathbb{R}^{2N}} f|\xi|^2\,dx\,d\xi \right)^{1-\theta},$$

avec $\theta = \dfrac{2p}{(N+2)p - N}$,

$$\text{(A.7')} \qquad \|j\|_{L^r(\mathbb{R}^N)} \leq C_1' \|f\|^{\theta}_{L^p(\mathbb{R}^N)} \left(\iint_{\mathbb{R}^{2N}} f|\xi|^2\,dx\,d\xi \right)^{1-\theta},$$

avec $\theta = \dfrac{p}{(N+2)p - N}$.

DÉMONSTRATION. Nous commençons par le

Lemme. *Soit $V \in L^{N/2+\delta}(\mathbb{R}^N)$ avec $\delta > 0$ si $N \geq 2$, $\delta > 1/2$ si $N = 1$. Alors, si on note $\mu_1 \leq \mu_2 \leq \ldots$ les valeurs propres négatives de l'opérateur $(\overline{H} + V)$, on a*

$$\text{(A.8)} \qquad \sum_j |\mu_j|^{\delta} \leq C(N,\delta) \int_{\mathbb{R}^N} |V^-|^{N/2+\delta}\,dx\,.$$

Ce résultat est dû à E. H. Lieb et W. Thirring [26] dans le cas où $N \geq 3$. Il suffit pour l'établir si $N = 1$ ou si $N = 2$ de reprendre la démonstration (présentée dans B. Simon [38] ou R. Dautray [9]) qui donne pour tout $\mu > 0$

$$\#\{j\,:\,\mu_j < -\mu\} \leq C \int_0^{\infty} e^{-\mu t/2}\, t^{-N/2-1}\,dt \int_{\mathbb{R}^N} f(tV^-(x))\,dx\,,$$

où $f(s) = (s-1)^+$, d'où

$$\begin{aligned} \sum_{j\geq 1} |\mu_j|^{\delta} &= \delta \int_0^{\infty} s^{\delta-1} \big(\#\{j\,:\,\mu_j < -s\}\big)\,ds \\ &\leq C \int_0^{\infty} t^{-(1+\delta+N/2)}\,dt \int_{\mathbb{R}^N} f(tV^-(x))\,dx \end{aligned}$$

$$= C \int_{\mathbb{R}^N} |V^-|^{N/2+\delta}\, dx\,,$$

où C désigne diverses constantes positives ne dépendant que de N et δ (la condition sur δ assure que $\int_0^\infty \sigma^{-(1+\delta+N/2)}\,(\sigma-1)^+\,d\sigma < \infty$).

A l'aide du lemme, nous allons tout d'abord démontrer (A.6) en s'inspirant de l'argument de E. H. Lieb et W. Thirring [26]. On considère l'opérateur $\overline{H} - t\rho^\alpha$ où $t > 0$ est une constante qui sera déterminée dans la suite et où $\alpha = (2(p-1))/(2+N(p-1))$. Nous allons établir que

$$\begin{aligned} \mathrm{Tr}\,((\overline{H} - t\rho^\alpha)\rho) &= \sum_{j\geq 1} \lambda_j \Big(\int_{\mathbb{R}^N} |\nabla\psi_j|^2 - t\rho^\alpha |\psi_j|^2\, dx \Big) \\ &\geq -\sum_{j\geq 1} \lambda_j |\mu_j|\,, \end{aligned} \tag{A.9}$$

où μ_j sont les valeurs propres négatives par ordre croissant de l'opérateur $\overline{H} - t\rho^\alpha$. En admettant provisoirement (A.9), on voit que

$$t \int_{\mathbb{R}^N} \rho^{\alpha+1}\, dx \;\leq\; \mathrm{Tr}\,(\overline{H}\rho) + \|\rho\|_p \Big(\sum_{j\geq 1} |\mu_j^-|^{p'} \Big)^{1/p'}\,,$$

d'où d'après le lemme précédent

$$t \int_{\mathbb{R}^N} \rho^{\alpha+1}\, dx \;\leq\; \mathrm{Tr}\,(\overline{H}\rho) + \|\rho\|_p\, C \Big(\int_{\mathbb{R}^N} (t\rho^\alpha)^{N/2+p'} \Big)^{1/p'}\,.$$

On observe ensuite que

$$\begin{aligned} \alpha + 1 &= \frac{2\,(p-1)}{2+N(p-1)} + 1 \\ &= \frac{(N+2)\,p - N}{Np - (N-2)} (= q) \;=\; \frac{N\alpha}{2} + \alpha p'\,, \end{aligned}$$

d'où

$$t \Big(\int_{\mathbb{R}^N} \rho^q\, dx \Big) \;\leq\; \mathrm{Tr}\,(\overline{H}\rho) + C \|\rho\|_p\, t^{1+N/(2p')} \Big(\int_{\mathbb{R}^N} \rho^q\, dx \Big)^{1/p'}\,.$$

On choisit alors

$$t^{N/(2p')} = (2C\|\rho\|_p)^{-1}\left(\int_{\mathbb{R}^N} \rho^q \, dx\right)^{1/p}$$

et on trouve

$$\left(\int_{\mathbb{R}^N} \rho^q \, dx\right)^{1+2p'/(Np)} \leq C \, \|\rho\|_p^{2p'/N} \operatorname{Tr}(\overline{H}\rho)$$

et (A.6) est démontrée.

La preuve de (A.9) est un exercice élémentaire de réarrangement. En effet, si on note φ_k une collection (finie ou dénombrable) orthonormée de vecteurs propres associés aux μ_k et si on pose $p_{jk} = \int_{\mathbb{R}^N} \psi_j \varphi_k^*$, on voit que

$$\operatorname{Tr}\left((\overline{H} - t\rho^\alpha)\rho\right) \geq \sum_{j,k} |p_{jk}|^2 \, \mu_k \lambda_j \, .$$

Or $\mu_1 \leq \mu_2 \leq \cdots < 0$, $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$, $\sum_j |p_{jk}|^2 = 1$, $\sum_k |p_{jk}|^2 \leq 1$. On déduit alors aisément (A.9) de ces informations.

Il reste à démontrer (A.7). Pour ce faire, on observe que $|j| \leq \rho^{1/2} e^{1/2}$ p.p. où $e = \sum_{j\geq 1} \lambda_j |\nabla \psi_j|^2$. Et on déduit aisément (A.7) de (A.6) grâce à l'inégalité de Cauchy-Schwarz en remarquant que

$$\frac{1}{r} = \frac{(N+1)p - (N-1)}{(N+2)p - N} = \frac{1}{2}\frac{Np - (N-2)}{(N+2)p - N} + \frac{1}{2} \, .$$

Références.

[1] Aizenman, M. et Simon, B., Brownian motion and Harnack's inequality for Schrödinger operators. *Comm. Pure Appl. Math.* **35** (1982), 209-271.

[2] Arnold, A., Degond, P., Markowich P. A. et Steinrück, H., The Wigner-Poisson problem in a crystal. *Appl. Math. Lett.* **2** (1989), 187-191.

[3] Arnold, A. et Markowich, P. A., Quantum transport models for semiconductors, in *Applied and Industrial Mathematics*, éd. R. Spigler, Kluwer, 1991.

[4] Arnold A. et Nier, F., The two-dimensional Wigner-Poisson problem for an electron gas in the charge neutral case. *Math. Methods Appl. Sci.* **14** (1991), 595-613.

[5] Arnold A. et Steinrück, A., The "electromagnetic" Wigner equation for an electron with spin. *Z.A.M.P.* **40** (1989), 793-815.

[6] Brezzi F. et Markowich, P. The three-dimensional Wigner-Poisson problem: existence, uniqueness and approximation. *Math. Methods Appl. Sci.* **14** (1991), 35-61.

[7] Cazenave, Th. et Haraux, A., *Introduction aux problèmes d'évolution semilinéaires.* Mathématiques et Applications, Ellipses, 1990.

[8] Córdoba A. et Fefferman, C., Fourier integral operators and propagation of wave packets. *Comm. Partial Diff. Equations* **3** (1978), 979.

[9] R. Dautray (éd.), *Méthodes probabilistes pour les équations de la Physique.* Collection CEA, Eyrolles, 1989.

[10] Degond, P. et Markowich, P., A mathematical analysis of quantum transport in three dimensional crystals. *Annali Mat. Pura Appl.* **160** (1991), 171-191.

[11] Degond, P. et Markowich, P., A quantum-transport model for semiconductors: the Wigner-Poisson problem on a bounded Brillouin zone. *RAIRO Mode. Math. Anal. Num.* **24** (1990), 697-710.

[12] DiPerna, R. J. et Lions. P. L., Solutions globales d'équations du type Vlasov-Poisson. *C.R. Acad. Sci. Paris* **307** (1988), 655-658.

[13] DiPerna R. J. et Lions, P. L., Global weak solutions of kinetic equations. *Sem. Mat. Torino* **46** (1988), 259-288.

[14] DiPerna, R. J. et Lions, P. L., Ordinary differential equations, Sobolev spaces and transport theory. *Invent. Math.* **98** (1989), 511-547.

[15] DiPerna, R. J. et Majda, A., Concentration regularizations for $2-D$ incompressible flow. *Comm. Pure Appl. Math.* **40** (1987), 301-345.

[16] DiPerna, R. J. et Majda, A., Reduced Hausdorff dimension and concentration-cancellation for $2-D$ incompressible flow. *J. Amer. Math. Soc.* **1** (1988), 59-95.

[17] Frensley, W. R., Wigner function model of a resonent-tunneling semiconductor device. *Phys. Rev. B* **36** (1987), 1570-1580.

[18] Gérard, P., Microlocal defect measures. *Comm. Partial Diff. Equations* **16** (1991), 1761-1794.

[19] Gérard, P., Mesures semi-classiques et ondes de Bloch, in *Séminaire EDP* 1990-1991, Ecole Polytechnique, Palaiseau, 1991.

[20] Gérard, P., Lions, P. L. et Paul, T., travail en préparation.

[21] Ginibre J. et Velo, G., On the global Cauchy problem for some nonlinear Schrödinger equations. *Ann. Inst. H. Poincaré. Analyse non linéaire* **1**

(1984), 309-323.

[22] Ginibre J. et Velo, G., On a class of nonlinear Schrödinger équations with nonlocal interaction. *Math. Z.* **170** (1980), 109-136.

[23] Grimwall, G., *The electron-phonon interaction in metals.* North-Holland, 1981,

[24] Kato, T., Schrödinger operators with singular potentials. *Israel J. Math.* **13** (1973), 135-148.

[25] Klauder, J. R. et Skagerstam, *Coherent states.* World scientific, 1981.

[26] Lieb, E. H. et Thirring, W., Inequalities for the moments of the eigenvalues of the Schrödinger Hamiltonian and their relation to Sobolev inequalities, in *Studies in Mathematical Physics, Essays in Honor of Valentine Bargmann*, E. H. Lieb, B. Simon, W. Thirring (éds), Princeton Univ. Press, 1976.

[27] Lions, P. L., The concentration-compactness principle in the Calculus of Variations. The locally compact case. I, *Ann. Inst. H. Poincaré, Analyse non linéaire* **1** (1984), 109-145; II, *Ann. Inst. H. Poincaré Analyse non linéaire* **I** (1984), 223-283.

[28] Lions, P. L., The concentration-compactness principle in the Calculus of Variations. The limit case. I, *Revista Mat. Iberoamericana* **1** (1985), 145-201; II, *Revista Mat. Iberoamericana* **1** (1985), 45-121.

[29] Lions, P. L. et Perthame, B., Propagation of moments and regularity for the 3-dimensional Vlasov-Poisson system. *Invent. Math.* **105** (1991), 415-430.

[30] Lions, P. L. et Perthame, B., Lemmes de moments, de moyenne et de dispersion. *C. R. Acad. Sci. Paris, Série I* **314** (1992), 801-806.

[31] Markowich, P., Boltzmann distributed quantum steady states and their classical limit. Preprint.

[32] Markowich, P., On the equivalence of the Schrödinger and the quantum Liouville equation. *Math. Meth. Appl. Sci.* **11** (1989), 459-469.

[33] Markowich, P. et Mauser, N., communication personnelle.

[34] Markowich, P. et Ringhofer, C., An analysis of the quantum Liouville equation. *Z.A.M.M.* **69** (1989), 121-127.

[35] Maslov, V. P., Non-standard characteristics in asymptotical problems, in *Proceedings of the International Congress of Mathematicians,* (ICM Warszawa 1982), Warsaw, 1983, Vol. I, North-Holland, 1984.

[36] Ravaioli, U., Osman, M. A., Pötz, W., Kluksdahl, N. et Ferry, D. K., Investigation of ballistic transport through resonant tunneling quantum wells using Wigner function approach. *Physica* **134B** (1985), 36-40.

[37] Schrödinger, E., Quantizierung als Eigenwertproblem. *Annalen der Physik* (1926), **79** 361-376 et 489-527, **80** 437-490 et **81** 109-139.

[38] Simon, B., *Functional integration and quantum physics.* Academic Press, 1979.
[39] Steinrück, H., The one-dimensional Wigner-Poisson problem and its relation to the Schrödinger-Poisson problem. *SIAM* (1990).
[40] Steinrück, H., The Wigner-Poisson problem in a crystal. Existence, uniqueness, semi-classical limit in the one-dimensional case. *Z.A.M.M.* (1990).
[41] Tartar, L., H-measures, a new approach for studying homogeneization, oscillations and concentration effects in partial differential equations. *Proc. Roy. Soc. Ed.* **115 A** (1990), 193-230.
[42] Tatarskii, V. I., The Wigner representation of quantum mechanics. *Sov. Phys. Usp.* **26** (1983), 311-327.
[43] Thirring, W. *Quantum Mechanics of large systems.* A course in Mathematical Physics, **4**. Springer, 1983.
[44] Weyl, H., *The theory of groups and quantum mechanics*, Dover, 1950 (orig.1931).
[45] Wigner, E., On the Quantum Correction for Thermodynamic Equilibrium. *Phys. Rev.* **40** (1932), 749-759.
[46] Yvon, J., Sur les rapports entre la théorie des mélanges et la statistique classique. *C. R. Acad. Sci. Paris* **223** (1946), 347-349.
[47] Yvon, J., Théorie quantique et classique, in *Mécanique Statistique*, éd. A. Blanc-Lapierre, Masson, 1967.

Recibido: 3 de diciembre de 1.992

Pierre Louis Lions et Thierry Paul
CEREMADE
Université Paris-Dauphine
Place de Lattre de Tassigny
75775 Paris Cedex 16, FRANCE