BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 9, Number 2, September 1983

# THE UNCERTAINTY PRINCIPLE

BY CHARLES L. FEFFERMAN[1]

ABSTRACT. If a function $\psi(x)$ is mostly concentrated in a box $Q$, while its Fourier transform $\hat{\psi}(\xi)$ is concentrated mostly in $Q'$, then we say $\psi$ is microlocalized in $Q \times Q'$ in $(x, \xi)$-space. The uncertainty principle says that $Q \times Q'$ must have volume at least 1. We will explain what it means for $\psi$ to be microlocalized to more complicated regions $\mathcal{B}$ of volume $\sim 1$ in $(x, \xi)$-space. To a differential operator $P(x, D)$ is associated a covering of $(x, \xi)$-space by regions $\{\mathcal{B}_\alpha\}$ of bounded volume, and a decomposition of $L^2$-functions $u$ as a sum of "components" $u_\alpha$ microlocalized to $\mathcal{B}_\alpha$. This decomposition $u \to (u_\alpha)$ diagonalizes $P(x, D)$ modulo small errors, and so can be used to study variable-coefficient differential operators, as the Fourier transform is used for constant-coefficient equations. We apply these ideas to existence and smoothness of solutions of PDE, construction of explicit fundamental solutions, and eigenvalues of Schrödinger operators. The theorems are joint work with D. H. Phong.

## CHAPTER I: THE *SAK* PRINCIPLE

The uncertainty principle says that a function $\psi$, mostly concentrated in $|x - x_0| < \delta_x$, cannot also have its Fourier transform $\hat{\psi}$ mostly concentrated in $|\xi - \xi_0| < \delta_\xi$, unless $\delta_x \cdot \delta_\xi \geq 1$. This simple fact has far-reaching consequences for PDE, but until recently it was used only in a very crude form. The most significant classical application concerned the eigenvalues of a selfadjoint differential operator

$$A(x, D) = \sum_{|\alpha| \leq m} a_\alpha(x) \left( \frac{1}{i} \frac{\partial}{\partial x} \right)^\alpha$$

with symbol $A(x, \xi) = \sum_{|\alpha| \leq m} a_\alpha(x) \xi^\alpha$. According to the uncertainty principle, each box

$$\mathcal{B} = \{(x, \xi) | \, |x - x_0| < \delta, \, |\xi - \xi_0| < \delta^{-1}\}$$

should count for one eigenvalue, so the number of eigenvalues of $A(x, D)$ which are less than $K$ should be given approximately as the volume of the set $S(A, K) = \{(x, \xi) \mid A(x, \xi) < K\}$. If $A$ is elliptic and $K \to \infty$, then this "volume-counting" is asymptotically correct (see Weyl [**41**], Carleman [**5**], Hörmander [**23**]). However, volume-counting can also produce grossly inaccurate estimates for systems as simple as two uncoupled harmonic oscillators.

---

This is an expanded version of a Colloquium Lecture of the same title presented at the 89th annual meeting of the American Mathematical Society in Denver, Colorado, January 5–8, 1983; received by the editors February 1, 1983.

1980 *Mathematics Subject Classification.* Primary 35-02, 35H05, 35P15, 35S05, 42B20, 42B25, 81H05.

[1]Supported by National Science Foundation Grant MCS 80-03072.


0273-0979/83 $1.00 + $.25 per page

We can do much better by taking the uncertainty principle more literally: Instead of measuring the importance of a subset $E \subseteq R^n \times R^n$ by its volume, we use instead the number of distorted unit cubes $\mathcal{B}$ which can be packed disjointly inside $E$. We shall see many examples in which *no* distorted unit cubes can be packed inside $E = S(A, K)$, even though $E$ has large volume. In this case $E$ counts for no eigenvalues, even though the classical approximation assigns it many. Packing distorted cubes into $S(A, K)$ rather than taking the volume amounts to a sharper form of the uncertainty principle. We shall call it the $SAK$ principle.

Now the $SAK$ principle is also important in questions of existence and regularity of solutions of PDE. For, these questions may be reduced by standard functional analysis to a priori estimates which take the form

$$(\dagger) \qquad c\|P(x,D)u\| \leq \|Q(x,D)u\| + \text{ small error},$$

where $P$ and $Q$ are differential (or slightly more general pseudodifferential) operators. We would like very much to know whether a given estimate of the form (†) holds for $u \in L^2$. The most naïve idea is to compare the symbols $P(x,\xi)$ and $Q(x,\xi)$ and guess that (†) holds if

$$(\dagger\dagger) \qquad |P(x,\xi)| \leq Q(x,\xi) + \text{ small error}.$$

This is true, although the proof is hard. However, the $SAK$ principle suggests that we do not need (††) in order to have the estimate (†). Indeed, a function $u$ can be localized in $(x,\xi)$-space no further than to a distorted unit box $\mathcal{B}$, and therefore the necessary and sufficient condition for (†) will be

$$c \max_{\mathcal{B}} |p| \leq \max_{\mathcal{B}} q + \text{ small error} \quad \text{for each } \mathcal{B},$$

which is weaker than (††). From these results we will give a unified discussion of some of the main results in linear PDE.

The application of $SAK$ to differential equations goes beyond a priori estimates. Our real goal is to diagonalize a variable-coefficient differential operator modulo small errors. Clearly this will give a powerful hold on existence, regularity, and a priori estimation of solutions, and on eigenvalues in the selfadjoint case. Moreover, it should make possible the construction of explicit approximate solutions.

Now the approximate diagonalization proceeds by cutting phase space $R^n \times R^n$ into suitable distorted boxes $\{\mathcal{B}_\nu\}$ of volume $\sim 1$. We shall write an arbitrary $u \in L^2$ as a sum of pieces, $u = \sum_\nu u_\nu$, so that $u_\nu$ together with its Fourier transform $\hat{u}_\nu$ are somehow "localized" inside $B_\nu$. Since the given differential operator $L$, acting on each piece $u_\nu$, is approximately multiplication by a scalar $\Lambda_\nu$, our decomposition will approximately diagonalize $L$.

To illustrate the ideas we make a first crude attempt to diagonalize

$$L(x,D) = \sum_{|\alpha| \leq m} a_\alpha(x) \left(\frac{1}{i}\frac{\partial}{\partial x}\right)^\alpha.$$

Let $L(x,\xi) = \sum_{|\alpha|\le m} a_\alpha(x)\xi^\alpha$ be the symbol of $L(x,D)$. The operator $L$ is made up of two ingredients: differentiation; and multiplication by smooth functions. To construct an approximate eigenfunction $u$ of $L(x,D)$, we want to come as close as possible to diagonalizing these operators simultaneously. Now $(1/i)(\partial/\partial x_k)$ is approximately multiplication by $\xi_k^0$ if the operators act on a function $u$ with Fourier transform concentrated near $\xi^0 \in R^n$. On the other hand, $u \to a(x)u$ is approximately multiplication by $a(x^0)$ if the operator acts on a function $u$ which is concentrated near $x^0$. So our approximate eigenfunction $u$ should be concentrated near $x^0$, while its Fourier transform $\hat{u}$ should be concentrated near $\xi^0$. The uncertainty principle tells us how well we can succeed in realizing the two conflicting goals. Basically, the best we can do is take $\phi((x-x_0)/\delta)e^{i\xi^0\cdot x}$, with $\phi$ a fixed Schwartz function; this is "microlocalized" to a box $\{|x-x^0| < \delta,\ |\xi-\xi^0| < \delta^{-1}\} = \mathcal{B}$, and we give our localized function $u$ the name $\phi_{\mathcal{B}}$.

Now to diagonalize $L(x,D)$ approximately, we cut phase space $R^n \times R^n$ into suitable boxes $\mathcal{B}_\nu = \{(x,\xi) |\ |x-x_\nu| < \delta_\nu,\ |\xi-\xi_\nu| < \delta_\nu^{-1}\}$ as in Figure 1, and to each $\mathcal{B}_\nu$ we associate the typical function $\phi_{\mathcal{B}_\nu}$ microlocalized to $\mathcal{B}_\nu$.

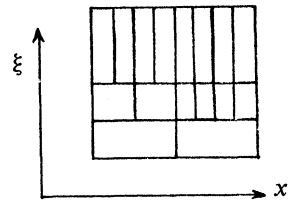



FIGURE 1

Roughly speaking, the $\{\phi_{\mathcal{B}_\nu}\}$ are orthogonal and form a basis for $L^2$, while

$$L(x,D)\phi_{\mathcal{B}_\nu} = L(x_\nu,\xi_\nu)\cdot\phi_{\mathcal{B}_\nu} + \text{Error}_\nu \tag{1}$$

with $\|\text{Error}_\nu\| \le O(|\xi_\nu|^{m-s})\|\phi_{\mathcal{B}_\nu}\|$. Here, $s$ depends on the geometry of the partition $\{\mathcal{B}_\nu\}$. By picking a good partition, we can make $s = 1/2$ (see [25, 8]).

So we have succeeded quite simply in diagonalizing an $m$th order operator $L(x,D)$ modulo errors of order $m-1/2$. In the easiest case of an elliptic operator (such as the Laplacian), the symbol $L(x_\nu,\xi_\nu)$ is of size $|\xi_\nu|^m$, so the error in (1) is negligibly small compared to the main term for large $|\xi|$. This approximate diagonalization easily gives another proof of the standard elliptic regularity theorem. The trouble comes when $L(x,D)$ is nonelliptic. The interesting phenomena in PDE are governed by the behavior of the symbol near the characteristic variety $V = \{(x,\xi) \mid L(x,\xi) = 0\}$. If in our approximate diagonalization we look at a box $\mathcal{B}_\nu$ that meets $V$, then in effect the "main" term in (1) is zero, and all the interesting phenomena are decided by the behavior of the "negligible" $\text{Error}_\nu$. Clearly, we have to do better.

What we will do is to cut phase space differently, using bent boxes $\mathcal{B}_\nu$ as in Figure 2.

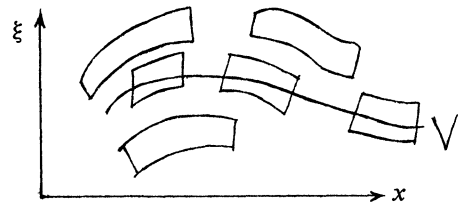



FIGURE 2

The boxes $\mathcal{B}_\nu$ still have volume $\sim 1$, but now they are bent into strange shapes, and their geometry is related to the characteristic variety $V$. This time we will find that the error terms in the analogue of (1) are small compared to the main terms, so the approximate diagonalization has nontrivial applications. Note that it isn't immediately clear what it means to say that a given $u \in L^2$ is microlocalized into a curved box $\mathcal{B}_\nu$ as in Figure 2. So, given a symbol $L(x,\xi)$, we shall have to answer the following questions:

How should we cut $R^n \times R^n$ into bent boxes $\mathcal{B}_\nu$?

How can we associate to $\mathcal{B}_\nu$ a natural projection operator $\pi_\nu$ whose image consists of functions "microlocalized" to $\mathcal{B}_\nu$?

How does $L(x,D)$ act on functions microlocalized to $\mathcal{B}_\nu$?

These questions are not easy. To understand them we need a technique for cutting and bending symbols $L(x,\xi)$. The technique can be understood on three different levels, of which the simplest is as follows.

LEVEL I (*Cutting all operators at once into big pieces modulo lower-order errors*). This is what specialists in PDE usually call microlocal analysis. It provides a powerful method, the "algorithm of the '70s" to prove theorems on PDE. The method is analogous to studying a nondegenerate vector field $X$ by first using a partition of unity to reduce matters to a local question, and then straightening out the vector field locally by a smooth change of coordinate, so the local question is reduced to the trivial case $X = \partial/\partial x_1$. We shall make partitions of unity and changes of variable in $(x,\xi)$-space by using pseudodifferential and Fourier integral operators, which we now briefly recall.

*Pseudodifferential operators.* The Fourier inversion formula shows that a differential operator $L(x,D)$ with symbol $L(x,\xi)$ is given by

$$L(x,D)u(x) = \int e^{ix\cdot\xi} L(x,\xi)\hat{u}(\xi)\,d\xi.$$

This formula makes sense even when $L(x,\xi)$ is not a polynomial in $\xi$, and $L(x,D)$ is called the pseudodifferential operator with symbol $L(x,\xi)$. If the symbols $L(x,\xi)$ satisfy suitable estimates, then the pseudodifferential operators $L(x,D)$ can be manipulated just like differential operators. The estimates on $L(x,\xi)$ are important because they determine how finely we can cut up phase space. Classically, one says that $L(x,\xi)$ is an $m$th order symbol ($L \in S^m$) if

$$|\partial_x^\alpha \partial_\xi^\beta L(x,\xi)| \le C_{\alpha\beta}(1+|\xi|)^{m-|\beta|}. \tag{2}$$

These estimates hold if $m$ is a positive integer and $L$ is the symbol of an $m$th order differential operator.

If $A \in S^{m_1}$ and $B \in S^{m_2}$, then the composed operator $A(x,D)B(x,D)$ is again a pseudodifferential operator whose symbol $A \circ B$ is given asymptotically, modulo symbols of arbitrarily large negative order, by Leibnitz' rule

$$A \circ B \sim \sum_{\alpha} \frac{1}{\alpha!}\left[\left(\frac{1}{i}\frac{\partial}{\partial \xi}\right)^{\alpha} A\right] \cdot \left[\left(\frac{\partial}{\partial x}\right)^{\alpha} B\right],$$

while the adjoint $A(x,D)^*$ is a pseudodifferential operator with symbol

$$A^{\#} \sim \sum_{\alpha} \frac{1}{\alpha!}\left(\frac{1}{i}\frac{\partial}{\partial \xi}\right)^{\alpha}\left(\frac{\partial}{\partial x}\right)^{\alpha} \overline{A}.$$

In particular, $A(x,D)B(x,D) = AB(x,D)$ modulo terms of lower order, and $[A(x,D), B(x,D)] = i\{A,B\}(x,D)$ modulo terms of lower order. Here

$$\{A,B\} = \sum_{k}\left(\frac{\partial A}{\partial \xi_k}\frac{\partial B}{\partial x_k} - \frac{\partial B}{\partial \xi_k}\frac{\partial A}{\partial x_k}\right)$$

is the Poisson bracket, which we shall meet again many times.

The above remarks justify the statement that pseudodifferential operators can be manipulated like differential operators. Their proofs involve a straightforward application of the method of stationary phase to evaluate some integrals of rapidly oscillating exponentials (see [3]). We should also point out that pseudodifferential operators ($\psi d0$) of order zero are bounded on $L^2$.

The earliest application of $\psi d0$ was to invert elliptic differential operators. If $A(x,\xi) \in S^m$ is an elliptic symbol, i.e., $|A(x,\xi)| \geq c(1+|\xi|)^m$, then $A^{-1}(x,\xi)$ is a symbol in $S^{-m}$, so the composition law for $\psi d0$ yields $A(x,D)A^{-1}(x,D) = A^{-1}(x,D)A(x,D) = I$ modulo symbols of order $-1$. An easy successive approximation argument lets us add lower-order corrections to the symbol $A^{-1}(x,\xi)$ so that $A(x,D)$ is inverted modulo symbols of arbitrarily large negative order. Thus $A(x,D)u = f$ is solved explicitly, modulo smooth errors.

Using $\psi d0$ we can decompose a differential operator $L(x,D)$ as an approximate direct sum by cutting phase space into blocks $\mathcal{B}_\nu$ as in Figure 3.

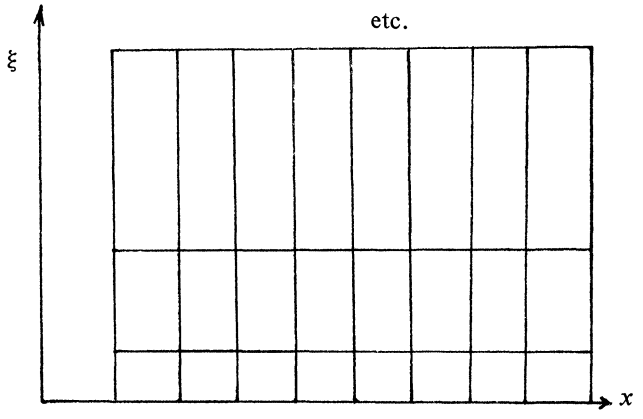



FIGURE 3

Here, each $\mathcal{B}_\nu$ is centered at $(x_\nu, \xi_\nu)$ and has sides 1 in the $x$-directions, and $\sim \frac{1}{4}(1+|\xi_\nu|)$ in the $\xi$-directions. Note that most of the $\mathcal{B}_\nu$ have volume $\gg 1$. We decompose $L(x,D)$ by means of a partition of unity $1 = \sum_\nu \phi_\nu^2(x,\xi)$ in phase space, with $\phi_\nu$ supported essentially on $\mathcal{B}_\nu$ and as smooth as possible.

The point is that Figure 3 gives exactly the finest possible cutting of phase space so that the $\phi_\nu$ belong uniformly to $S^0$; in fact $(x,\xi) \to (\phi_\nu(x,\xi))_{\nu\in Z}$ belongs to $S^0$ as a vector-valued symbol. Therefore, the operator $u\colon f \to (f_\nu)_{\nu\in Z} = (\psi_\nu(x,D)f)_{\nu\in Z}$ is a vector-valued $\psi d0$. Now from the formulas for composition and adjoints of $\psi d0$'s we obtain $U^*U = I$ and $UL(x,D) = L(x,D)U$ modulo lower-order errors. Since $f_\nu$ is microlocalized to $\mathcal{B}_\nu$, we have succeeded in approximately splitting $L(x,D)$ as a direct sum of microlocalized operators acting on $f_\nu$. As claimed, $\psi d0$ let us use partitions of unity in $(x,\xi)$-space. Of course, we are still far from diagonalizing $L(x,D)$, since the blocks $\mathcal{B}_\nu$ of Figure 3 have large volume.

In addition to cutting symbols we shall also bend them, using

*Fourier integral operators.* These generalize a simple change of variable $y = \phi(x)$ to allow changes of variable in $(x,\xi)$ together. Under $y = \phi(x)$, the differential equation $L(x,D)u = f$ goes over to $\tilde{L}(y,D)\tilde{u} = \tilde{f}$, with $\tilde{L}$ given modulo lower-order terms by

$$(0)\quad \tilde{L}(y,\eta) = L\circ\Phi(y,\eta), \qquad \Phi\colon (y,\eta)\to(x,\eta) \quad \text{with } x = \phi^{-1}(y),\ \xi = (\phi'(x))^{\dagger}\eta.$$

The transformation $\Phi$ has a very special property: it preserves Poisson brackets, i.e.,

$$(\mathrm{x})\qquad \{F,G\}\circ\Phi = \{F\circ\Phi, G\circ\Phi\}.$$

This is natural in view of the formula for commutators of differential operators in terms of $\{\ ,\ \}$. Transformations $\Phi$ which satisfy (x) are called *canonical.* There are many canonical transformations which do not arise from a simple change of coordinate $y = \phi(x)$. Canonical transformations preserve volume in $R^n \times R^n$.

To repeat, we know that (0) defines a canonical transformation $\Phi$, and that the equations $L(x,D)u = f$, $\tilde{L}(y,D)\tilde{u} = \tilde{f}$ are equivalent, where $\tilde{L}(y,\eta) = L\circ\Phi(y,\eta)$ modulo lower terms. The equivalence is given by $\tilde{u} = Uu$, $\tilde{f} = Uf$, $\tilde{L}(y,D) = UL(x,D)U^{-1}$, and $Uf(y) = |\det D\phi^{-1}(y)|^{1/2} f\circ\phi^{-1}(y)$. (We inserted the harmless determinant factor to make $U$ unitary.)

Now Egorov had the simple, deep idea that the same kind of equivalence connects $L(x,D)$ and $\tilde{L}(y,D)$, even when $\tilde{L}(y,\eta) = L\circ\Phi$ for canonical transformations not arising from a coordinate change $y = \phi(x)$. To state the result precisely, we work on a block $\mathcal{B}$ of size $1\times M$ taken from Figure 3. Suppose the block is centered at $(x^0,\xi^0)$ and denote by $i$ the natural change of scale $i\colon (x,\xi)\to(x-x^0,(\xi-\xi^0)/M)$ which carries $\mathcal{B}$ to the unit cube.

A canonical transformation $\Phi\colon (y,\eta)\to(z,\varsigma)$ defined on $\mathcal{B}$ will be said to satisfy "natural estimates" if $i\Phi i^{-1}$ is a $C^\infty$ map with derivatives of all orders bounded independent of $M$.

THEOREM (EGOROV). *Let $\Phi$ be a canonical transformation satisfying natural estimates and carrying $\mathcal{B}$ into its double $\mathcal{B}^*$. Let $A(z,\zeta) \in S^m$ be a symbol supported in $\Phi(\mathcal{B})$ and define $\tilde{A}(y,\eta) = A \circ \Phi(y,\eta)$. Then the operators $A(z,D)$ and $\tilde{A}(y,D)$ are related by*

$$\tilde{A}(y,D) = UA(z,D)U^{-1} + \text{lower-order terms}$$

*for a suitable unitary transformation $U$.*

For "most" $\Phi$, the operator $U$ is given explicitly as a Fourier integral operator

$$Uf(y) = \int e(y,\zeta)e^{iS(y,\zeta)}\hat{f}(\zeta)\,d\zeta \quad \text{with } e \in S^0,\ S \in S^1.$$

In case $\Phi$ arises from $y = \phi(x)$ by (0), we set $e(y,\zeta) = |\det D\phi^{-1}(y)|^{1/2}$, $S(y,\zeta) = \phi^{-1}(y) \cdot \zeta$, and the Fourier inversion formula yields the familiar $Uf(y) = |\det(etc)|^{1/2} f \circ \phi^{-1}(y)$. For more general $\Phi$, the function $S$ is related to $\Phi$ by

$$\{(y,\eta,z,\zeta) | \Phi(y,\eta) = (z,\zeta)\} = \left\{(y,\eta,z,\zeta) \mid \eta_k = \frac{\partial S(y,\zeta)}{\partial y_k},\ z_k = \frac{\partial S(y,\zeta)}{\partial \zeta_k}\right\}.$$

As in the calculus of $\psi d0$, proving Egorov's theorem amounts to calculating some explicit integrals of rapidly oscillating exponentials, and the argument is quite easy.

So now we know that $L(x,D)u = f$ and $\tilde{L}(x,D)\tilde{u} = \tilde{f}$ are equivalent if the symbols are related by a suitable canonical transformation. In other words, we can bend symbols as well as cut them.

Now we can describe the "algorithm of the '70s" for proving theorems in PDE. First solve your favorite PDE, say $\partial u/\partial x_1 = f$. Next formulate a condition on symbols that locally characterizes the example up to canonical transformations. For instance, a real symbol with only simple zeros is locally equivalent to $\xi_1$ after a canonical transformation and multiplication by an elliptic symbol. Finally, we conclude that all PDE whose symbols satisfy the given condition can be solved. The reason is that we can first use $\psi d0$ to cut the original problem into pieces microlocalized to the boxes $\mathcal{B}_\nu$ of Figure 3, and then in each $\mathcal{B}_\nu$ use Egorov's theorem to bring the problem back to the example we started with. The method is remarkably powerful.

Before leaving standard microlocal analysis, we should point out an analogy between PDE and quantum mechanics. This makes it plausible that the uncertainty principle has something to do with PDE. We start by reviewing classical mechanics. The state of a classical system is specified by the coordinates $x_i$ and momenta $\xi_i = m_i(dx_i/dt)$ of its particles. An observable quantity (e.g. angular momentum) is given by a function $F(x,\xi)$. If we observe $F$ when the system is in state $(x^0,\xi^0)$, we get a deterministic answer $F(x^0,\xi^0)$. Of particular importance is the observable

$$H = \sum_{i=1}^{N} \frac{1}{2m_i}\xi_i^2 + V(x_1,\ldots,x_N),$$

the total energy or Hamiltonian. Here $V$ is the potential in which the particles move. Newton's equations of motion say that the classical system evolves by $m_i(d^2x_i/dt^2) = -\partial V/\partial x_i$, which we rewrite as

$$\frac{dx_i}{dt} = \frac{1}{m_i}\xi_i, \qquad \frac{d\xi_i}{dt} = -\frac{\partial V}{\partial x_i}, \tag{3}$$

Hamilton's equations. The time evolution of any observable $F(x,\xi)$ is given in terms of the Poisson bracket by

$$dF/dt = \{H, F\}. \tag{4}$$

This amounts to Hamilton's equations when $F = x_i$ or $\xi_i$, and then follows in general by the chain rule. So (4) provides a complete description of how the system moves. The laws of mechanics are now clearly invariant under canonical transformations.

On the other hand, in quantum mechanics, the state of a system is described by a vector $\psi \in L^2$. An observable quantity is a selfadjoint operator $A$ on Hilbert space. For instance, position corresponds to the operator $\psi \to x_j\psi$, while momentum corresponds to $\psi \to (1/i)(\partial/\partial x_j)\psi$. If we measure the observable $A$ when the system is in state $\psi$, then the outcome is probabilistic, but the average observed value will be $\langle A\psi, \psi\rangle$. In the Schrödinger picture (which we have been describing), the state $\psi$ evolves in time according to $d\psi/dt = iH\psi$, where the Hamiltonian operator $H$ specifies the physics of the system. In the equivalent Heisenberg picture, the state remains constant, but observables $A$ evolve according to $dA/dt = i[H, A]$.

Note that the laws of quantum mechanics are invariant under the action of a unitary transformation $U\colon L^2 \to L^2$. In fact, sending $\psi \to U\psi = \tilde{\psi}$, $A \to UAU^{-1} = \tilde{A}$ for observables $A$ preserves all the equations and predicts the same outcome of any experiment.

We can summarize this elementary discussion by a table:

| | *Classical* | *Quantum* |
|---|---|---|
| • State of system | $(x, \xi)$ | $\psi \in L^2$ |
| • Observable | $F$ function | $A$ operator |
| • Result of measuring observable | Deterministic; always $F(x,\xi)$ | Probabilistic; on average $\langle A\psi, \psi\rangle$ |
| • Object controlling dynamics | Poisson bracket $\{\ ,\ \}$ | Commutator $i[\ ,\ ]$ |
| • Change to equivalent viewpoint | Canonical transformation | Unitary operator |

Clearly, standard microlocal analysis amounts to quantization. By $\psi d0$ we pass from functions of $(x,\xi)$ to operators in such a way that Poisson brackets go over to commutators. Fourier integral operators let us pass from canonical transformations to unitary operators. It is very natural that the uncertainty principle should play an important role in PDE.

In particular, we can look ahead to the decomposition of phase space $R^n \times R^n$ into curved boxes of volume $\sim 1$. We can now begin to describe what the curved boxes look like: They are images of a cube of side $\sim 1$ under canonical transformations.

Let us return to the techniques of cutting and bending symbols.

The calculus of $\psi d0$ together with Egorov's theorem yield the finest possible cutting and bending that work simultaneously on all symbols in $S^m$.

LEVEL II (*Cutting a single operator into smaller pieces modulo a lower-order error*). To make further progress we have to cut phase space into much smaller pieces than the blocks $\{\mathcal{B}_\nu\}$ of Figure 3. This time the cutting will depend on the particular symbol $A(x,\xi)$ we are trying to understand. The idea is to bisect repeatedly the $\mathcal{B}_\nu$ until we arrive at a family of blocks $\{\mathcal{B}_{\nu j}\}$ on which $A(x,\xi)$ is somehow "nondegenerate". So the $\{\mathcal{B}_{\nu j}\}$ form a Calderón-Zygmund decomposition of $\mathcal{B}_\nu$ as in Figure 4A, and the whole of phase space is cut up as in Figure 4B.

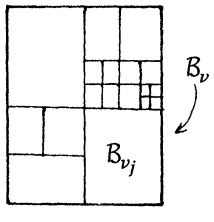



FIGURE 4A

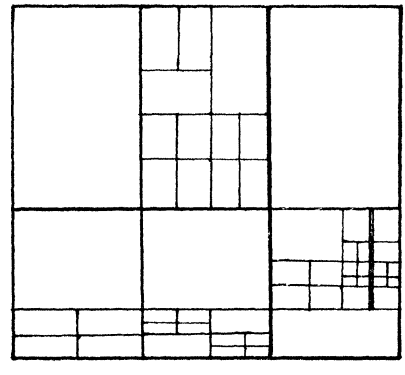


FIGURE 4B

For a general symbol $L(x,\xi)$ unrelated to $A(x,\xi)$, the decomposition of Figure 4B would be too fine: If we try to represent $L(x,D)$ as an approximate direct sum of operators $L_{\nu j}(x,D)$ microlocalized to the boxes of Figure 4B, then we would find that the error terms are large. However, the decomposition is very fine precisely where the symbol $A(x,\xi)$ is small, and therefore $A(x,D)$ is well approximated by a direct sum of microlocalized pieces. In particular, the error terms are of lower order than the main terms, just as in the standard microlocalization of Level I. Figure 4B actually gives the finest possible microlocalization of $A(x,D)$ modulo lower-order errors.

It is strong enough for some useful applications (the Nirenberg-Trèves conjecture ($\mathcal{P}$), Hörmander's theorem on squares of vector fields), but we are still far from diagonalizing the operator $A(x,D)$, since the pieces $\mathcal{B}_{\nu j}$ in Figure 4B still may have large volume. So we pass to

LEVEL III (*Cutting a single operator into small enough pieces modulo a one-percent error*). In the final picture, phase space $R^n \times R^n$ is cut into curved

boxes $\mathcal{B}_\alpha$ of volume $\sim 1$ which sit inside the $\mathcal{B}_{\nu j}$ of Level II. The family $\{\mathcal{B}_\alpha\}$ again depends on the particular symbol $A(x,\xi)$ to be analyzed. Each $\mathcal{B}_\alpha$ is essentially the image of the unit cube under a canonical transformation $\Phi_\alpha$. Corresponding to the decomposition of phase space into boxes of bounded volume, we can finally approximate $A(x,D)$ by an operator which is explicitly diagonalized. Under natural hypotheses (for instance, if the symbol $A(x,\xi)$ is positive), the eigenvalue corresponding to the box $\mathcal{B}_\alpha$ is of magnitude $\sim \mathrm{MAX}_{(x,\xi)\in\mathcal{B}_\alpha} |A(x,\xi)|$. This agrees with the *SAK* principle as stated at the beginning of the chapter. The canonical transformations $\Phi_\alpha$ are far too wild to allow direct use of Egorov's theorem. We could never carry out such violent cutting and bending with errors of lower order than the main terms. Instead, we are forced to let the error grow as large as a fixed small constant times the main term. Fortunately, such errors have no effect on the applications to PDE. In particular, we can write an approximate inverse for $A(x,D)$, given sharp a priori estimates, and describe the eigenvalues.

Next we review a few of the main problems and results in linear PDE. We have picked out the topics for which approximate diagonalization and the *SAK* principle have immediate applications.

*Existence of solutions of PDE.* This question was radically transformed by the discovery of H. Lewy that the equation

$$(\dagger) \qquad \left[\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right) + (x - iy)\frac{\partial}{\partial t}\right] u = f$$

has no solutions for general $f \in C^\infty$, even if we ask only for distribution solutions $u$ defined in a small neighborhood. Equation (†) is not a cooked-up example, but arises as an analogue of the Cauchy-Riemann equations on the unit sphere in $\mathbf{C}^2$. Lewy's work led to a new question: How can we recognize those $L$ for which $Lu = f$ has local solutions?

After some preliminary work by Hörmander [involving commutators of $L$ with $L^*$], Nirenberg and Trèves found the correct conjectures and gave overwhelming evidence by proving them in many cases [**33**, **34**]. The Nirenberg-Trèves conjectures relate local solvability to the geometry of the symbol $L(x,\xi)$. To understand their condition, and to see why an equation can fail to be locally solvable, we look at a simple example:

$$L = \frac{\partial}{\partial t} + i\sum_{k=1}^{n} a_k(t)\frac{\partial}{\partial x_k}.$$

We can solve $Lu = f$ formally by making a partial Fourier transform in the $x$-variables. Thus our PDE goes over to an elementary ODE $[\partial/\partial t + \sum_k a_k(t)\xi_k]\hat{u}(t,\xi) = \hat{f}(t,\xi)$, which we solve easily using the integrating factor $\exp[\int^t \sum_k a_k(s)\xi_k\, ds]$. The trouble is that the integrating factor grows exponentially in $\xi$, so we may easily end up with a formal solution $\hat{u}(t,\xi)$ which also grows exponentially. In this case the partial Fourier transform $u$ cannot be inverted, and $Lu = f$ has no solutions. Working out the details, we arrive at the necessary and sufficient condition for solvability, namely

$(\mathcal{P})$ For $\xi \neq 0$, the function $t \to \sum_k a_k(t)\xi_k$ never changes sign.

More generally, if $L$ is a differential operator with principal symbol $p+iq$, one associates certain curves to $p$ (the null bicharacteristics), and the Nirenberg-Trèves condition is

($\mathcal{P}$) $q$ never changes sign on the curves associated to $p$.

For a large class of PDE (principal type), this is equivalent to local solvability. The proof requires microlocal analysis on what we have called Level II, and it formed the original motivation for Calderón-Zygmund decomposition of symbols.

*Hypoellipticity.* Solutions $u$ of Laplace's equation $\Delta u = f$ can be singular only where $f$ is singular. On the other hand, solutions of the wave equation $(\partial^2/\partial t^2 - \Delta)u = f$ have singularities which propagate from the singularities of $f$ along light cones. In general, an equation $Lu = f$ is called hypoelliptic if $u$ is always $C^\infty$ except where $f$ is already not $C^\infty$. It is an interesting problem to decide whether a given PDE is hypoelliptic, and to understand precisely how smooth $u$ must be if we know how smooth $f$ is.

A basic nontrivial example of a hypoelliptic equation is a sum of squares of vector fields. Already in Kohn [**26**] it was clear that commutators played an important role; see also Kolmogorov [**28**]. Hörmander generalized these examples in his celebrated

THEOREM. $L = \sum_{j=1}^{N} X_j^2 + X_0$ *is hypoelliptic if the vector fields* $X_0, \ldots, X_N$ *and their repeated commutators span the tangent space at each point.*

In view of the connection of $L$ with the Bergman and Szegö kernels in complex variables, one wants to write explicitly an approximate inverse for $L$.

To invert $L$ and give sharp estimates involves not so much the algebra of commutators, but rather a geometric study of certain non-Euclidean "balls" $B_L(x, \rho)$ associated to $L$. This key discovery is due to Stein [**37**, **19**, **20**] who used nilpotent groups as the bridge between commutators and geometry. On a nilpotent group $N$ one has natural examples of noncommuting vector fields $X_j$, namely the (left) translation-invariant vector fields that make up the Lie algebra $n$.

Since $L = \sum_j X_j^2 + X_0$ is then translation-invariant on $N$, we know that $L^{-1}$ is given as a convolution operator on $N$. Also, the convolution kernel $K(x)$ must be homogeneous with respect to the natural dilations $\delta_t$ which act on $N$. So the nature of $L^{-1}$ is well understood in this case. On the other hand, the group $N$ is equipped with a family of non-Euclidean balls: To define the ball of radius $\rho$ about the identity in $N$, we just apply the dilation $\delta_\rho$ to a fixed neighborhood of the identity (which serves as a unit ball). The fundamental solution of $L$ is intimately tied to the shape of the non-Euclidean balls in $N$.

Now an arbitrary family of noncommuting vector fields may be recovered from the special case of a nilpotent group by using a process called the Rothschild-Stein "lifting".

As a result, sharp estimates [**17**] and now also the fundamental solution of $L$ can be read off from the geometry of the balls $B_L(x,\rho)$. The fundamental solution is due, independently, to Nagel, Stein and Wainger [**32**] and Sanchez [**38**].

We also want to understand second-order operators $L$ not given as sums of squares. For an operator

$$L = -\sum_{jk} \frac{\partial}{\partial x_j} a_{jk} \frac{\partial}{\partial x_k} + \sum_j b_j \frac{\partial}{\partial x_j} + C$$

with $(a_{jk}) \geq 0$ and $a_{jk}$, $b_j$, $c$ real, Oleinik and Radkevitch [**35**] gave a condition in terms of commutators which is sufficient and close to necessary for hypoellipticity. The Oleinik-Radkevitch condition is as follows.

Start with the symbols

$$L_0 = \sum_j b_j(x)\xi_j, \quad L_j = \sum_k a_{jk}(x)\xi_k, \quad L_\mu = \sum_{jk} \frac{\partial a_{jk}}{\partial x_\mu} \xi_j \xi_k |\xi|^{-1}.$$

Next form all repeated Poisson brackets of these symbols up to order $k$; say $P_1, \ldots, P_N$ are the symbols obtained in this way.

THEOREM [**35**]. *$L$ is hypoelliptic if $\sum_1^N |P_s(x,\xi)| \geq c|\xi|$.*

This is, of course, analogous to saying that the commutators of vector fields span the tangent space.

The Oleinik-Radkevitch theorem is proved using $\psi d0$ calculus, but now the techniques of nilpotent groups are no longer available. So one has neither an explicit solution nor sharp estimates. We shall study these problems (and also get a simple proof of hypoellipticity of sums of squares) using approximate diagonalization and $SAK$.

*Boundary-value problems.* To fix the ideas look at the equation $\Delta u = 0$ in $\Omega$, $Xu = f$ on $\partial\Omega$, where $X$ is a complex vector field. We can suppose $\partial\Omega$ has been straightened out, so $\Omega = \{(x,t) \in R^n \times R^1 \mid t > 0\}$. Thus $u(x,t)$ is the Poisson integral of $\overline{u}(x) = u(x,0)$, and our problem is to find $\overline{u}(x)$. At the boundary the vector field $X$ splits up into $a(x)(\partial/\partial t) + X_{\mathrm{TAN}}$, where $a(x)$ is complex-valued and $X_{\mathrm{TAN}}$ is a complex vector field tangent to $\partial\Omega$. For $u(x,t) =$ Poisson integral of $\overline{u}(x)$, we compute that $-\partial u/\partial t|_{t=0} = (-\Delta)^{1/2}\overline{u}$, where now $\Delta$ denotes the Laplacian on $\partial\Omega$. Therefore, our boundary problem reduces to

$$[-a(x)(-\Delta)^{1/2} + X_{\mathrm{TAN}}]\overline{u} = f \quad \text{on } R^n.$$

In other words, a boundary-value problem for the Laplacian on $\Omega$ reduces to a pseudodifferential equation on the boundary $\partial\Omega$. This makes it important to solve pseudodifferential equations.

Given a symbol $L(x,\xi) = p + iq$, we now ask whether $L(x,D)u = f$ is hypoelliptic or locally solvable. A crucial condition here is the analogue of $(\mathcal{P})$

for pseudodifferential operators, namely:

($\Psi$) As we flow along the null bicharacteristics of $p$, the symbol $q$ can change sign only from minus to plus.

(This is equivalent to ($\mathcal{P}$) in the special case of differential operators, for then $p+iq$ must be even or odd, so any sign change would lead to a forbidden sign change.)

Moyer [**31**] has shown that ($\Psi$) is necessary for local solvability of $p-iq$ and for hypoellipticity of $p+iq$. One supposes ($\Psi$) implies local solvability for equations of principle type, but this is unknown and probably quite hard. Regarding hypoellipticity under condition ($\Psi$), there is a very strong result of Egorov. To state Egorov's theorem we may assume $p+iq$ is first-order, since this may be achieved by multiplying by an elliptic symbol.

THEOREM ([EGOROV [**10**]; SEE ALSO HÖRMANDER [**24**]). *Suppose $p+iq$ is a first-order symbol satisfying ($\Psi$). Let $p_1, p_2, \ldots, p_N$ denote $p$, $q$ and their repeated Poisson brackets up to order $m$. If $\sum_1^N |p_k(x,\xi)| \geq c|\xi|$ for large $\xi$, then $L=(p+iq)(x,D)$ is hypoelliptic.*

More precisely, $L$ satisfies the sharp subelliptic estimate $||u|| + ||Lu|| \geq c||u||_{1/(m+1)}$; this estimate is actually equivalent to the hypotheses of Egorov's theorem.

Egorov's original paper [**10**] is the first place a problem in PDE is solved by cutting phase space into curved boxes $\mathcal{B}_\nu$. (Egorov's boxes have large volume, however, so they do not diagonalize the equation.) Unfortunately, while giving a simple solution of the localized problems on the $\mathcal{B}_\nu$, Egorov provides no rigorous discussion of how the microlocalized results can be patched together to solve the original problem. This is a highly nontrivial task involving Level III microlocalization. Later, Hörmander [**24**] gave a careful justification of Egorov's main ideas.

From our work on approximate diagonalization and $SAK$, Egorov's theorem may be read off as a simple consequence. Approximate diagonalization thus gives a clue as to what is really going on in the very complicated arguments in [**10**, **24**].

This concludes our introduction to the $SAK$ technique. In the next chapter we apply our philosophy to the study of eigenvalues of Schrödinger operators (which is a natural starting point in view of the connection to quantum mechanics). Then we return to general PDE and state precisely our theorems on approximate diagonalization. We shall explain how to get the applications as consequences of our main theorem. The proof of our main result is very hard. Here we will do little more than sketch the ideas.

It seems to me that these techniques give strong results for a single hypoelliptic PDE. For hypoelliptic systems the analogues of our results are completely open. There are fascinating new geometric questions brought out in the context of $\overline{\partial}$ by Kohn [**27**] and Catlin [**6**]. A natural goal for future study is to understand the Bergman and Szegö kernels on weakly pseudoconvex domains. The inversion of hypoelliptic scalar operators may be regarded optimistically as a first step in attacking this very hard problem.

## CHAPTER II: SCHRÖDINGER OPERATORS

To test our philosophy we will study eigenvalues of Schrödinger operators $L = -\Delta + V(x)$ on $R^n$. Our goal will be to relate the eigenvalues and eigenfunctions of $L$ to the growth of the symbol $|\xi|^2 + V(x)$ on testing boxes $\mathcal{B} = \{(x,\xi) \in R^n \times R^n \,|\, |x - x_0| < \delta,\ |\xi - \xi_0| < \delta^{-1}\}$.

We begin by studying polynomial potentials $V$ for which the estimates are elementary. In this case, the number of eigenvalues of $L$ which are $< E$ is essentially the largest number of pairwise disjoint testing boxes $\mathcal{B}$ which fit inside $\{|\xi|^2 + V(x) < E\}$. As a consequence, we can read off the order of magnitude of each eigenvalue $\lambda_N$ for nonnegative polynomials $V$, and our philosophy is confirmed. The proofs will form a simple model for our later analysis of pseudodifferential operators.

Next we abandon polynomials and study the eigenvalues of $-\Delta + V(x)$ for completely arbitrary potentials $V(x) \leq 0$. Remarkably, estimates analogous to the easy polynomial case hold for arbitrary $V$. This time the proofs are deep and rely on the techniques of Fourier analysis on $R^n$ developed in the 1970s.

Finally, we study the special Schrödinger operator arising from the Coulomb forces among $N$ electrons and $N$ nucleii in $R^3$. Since ordinary objects have many electrons and nucleii, we look for estimates independent of $N$. Classical results of Dyson and Lenard [**9a**] and Lieb and Thirring [**30**] show that bulk matter occupies a volume proportional to the number of particles. In the spirit of the *SAK* principle, we will sharpen these results by proving that ordinary matter is made of atoms which bind together to form molecules.

Before stating our results, we summarize the classical eigenvalue estimates and show in several examples that they are not sharp. The standard philosophy is that the number $N(\lambda, L)$ of eigenvalues $< \lambda$ is approximately the phase space volume $V(\lambda, L) = |\{(x,\xi)|\, |\xi|^2 + V(x) < \lambda\}|$. Thus, the $N$th eigenvalue should be roughly the smallest $\lambda$ for which $\mathrm{Vol}(\lambda, L) = N$. Is this true? A basic theorem of Cwickel, Lieb and Rosenblum [**40**] is as follows.

THEOREM 1. *In $R^n$ $(n \geq 3)$ one has the estimate $N(\lambda, L) \leq C_n \mathrm{Vol}(\lambda, L)$.*

COROLLARY 1. *If $\mathrm{Vol}(\lambda, L) < C_n^{-1}$, then $L \geq \lambda$.*

COROLLARY 2. *The sum of the absolute values of the negative eigenvalues of $-\Delta + V(x)$ is at most $C'_n \int_{\{V(x)<0\}} |V(x)|^{(n+2)/2}\, dx$.*

Corollary 2 follows by integrating the estimate of Theorem 1 over all negative $\lambda$; in fact Corollary 2 is much easier than Theorem 1 and was proved first in Lieb and Thirring [**30**]. It has a very important application which we shall discuss later.

We pause to note that Corollary 1 is nothing but Sobolev's inequality. In fact, we may assume $\lambda = 0$ and $V \leq 0$. Corollary 1 asserts that

$$\langle Lu, u\rangle = \|\nabla u\|^2 - \int |V|\,|u|^2\, dx \geq 0$$

if

$$\mathrm{Vol}(0,1) = \int_{R^n} |V(x)|^{n/2}\,dx \le C_n^{-1}.$$

That is,

$$\int |V|\,|u|^2\,dx \le \|\nabla u\|^2 \quad \text{if } \|V\|_{L^{n/2}} \le c(n),$$

which simply means $\nabla u \in L^2$ implies $|u|^2 \in L^{n/(n-2)}$. So Theorem 1 is really a sharpening of Sobolev's inequality.

From Theorem 1 we see that the real question is whether $N(\lambda, L) \sim \mathrm{Vol}(\lambda, L)$ or, instead, $N(\lambda, L) \ll \mathrm{Vol}(\lambda, L)$; and in the second case, how big is $N(\lambda, L)$? Now let us look at a few examples.

EXAMPLE I (*Two uncoupled harmonic oscillators*). $L = -\Delta + \mu_1 x^2 + \mu_2 y^2$ on $R^2$. We fix $\mu_1 = 1$ and take $\mu_2$ to be very small. The lowest eigenvalue is $\lambda_1(L) = 1 + \sqrt{\mu_2} \approx 1$. On the other hand, given $\epsilon > 0$, $N > 1$, we can take $\mu_2$ so small that $\mathrm{Vol}(\epsilon, L) > N$. Hence volume-counting predicts huge numbers of very small eigenvalues for $L$, even though the true lowest eigenvalue is approximately 1.

The next two examples are due to B. Simon.

EXAMPLE II. $L = -\Delta + x^2 y^2$ on $R^2$. Here, the phase-space volume $\mathrm{Vol}(\lambda, L) = +\infty$ for every $\lambda > 0$. Nevertheless, $L$ has discrete eigenvalues $\lambda_N$ tending to infinity. Our results give the order of magnitude of $\lambda_N$. For the closely related Dirichlet Laplacian on $\Omega = \{|xy| < 1\} \subseteq R^2$, Simon has given precise eigenvalue asymptotics.

EXAMPLE III. Let $\mathcal{A}$ be a Lie algebra of compact type, and on $R^N = \mathcal{A} \oplus \mathcal{A} \oplus \cdots \oplus \mathcal{A}$ set $L = -\Delta + \sum_{j<k} |||[A_j, A_k]|||^2$. Here, the triple norm is given in terms of the Killing form on $\mathcal{A}$. Again $\mathrm{Vol}(\lambda, L) \equiv +\infty$, but Simon has shown that $L$ has discrete eigenvalues tending to infinity. The example arises as a (grossly oversimplified) model of quantum gauge theories. In a gauge theory the classical field is given in terms of the potentials $A_j$ by $E_{jk} = \partial A_j/\partial x_k - \partial A_k/\partial x_j + [A_j, A_k]$. If the potentials are slowly varying, one can put $E_{jk} \approx [A_j, A_k]$, and quantization leads to the Schrödinger operator $L$. Mathematically, $L$ is a more complicated version of Example II.

So far our potentials have been polynomials. Next we look at potentials with a different shape.

EXAMPLE IV. $L = -\Delta - k/|x|^2$ on $R^n$ $(n \ge 3)$. This remarkable example is well known to people interested in Schrödinger equations. For all finite $\lambda$ one has $\mathrm{Vol}(\lambda, L) = +\infty$, so it is natural to guess that $L$ is unbounded below. The correct result is that $L \ge 0$ for $k \le k_{\mathrm{critical}}(n)$, while $L$ is unbounded below for $k > k_{\mathrm{critical}}(n)$. This may be understood in terms of Corollary 1 to Theorem 1 and a sharper form of Sobolev's inequality due to R. Hunt [**21**]. Our results on singular potentials cover this example, but do not give the value of $k_{\mathrm{critical}}(n)$.

EXAMPLE V (*Particle in a box*). Let $I = I_1 \times I_2 \times \cdots \times I_n$ be a rectangular box in $R^n$ $(n \ge 3)$ whose sides $I_1, I_2, \ldots, I_n$ have lengths $\delta_1 \le \delta_2 \le \cdots \le \delta_n$. For $E > 0$ small enough, the Schrödinger operator $L = -\Delta - E\chi_I \ge 0$, while for $E > E_{\mathrm{critical}}(\delta_1, \delta_2, \ldots, \delta_n)$ we find that $L$ has negative eigenvalues, so the potential well can capture a particle. We ask how the energy $E_{\mathrm{critical}}$

depends on the sides of the box. Volume-counting leads to the guess $E_{\text{critical}} \sim (\delta_1^{-2}\delta_2^{-2}\cdots\delta_n^{-2})^{1/n}$. This is completely wrong. Curiously, the order of magnitude of $E_{\text{critical}}$ depends only on the three shortest sides:

$$E_{\text{critical}} \sim \frac{1}{\delta_1\delta_2\log((\delta_2+\delta_3)/\delta_2)}.$$

Our general results imply

$$\frac{C}{\delta_1\delta_2} \geq E_{\text{critical}} \geq \frac{c_\epsilon}{\delta_1\delta_2(\delta_3/\delta_1)^\epsilon},$$

so that $E_{\text{critical}}$ depends strongly on $\delta_1$, $\delta_2$, and weakly on $\delta_3$.

Hence, volume-counting can lead to gross errors. Later, we will give a kind of converse to Theorem 1, which characterizes these potentials $V$ for which volume-counting is approximately right.

Now we can state our eigenvalue estimates for polynomial potentials. Let $V$ be a polynomial of degree $\leq d$ on $R^n$. We guess that the number $N(\lambda, L)$ of eigenvalues $< \lambda$ for $L = -\Delta + V(x)$ is approximately the number $N_{UP}(\lambda, L)$ of pairwise disjoint testing boxes $\mathcal{B} = \{|x - x_0| < \delta, |\xi - \xi_0| < \delta^{-1}\}$ which fit inside $\{|\xi|^2 + V(x) < \lambda\}$. This leads us to estimate the lowest eigenvalue $\lambda_1(L)$ by

$$\lambda_{UP}(L) = \inf_{\mathcal{B}} \max_{(x,\xi)\in\mathcal{B}} (\xi^2 + V(x)) \sim \inf_{x_0,\delta} \left\{\delta^{-2} + \max_{|x-x_0|<\delta} V(x)\right\};$$

$\lambda_{UP}(L)$ is the lowest number $\lambda$ for which $N_{UP}(\lambda, L) \geq 1$.

THEOREM 2. *If $V \geq 0$ is a polynomial, then $c\lambda_{UP}(L) \leq \lambda_1(L) \leq C\lambda_{UP}(L)$. Here $c$ depends only on $n$, $d$, while $C$ depends only on $n$.*

To estimate $N(\lambda, L)$ for $V \geq 0$, we divide $R^n$ into a grid of cubes $\{Q_\nu\}$ of side $\lambda^{-1/2}$, and redefine $N_{UP}(\lambda, L)$ as the number of $Q_\nu$ on which $\text{MAX}_{Q_\nu} V \leq \lambda$. This is consistent with the earlier, less computable, definition of $N_{UP}$. For the higher eigenvalues of $L$ we have

THEOREM 3. *If $V \geq 0$ is a polynomial, then $N_{UP}(c\lambda, L) \leq N(\lambda, L) \leq N_{UP}(C\lambda, L)$. Here $C$ depends only on $n$ and $d$, while $c$ depends only on $n$.*

This yields the order of magnitude of the $N$th eigenvalue uniformly in $N$. There is a significantly sharper form of Theorem 2 for polynomial potentials $V(x)$ which are not assumed to be positive at all points $x \in R^n$.

THEOREM 4. *If $V$ is any polynomial on $R^n$, then*

$$\inf_{x_0,\delta}\left\{c\delta^{-2} + \underset{|x-x_0|<\delta}{\text{MAX}} V(x)\right\} \leq \lambda_1(L) \leq \inf_{x_0,\delta}\left\{C\delta^{-2} + \underset{|x-x_0|<\delta}{\text{MAX}} V(x)\right\}.$$

Here $C$ depends only on $n$, while $c$ depends only on $n$, $d$. We forego the analogous sharpened form of Theorem 3, though it is true also. Later we will give estimates in the spirit of Theorems 2 and 4 with constants independent of the dimension. This is of interest because one wants to pass to the limit and study infinite-dimensional problems.

Now we drop the assumption that $V$ is a polynomial. We take arbitrary $V \leq 0$ on $R^n$ and attempt to estimate the negative eigenvalues of $L = -\Delta + V(x)$. Since $V$ is no longer a polynomial and may be very singular, it is now natural to *average* the symbol $\xi^2 + V(x)$ over a testing box $\mathcal{B} = \{|x - x_0| < \delta, |\xi - \xi_0| < \delta^{-1}\}$, rather than maximize the symbol. (For instance, suppose $V(x) = 0$ in $|x - x_0| < \delta/10^9$, $V(x) = -\delta^{100}$ in $\delta/10^9 \leq |x - x_0| < \delta$.) This leads to a guess

$$\sim \delta^{-2} + \mathrm{Av}_{B(x,\delta)} V = -(\mathrm{Av}_{B(x,\delta)}|V| - \delta^{-2})$$

for $\lambda_1(L)$, rather than $\delta^{-2} + \mathrm{MAX}_{B(x,\delta)} V$. In fact, the following is true.

THEOREM 5. *For $1 < p < \infty$ and constants depending only on $n$, $p$, one has the estimates $cE_{\mathrm{sm}} \leq -\lambda_1(L) \leq CE_{\mathrm{big}}$, where*

$$E_{\mathrm{sm}} = \sup_{x,\delta}[(\mathrm{Av}_{B(x,\delta)}|V|) - C\delta^{-2}],$$

$$E_{\mathrm{big}} = \sup_{x,\delta}[(\mathrm{Av}_{B(x,\delta)}|V|^p)^{1/p} - c\delta^{-2}].$$

COROLLARY. *If $(\mathrm{Av}_{B(x,\delta)}|V|^p)^{1/p} \leq c\delta^{-2}$ for every $x$, $\delta$, then $-\Delta + V \geq 0$.*

To estimate the number of negative eigenvalues, we look for collections of pairwise disjoint testing boxes on which the symbol $|\xi|^2 + V(x)$ has a negative average. This motivates the following result, valid in $R^n$, $n \geq 3$. Again we take $1 < p < \infty$ and use constants $C$, $c$ depending only on $n$ and $p$.

THEOREM 6. (A) *Let $Q_1, Q_2, \ldots, Q_N$ be a collection of cubes whose doubles are disjoint. Suppose $(\mathrm{Av}_{Q_j}|V|) \geq C(\mathrm{diam}\, Q_j)^{-2}$ for each of the cubes. Then $L = -\Delta + V$ has at least $N$ negative eigenvalues.*

(B) *Conversely, suppose $-\Delta + V$ has at least $CN$ negative eigenvalues. Then there is a collection of pairwise disjoint cubes $Q_1, \ldots, Q_N$ for which*

$$(\dagger) \qquad (\mathrm{Av}_{Q_j}|V|^p)^{1/p} \geq c\,(\mathrm{diam}\, Q_j)^{-2}.$$

In practice it is surprisingly easy to find essentially the largest possible collection of $\{Q_j\}$ as in Theorem 6.

Theorems 5 and 6 are somewhat sharp because the upper bounds are of roughly the same form as the lower bounds. It would be interesting to give a sharp limiting form of Theorems 5 and 6 corresponding to $p = 1$; we discuss this point later. We also put off for later the application to physics, and content ourselves here simply with pointing out that Theorem 6 implies Theorem 1. In fact, Theorem 6 evidently gets sharper as $p$ decreases, and we shall use $p = n/2$. To prove Theorem 1 we may evidently take $\lambda = 0$. Also, we may suppose $V \leq 0$, since changing $V$ to $\min\{V, 0\}$ only lowers the eigenvalues of $L$. So we are in the situation of Theorem 6(B), which now produces a collection $Q_1, \ldots, Q_N$ of disjoint cubes satisfying (†), with $CN \geq N(0, L)$. Since $p = n/2$, estimate (†) takes the form

$$\frac{1}{|Q_j|}\int_{Q_j} |V|^{n/2}\, dx \geq [c\,(\mathrm{diam}\, Q_j)^{-2}]^{n/2} = \frac{c'}{|Q_j|},$$

that is, $\int_{Q_j} |V|^{n/2}\,dx \geq c'$. Summing over $j$, we find that

$$\mathrm{Vol}(0,L) = \int_{R^n} |V|^{n/2}\,dx \geq \sum_{j=1}^{N} \int_{Q_j} |V|^{n/2}\,dx \geq c'N \geq \frac{c'}{C} N(0,L),$$

which is the conclusion of Theorem 1. Similarly, the corollary to Theorem 5 (with $p = n/2$) immediately implies the Sobolev inequality.

We now come to the proofs of Theorems 2, 3 and 4 on polynomial potentials.[3] These results all rest on the following elementary estimate.

MAIN LEMMA. *Assume $V(x) \geq 0$ is a polynomial of degree $\leq d$ on a cube $Q$ in $R^n$. Suppose $(\mathrm{Av}_Q V) \geq (\mathrm{diam}\, Q)^{-2}$. Then for functions $u$ in $Q$ we have*

$$\int_Q \{|\nabla u(x)|^2 + V(x)|u(x)|^2\}\,dx \geq c(\mathrm{diam}\, Q)^{-2} \int_Q |u(x)|^2\,dx.$$

*The constant $c$ depends only on $n$ and $d$.*

First we prove the Main Lemma, then explain how to deduce Theorems 2, 3 and 4. The analogue of the Main Lemma for operator-valued potentials will be important in the later discussion of pseudodifferential operators.

PROOF OF THE MAIN LEMMA. We exploit the following simple properties of polynomials $P(x)$ of degree $\leq d$.

(a) $\mathrm{Av}_Q|P| \leq \mathrm{MAX}_Q|P| \leq C\,\mathrm{Av}_Q|P|$.

(b) $\mathrm{MAX}_Q|\nabla P| \leq C(\mathrm{diam}\, Q)^{-1}\,\mathrm{MAX}_Q|P|$.

(c) Suppose $P \geq 0$ on $Q$. Then there is a subcube $Q' \subseteq Q$ with $(\mathrm{diam}\, Q') \geq c(\mathrm{diam}\, Q)$ on which we have $\mathrm{MIN}_{Q'} P \geq \frac{1}{2}\mathrm{MAX}_Q P$.

To check these we may assume $Q =$ unit cube.

Property (a) just asserts the equivalence of two norms on a finite-dimensional vector space, property (b) says that a linear map of finite-dimensional spaces is bounded in norm, while (c) follows from (b) if we pick $Q'$ to include a point of $Q$ where $P$ takes its maximum. Thus, (a), (b), (c) are trivial.

Now let $u$ be a function on $Q$. We start with the trivial estimate

$$\int_Q |\nabla u(x)|^2\,dx \geq \frac{c(\mathrm{diam}\, Q)^{-2}}{|Q|} \int_{Q\times Q} |u(x) - u(y)|^2\,dx\,dy.$$

Also,

$$\int_Q V(x)|u(x)|^2\,dx = \int_Q V(y)|u(y)|^2\,dy = \frac{1}{|Q|}\int_{Q\times Q} V(y)|u(y)|^2\,dx\,dy.$$

[3]Since the results in this lecture haven't appeared before in print, we give the proofs now.

Putting this together yields

$$
\begin{aligned}
&\int_Q \{|\nabla u(x)|^2 + V(x)|u(x)|^2\}\,dx \\
&\quad \geq \frac{1}{|Q|}\int_{Q\times Q} [c(\operatorname{diam} Q)^{-2}|u(x)-u(y)|^2 + V(y)|u(y)|^2]\,dx\,dy \\
&\quad \geq \frac{1}{|Q|}\int_{Q\times Q} (\operatorname{MIN}\{V(y), c(\operatorname{diam} Q)^{-2}\}) \\
&\qquad \cdot [|u(x)-u(y)|^2 + |u(y)|^2]\,dx\,dy \qquad (\dagger) \\
&\quad \geq \frac{1}{|Q|}\int_{Q\times Q} (\operatorname{MIN}\{V(y), c(\operatorname{diam} Q)^{-2}\}) \cdot \left[\frac{1}{2}|u(x)|^2\right] dx\,dy \\
&\quad = \left[\frac{1}{|Q|}\int_Q \frac{1}{2}\operatorname{MIN}\{V(y), c(\operatorname{diam} Q)^{-2}\}\,dy\right] \cdot \int_Q |u(x)|^2\,dx.
\end{aligned}
$$

By property (c) and the hypothesis $(\operatorname{Av}_Q V) \geq (\operatorname{diam} Q)^{-2}$, we have

$$\tfrac{1}{2}\operatorname{MIN}\{V(y), c(\operatorname{diam} Q)^{-2}\} \geq c(\operatorname{diam} Q)^{-2}$$

for a fixed portion of the measure of $Q$. Hence

$$\frac{1}{|Q|}\int_Q \frac{1}{2}\operatorname{MIN}\{V(y), c(\operatorname{diam} Q)^{-2}\}\,dy \geq c'(\operatorname{diam} Q)^{-2},$$

and the Main Lemma follows from (†). Q.E.D.

See Simon [**39**] for another proof of the Main Lemma.

PROOF OF THEOREM 2. The upper bound for $\lambda_1(L)$ is trivial, since

$$\lambda_1(L) = \inf_{\phi\neq 0} \frac{\langle L\phi, \phi\rangle}{||\phi||^2} \leq \frac{\langle L\phi^0, \phi^0\rangle}{||\phi^0||^2}$$

for any fixed $\phi^0 \neq 0$. Letting $\phi^0$ run over all translates and dilates of a fixed smooth function supported in the unit ball, we at once obtain $\lambda_1(L) \leq C\lambda_{UP}(L)$.

The lower bound for $\lambda_1(L)$ amounts to the estimate

$$(\dagger) \qquad \langle Lu, u\rangle = ||\nabla u||^2 + \langle Vu, u\rangle \geq c\lambda_{UP}(L)||u||^2 \quad \text{for } u \in C^\infty(R^n).$$

Cut up $R^n$ into a grid of cubes $Q$, each having side $C_1[\lambda_{UP}(L)]^{-1/2}$. If we take $C_1$ large enough, each $Q$ will contain a ball $B(x_0, \delta)$ with $\delta = 2[\lambda_{UP}(L)]^{-1/2}$, so, by definition of $\lambda_{UP}$, we have

$$\lambda_{UP} \leq \delta^{-2} + \operatorname*{MAX}_{B(x_0,\delta)} V \leq \frac{\lambda_{UP}}{4} + \operatorname*{MAX}_Q V.$$

Thus,

$$\operatorname*{MAX}_Q V \geq \tfrac{3}{4}\lambda_{UP} \geq (\tfrac{3}{4}C_1^2)(\operatorname{diam} Q)^{-2}.$$

Since $\mathrm{Av}_Q V \sim \mathrm{MAX}_Q V$, by taking $C_1$ large we may arrange that

$$(*) \qquad \mathrm{Av}_Q V \geq (\mathrm{diam}\, Q)^{-2},$$

$$(**) \qquad (\mathrm{diam}\, Q)^{-2} = c'\lambda_{UP}(L).$$

Now $(*)$, $(**)$ and the Main Lemma yield

$$\int_Q \{|\nabla u(x)|^2 + V(x)|u(x)|^2\}\, dx \geq c\lambda_{UP}(L) \int_Q |u(x)|^2\, dx$$

for each cube of the grid. Summing over all $Q$ yields the needed estimate $(\dagger)$, and Theorem 2 is proved. Q.E.D.

PROOF OF THEOREM 3. Recall from elementary functional analysis that

(1) $N(\lambda, L) \geq N$ if we can find an $N$-dimensional subspace $H \subseteq L^2$ so that $\langle Lu, u\rangle \leq \lambda ||u||^2$ for $u \in H$.

(2) $N(\lambda, L) \leq N$ if we can find a codimension $N$ subspace $H \subseteq L^2$ so that $\langle Lu, u\rangle \geq \lambda ||u||^2$ for $u \in H$.

We first check that $N(\lambda, L) \geq N = N_{UP}(c\lambda)$. Corresponding to $(c\lambda)$ is a grid $\{Q_\nu\}$ of cubes of sides $(c\lambda)^{-1/2}$. Let $Q_1, \ldots, Q_N$ be those cubes of the grid on which $\mathrm{MAX}_Q V \leq c\lambda$. Translate and dilate a fixed smooth function $\phi$ supported in the unit cube to obtain functions $\phi_1, \ldots, \phi_N$ supported in $Q_1, \ldots, Q_N$. Then just define $H$ as the span of $\phi_1, \ldots, \phi_N$. Evidently $H$ is $N$-dimensional, and for $u = \sum_1^N \alpha_1 \phi_1 \in H$ we have

$$\langle Lu, u\rangle = ||\nabla u||^2 + \langle Vu, u\rangle = \sum_1^N \{|\alpha_j|^2 ||\nabla \phi_j||^2 + |\alpha_j|^2 \langle V\phi_j, \phi_j\rangle\}$$

$$\leq \sum_1^N C(\mathrm{diam}\, Q_j)^{-2} |\alpha_j|^2 ||\phi_j||^2 \leq \lambda ||u||^2.$$

These estimates hold because the $\phi_j$ are translates and dilates of a fixed function, while $\mathrm{MAX}_{Q_j} V \leq c\lambda$. So we have $N(\lambda, L) \geq N_{UP}(c\lambda, L)$ by virtue of (1). Now we show that $N(\lambda, L) \leq N_{UP}(C\lambda, L) = N$. Corresponding to $C\lambda$ is a grid of cubes $\{Q_\nu\}$ of side $(C\lambda)^{-1/2}$, and we let $Q_1, \ldots, Q_N$ be those cubes on which $\mathrm{MAX} V \leq C\lambda$. Define $H$ as the space of all $u \in L^2$ with integral zero over $Q_1, Q_2, \ldots, Q_N$. $H$ has codimension $N$, and we shall prove that

$$(*) \qquad \langle Lu, u\rangle = ||\nabla u||^2 + \langle Vu, u\rangle \geq \lambda ||u||^2 \quad \text{for } u \in H.$$

This will follow at once by summing the following estimates over $\nu$:

$$(**) \qquad \int_{Q_\nu} \{|\nabla u|^2 + V|u|^2\}\, dx \geq \lambda \int_{Q_\nu} |u|^2\, dx \quad \text{for } u \in H.$$

For $\nu \neq 1, 2, \ldots, N$ we have $\mathrm{MAX}_{Q_\nu} V \geq C\lambda$, so $(**)$ holds for any $u$ by virtue of the Main Lemma.

On the other hand, for $\nu = 1, 2, \ldots, N$ we argue as follows. Any $u \in L^2(Q)$ satisfies

$$\int_{Q_\nu} |\nabla u(x)|^2\, dx \geq c(\mathrm{diam}\, Q_\nu)^{-2} \int_{Q_\nu} |u(x) - \mathrm{av}_{Q_\nu} u|^2\, dx$$

$$\geq \lambda \int_{Q_\nu} |u(x) - \mathrm{av}_{Q_\nu} u|^2\, dx.$$

If $u \in H$, then $\mathrm{av}_{Q_\nu} u = 0$ and $(**)$ follows at once. Thus, $(**)$ holds for all $\nu$, and now $N(\lambda, L) \leq N_{UP}(\lambda, L)$ by virtue of (2). The proof of Theorem 3 is complete. Q.E.D.

PROOF OF THEOREM 4. The upper bound for $\lambda_1(L)$ is trivial as in Theorem 2. The lower bound for $\lambda_1(L)$ amounts to the following estimate.

Suppose $V$ is a polynomial of degree $\leq d$ on $R^n$ satisfying

$$(\dagger) \qquad \operatorname*{MAX}_{Q} V \geq -c_1(\operatorname{diam} Q)^{-2}$$

for every cube $Q$. Then $\|\nabla u\|^2 + \langle Vu, u\rangle \geq 0$.

To prove this we can suppose that $u$ is supported in a very large cube $Q^0$. Except in the trivial case $V =$ constant, we have

$$\left[\operatorname*{MAX}_{Q^0} V - \operatorname*{MIN}_{Q^0} V\right] \geq C(\operatorname{diam} Q^0)^{-2}$$

if we start out with $Q^0$ large enough.

Now make a Calderón-Zygmund decomposition of $Q^0$ by bisecting $Q^0$ into $2^n$ equal subcubes, bisecting each of these subcubes, etc. We stop cutting whenever we arrive at a cube $Q$ satisfying

$$(\dagger\dagger) \qquad \left[\operatorname*{MAX}_{Q} V - \operatorname*{MIN}_{Q} V\right] \leq C_1(\operatorname{diam} Q)^{-2}.$$

This will eventually happen, since each time we bisect $Q$ the left side of $(\dagger\dagger)$ shrinks, while the right side grows by a factor of 4. Consequently, the big cube $Q^0$ is partitioned into subcubes $\{Q_\nu\}$ each of which satisfies

$$(*) \qquad cC_1(\operatorname{diam} Q_\nu)^{-2} \leq \left[\operatorname*{MAX}_{Q_\nu} V - \operatorname*{MIN}_{Q_\nu} V\right] \leq C_1(\operatorname{diam} Q_\nu)^{-2}.$$

Here $c$ depends only on $n$ and $d$, and the first estimate of $(*)$ holds because $Q$ arose by bisecting a cube for which $(\dagger\dagger)$ fails.

If we take $C_1$ large enough depending on $n$ and $d$, then the Main Lemma applied to $\tilde{V}(x) = V(x) - \mathrm{MIN}_{Q_\nu} V$ shows that

$$(**) \qquad \int_{Q_\nu} \{|\nabla u|^2 + V|u|^2\}\, dx \geq \left(\operatorname*{MIN}_{Q_\nu} V + \bar{c}(\operatorname{diam} Q_\nu)^{-2}\right) \int_{Q_\nu} |u|^2\, dx.$$

However, if $(\dagger)$ holds with $c_1$ small enough, then we have

$$\operatorname*{MIN}_{Q_\nu} V \geq -\bar{c}(\operatorname{diam} Q_\nu)^{-2}.$$

To see this, suppose $\mathrm{MIN}_{Q_\nu} V$ occurs at $x^0 \in Q_\nu$, and let $Q$ be a subcube of $Q_\nu$ with $\operatorname{diam} Q = \beta(\operatorname{diam} Q_\nu)$ and $x^0 \in Q$. By observation (b) in the proof of the Main Lemma, we have

$$\operatorname*{MAX}_{Q} V \leq \operatorname*{MIN}_{Q_\nu} V + C\beta\left[\operatorname*{MAX}_{Q_\nu} V - \operatorname*{MIN}_{Q_\nu} V\right].$$

Using $(\dagger)$ on the left and $(*)$ on the right, we get

$$-c_1(\operatorname{diam} Q)^{-2} \leq \operatorname*{MIN}_{Q_\nu} V + C\beta C_1(\operatorname{diam} Q_\nu)^{-2},$$

i.e.

$$-c_1\beta^{-2}(\operatorname{diam} Q_\nu)^{-2} \le \underset{Q_\nu}{\operatorname{MIN}}\, V + CC_1\beta(\operatorname{diam} Q_\nu)^{-2}.$$

Picking $\beta \ll 1/CC_1$ and then picking $c_1 \ll \beta^2$, we see that $\operatorname{MIN}_{Q_\nu} V \ge -\bar{c}(\operatorname{diam} Q_\nu)^{-2}$ as claimed. Now (**) shows that $\int_{Q_\nu}\{|\nabla u|^2 + V|u|^2\}\,dx \ge 0$. Summing over $\nu$ yields $\|\nabla u\|^2 + \langle Vu, u\rangle \ge 0$, which is the desired estimate. Q.E.D.

Before leaving polynomial potentials, we should mention a version of Theorem 4 with constants independent of the dimension. To motivate the result, we look at the $\phi_2^4$-field from quantum field theory. Here

$$L = -\sum_{\nu=1}^{N} \frac{\partial^2}{\partial x_\nu^2} + \sum_{\nu=1}^{N} [-\alpha x_\nu^2 + \beta x_\nu^4 + \gamma + \delta(x_\nu - x_{\nu+1})^2]$$

and $\alpha, \beta, \gamma, \delta$ depend on $N$. $L$ describes the quantum-mechanical analogue of a classical mechanical system of $N$ particles with positions $x_1, \dots, x_N$ and moving in a potential well

$$V(x_1, \dots, x_N) = \sum_{\nu=1}^{N} [-\alpha x_\nu^2 + \beta_\nu^4 + \gamma + \delta(x_\nu - x_{\nu+1})^2],$$

If we could pass to the limit as $N \to \infty$, then the classical system goes over to a field $\phi(x)$ with "potential" $V(\phi) = \int_{-\infty}^{\infty}[\mu\phi^2 + \beta\phi^4 + |\nabla\phi|^2]\,dx$, while $L$ goes over to the corresponding quantized field. So estimates independent of the dimension are aimed at the passage from ordinary quantum mechanics to quantum field theory. We emphasize that so far our estimates are much too crude to deal with this problem. Nevertheless, we can pick out from the examples of quantum field theory two basic properties of the potential $V(x_1, \dots, x_N)$:

(A) $V(x_1, \dots, x_N)$ is a polynomial of degree $\le d$, $d$ independent of $N$.

(B) In $V(x_1, \dots, x_N)$ we find that $x_\nu$ is coupled directly only to $x_{\nu-1}$ and $x_{\nu+1}$. That is, $\partial^2 V/\partial x_\mu \partial x_\nu = 0$ unless $|\mu - \nu| \le 1$.

More generally, a potential $V(x_1, \dots, x_N)$ will be called type $(d, s)$ if

(A′) $V$ is a polynomial of degree at most $d$.

(B′) For each $\mu$ we have $\partial^2 V/\partial x_\mu \partial x_\nu = 0$ except for at most $s$ values of $\nu$.

We shall estimate the lowest eigenvalue of $L = -\Delta + V(x)$ in terms of the growth of $V$ on boxes $I = I_1 \times I_2 \times \cdots \times I_N$, where the lengths of the intervals $I_1, I_2, \dots, I_N$ need not be equal. Our result is as follows.

THEOREM 7. *Suppose $V(x_1, \dots, x_N)$ is of type $(d, s)$. Then the lowest eigenvalue $\lambda_1(L)$ may be estimated by*

$$\inf_{I_1\times\cdots\times I_N}\left\{c\sum_j |I_j|^{-2} + \max_{I_1\times\cdots\times I_N} V\right\} \le \lambda_1(L)$$

$$\le \inf_{I_1\times\cdots\times I_N}\left\{C\sum_j |I_j|^{-2} + \max_{I_1\times\cdots\times I_N} V\right\}.$$

*Here $C$ is a universal constant, while $c$ depends on $d$ and $s$ but not on $N$.*

For $V \geq 0$ we obtain the order of magnitude of $\lambda_1(L)$.

The proof of Theorem 7 is a delicate refinement of the proof of Theorem 4. Again we start with a huge cube $Q^0$ and make a Calderón-Zygmund decomposition. This time, however, instead of bisecting a given box into $2^N$ congruent smaller boxes, we cut only one side at a time. Thus, at each stage of the Calderón-Zygmund cutting, a box is cut into only two pieces but the shape changes. A critical part of the proof is to decide at each stage which side to cut. We won't give the details here. Instead, we move on to nonpolynomial potentials.

To prepare for the proofs of Theorems 5 and 6, we introduce some notation and background from Euclidean Fourier analysis. For $Q$ a dyadic cube, define

$$H_+^Q = \text{space of functions supported in } Q, \text{ and linear + constant on each of the dyadic subcubes obtained by bisecting } Q.$$

$$H_0^Q = \text{space of functions supported in } Q, \text{ and linear + constant on all of } Q.$$

$$H^Q = \text{the orthogonal complement of } H_0^Q \text{ in } H_+^Q.$$

An $L^2$-function $u$ can be expanded in a series $u = \sum_Q \hat{u}(Q)$ with each $\hat{u}(Q) \in H^Q$. This is a slight variant of a Haar series, and the corresponding variant of the dyadic square function is

$$S(u) = \left[\sum_Q \frac{\|\hat{u}(Q)\|^2}{|Q|}\chi_Q\right]^{1/2}.$$

Such functions play an important role in the analysis of singular integral operators. It is well known to Fourier analysts that the size of $S(u)$ controls very closely the size of $u$. In particular, we will need the following "weighted-norm" inequality.

LEMMA A. *Let $u$ be a function supported on a cube $Q^0$ and orthogonal to $H_0^{Q^0}$. Then $\int_{Q^0} |u|^2|V|\,dx \leq C\int_{Q^0} S^2(u)V^+\,dx$, where*

$$V^+(x) = \sup_{Q \ni x} (\mathrm{Av}_Q |V|^p)^{1/p}, \qquad p > 1.$$

SKETCH OF PROOF. One checks easily that $V^+$ satisfies Muckenhoupt's $(A_\infty)$ condition. Therefore the standard theory of weighted-norm inequalities tells us for $u \perp H_0^{Q^0}$ that $\int_{Q^0} |u^*|^2 V^+\,dx \leq C\int_{Q^0} S^2(u)V^+\,dx$, where $u^*$ is the dyadic maximal function of $u$. This is much stronger than the estimate of Lemma A, since we can just write

$$\int_{Q^0} |u|^2|V|\,dx \leq \int_{Q^0} |u^*|^2 V^+\,dx \leq C\int_{Q^0} S^2(u)V^+\,dx. \quad \text{Q.E.D.}$$

Next we relate the expansion $u = \sum_Q \hat{u}(Q)$ to the Laplacian by introducing $|||u|||^2 = \sum_Q (\operatorname{diam} Q)^{-2}\|\hat{u}(Q)\|^2$, and proving

LEMMA B. $|||u|||^2 \leq C\|\nabla u\|^2$ *for* $u \in C_0^\infty(R^n)$.

PROOF. Write $u = \sum_l u_l$ with $\hat{u}_l(\xi)$ supported in $|\xi| \sim 2^l$. Thus $u_l = \chi_l * u_l$, where

$$\hat{\chi}_l(\xi) = \begin{cases} 1 & \text{for } |\xi| \sim 2^l, \\ 0 & \text{for } |\xi| \le 2^{l-2} \text{ and for } |\xi| \ge 2^{l+2}, \\ \text{smooth in between.} \end{cases}$$

We can take $\chi_l$ to satisfy

$$|\partial^\alpha \chi_l(z)| \le C_{\alpha s} 2^{(n+|\alpha|)l} (2^{-l}/(|z| + 2^{-l}))^s, \tag{1}$$

since this amounts to saying that $2^{-nl}\chi_l(2^{-l}z)$ is a Schwartz function. Now $||\hat{u}(Q)|| \le \sum_l ||\hat{u}_l(Q)||$. We study first the case $\operatorname{diam} Q \le 2^{-l}$; say $\operatorname{diam} Q = 2^{-l-k}$, $k \ge 0$. For suitable $\psi = \psi_l^Q \in H^Q$ of unit $L^2$-norm, we have

$$\begin{aligned} ||\hat{u}_l(Q)|| &= |\langle \chi_l * u_l, \psi \rangle| = \left| \int \chi_l(x-y) u_l(y) \psi(x)\, dx\, dy \right| \\ &= \left| \int \left[ \chi_l(x-y) - \chi_l(x_Q - y) \right.\right. \\ &\qquad \left.\left. - \sum_\mu (x_\mu - x_{Q_\mu}) \cdot \partial_\mu \chi_l(x_Q - y) \right] u_l(y) \psi(x)\, dx\, dy \right|. \end{aligned} \tag{2}$$

Here $x_Q$ with coordinates $(x_{Q_\mu})$ is the center of $Q$, and the extra terms in brackets don't affect the integral because $\psi \in H^Q$ and, hence, $\psi$ annihilates constants and linear functions.

Estimates (1) for the second derivatives of $\chi_l$ show that the term in brackets is bounded by $C_s(\operatorname{diam} Q)^2 2^{(n+2)l}(2^{-l}/(|x_Q - y| + 2^{-l}))^s$, while $\int_Q |\psi(x)|\, dx \le |Q|^{1/2}$ since $\psi$ is supported in $Q$ and has unit $L^2$-norm. So (2) implies

$$\begin{aligned} ||\hat{u}_l(Q)|| &\le C|Q|^{1/2} (\operatorname{diam} Q)^2 2^{(n+2)l} \int_{R^n} \left( \frac{2^{-l}}{|x_Q - y| + 2^{-l}} \right)^s |u_l(y)|\, dy \\ &= C|Q|^{1/2} \cdot 2^{-2k} (\phi_l * |u_l|)(x_Q), \quad \text{where } \phi_l(z) = 2^{nl} \left( \frac{2^{-l}}{|z| + 2^{-l}} \right)^s. \end{aligned}$$

Since $\phi_l(x_Q - y)$ and $\phi_l(x - y)$ are of the same order of magnitude for $x \in Q$, $y \in R^n$, it follows that

$$||\hat{u}_l(Q)|| \le C'|Q|^{1/2} \cdot 2^{-2k} (\phi_l * |u_l|)(x) \quad \text{for any } x \in Q,$$

so that

$$||\hat{u}_l(Q)||^2 \le C 2^{-4k} \int_Q (\phi_l * |u_l|)^2\, dx.$$

Summing over all $Q$ of diameter $2^{-l-k}$ yields

$$\sum_{\operatorname{diam} Q = 2^{-l-k}} ||\hat{u}_l(Q)||^2 \le C 2^{-4k} \int_{R^n} (\phi_l * |u_l|)^2\, dx \le C' 2^{-4k} ||u_l||^2,$$

since $\phi_l$ has bounded $L^1$-norm. Now multiply by $2^{2l+2k}$ and sum over all $k \ge 0$ and all $l$. The result is

$$\sum_Q \sum_{2^{-l} \ge \operatorname{diam} Q} \frac{||\hat{u}_l(Q)||^2 (\operatorname{diam} Q)^{-2}}{2^l \operatorname{diam} Q} \le C \sum_l 2^{2l} ||u_l||^2 \le C' ||\nabla u||^2.$$

Since for a fixed $Q$ we have

$$\left(\sum_{2^{-l}\geq \operatorname{diam} Q} \|\hat{u}_l(Q)\|\right)^2 \leq C \sum_{2^{-l}\geq \operatorname{diam} Q} \frac{\|\hat{u}_l(Q)\|^2}{2^l \operatorname{diam} Q},$$

it follows that

$$\sum_Q \left(\sum_{2^{-l}\geq \operatorname{diam} Q} \|\hat{u}_l(Q)\|\right)^2 (\operatorname{diam} Q)^{-2} \leq C\|\nabla u\|^2. \tag{3}$$

On the other hand, suppose $2^{-l} < \operatorname{diam} Q$, say $\operatorname{diam} Q = 2^{-l+k}$ with $k \geq 1$. For fixed $l$ and $k$ we know that $\sum_{\operatorname{diam} Q = 2^{k-l}} \|\hat{u}_l(Q)\|^2 \leq \|u_l\|^2$. Multiplying by $2^{2l-k}$ and summing over all $k \geq 1$ and all $l$, we get

$$\sum_Q \sum_{2^{-l}<\operatorname{diam} Q} \|\hat{u}_l(Q)\|^2 (\operatorname{diam} Q)^{-2} \cdot \{2^l \operatorname{diam} Q\} \leq C \sum_l 2^{2l}\|u_l\|^2 \leq C'\|\nabla u\|^2.$$

Since for fixed $Q$ we have

$$\left(\sum_{2^{-l}<\operatorname{diam} Q} \|\hat{u}_l(Q)\|\right)^2 \leq C \sum_{2^{-l}<\operatorname{diam} Q} \|\hat{u}_l(Q)\|^2 \cdot \{2^l \operatorname{diam} Q\},$$

it follows that

$$\sum_Q \left(\sum_{2^{-l}<\operatorname{diam} Q} \|\hat{u}_l(Q)\|\right)^2 (\operatorname{diam} Q)^{-2} \leq C\|\nabla u\|^2.$$

This and (3) yield

$$\sum_Q \left(\sum_l \|\hat{u}_l(Q)\|\right)^2 (\operatorname{diam} Q)^{-2} \leq C\|\nabla u\|^2.$$

Since we already noted that $\|\hat{u}(Q)\| \leq \sum_l \|\hat{u}_l(Q)\|$, Lemma B is proved. Q.E.D.

Combining Lemmas A and B it is now easy to prove the corollary to Theorem 5. In fact one has

LEMMA C. *Assume* $(\operatorname{Av}_Q V^p)^{1/p} \leq \gamma(\operatorname{diam} Q)^{-2}$ *for all* $Q \subseteq Q^0$. *Then for any* $u$ *orthogonal to* $H_0^{Q^0}$ *we have*

$$\int_{Q^0} |u|^2 |V|\, dx \leq C\gamma |||u|||^2.$$

Lemmas B and C show that

$$\langle(-\Delta + V)u, u\rangle = \|\nabla u\|^2 - \int |u|^2|V|\, dx \geq c|||u|||^2 - \int |u|^2|V|\, dx \geq 0$$

for $u \perp H_0^{Q^0}$, if we pick $\gamma$ small. The set of $u \in C_0$ orthogonal to $H_0^{Q^0}$ for some large $Q^0$ is dense in $L^2$, so $-\Delta + V \geq 0$, which is the corollary to Theorem 5.

PROOF OF LEMMA C. First of all we claim that $\operatorname{Av}_Q V^+ \leq C\gamma(\operatorname{diam} Q)^{-2}$ for every cube $Q$. In fact, define

$$c(Q) = \max_{Q' \supset Q} (\operatorname{Av}_{Q'} |V|^p)^{1/p}, \qquad V_Q^+(x) = \sup_{\substack{Q' \ni x \\ Q' \subseteq Q}} (\operatorname{Av}_{Q'} |V|^p)^{1/p}.$$

The definitions at once yield $V^+(x) = \max[c(Q), V_Q^+(x)] \leq c(Q) + V_Q^+(x)$ for $x \in Q$.

The hypothesis of Lemma C gives $c(Q) \leq \gamma(\operatorname{diam} Q)^{-2}$, while the maximal theorem and hypothesis imply

$$(\operatorname{Av}_Q V_Q^+) \leq C_p(\operatorname{Av}_Q |V|^p)^{1/p} \leq C_p \gamma(\operatorname{diam} Q)^{-2};$$

these inequalities prove the claim.

Now for $u \perp H_0^{Q^0}$ we invoke Lemma A to write

$$\int_{Q^0} |u|^2 |V|\, dx \leq C \int_{Q^0} S^2(u) V^+\, dx = C \int_{Q^0} \left[ \sum_Q \frac{\|\hat{u}(Q)\|^2}{|Q|} \chi_Q \right] V^+\, dx$$

$$= C \sum_Q \|\hat{u}(Q)\|^2 (\operatorname{Av}_Q V^+) \leq C' \gamma \sum_Q \|\hat{u}(Q)\|^2 (\operatorname{diam} Q)^{-2}$$

by the claim. That is,

$$\int_{Q^0} |u|^2 |V|\, dx \leq C' \gamma |||u|||^2. \quad \text{Q.E.D.}$$

Proof of Theorem 6. Part (A) is trivial, in the spirit of Theorem 3. To prove part (B) we shall suppose $V$ is bounded, but make sure our estimates don't depend on the bound.

We can immediately define the cubes $Q_1, \ldots, Q_N$ of part (B). They are simply the minimal dyadic cubes which satisfy $(\operatorname{Av}_{Q_j} |V|^p)^{1/p} \leq \gamma(\operatorname{diam} Q_j)^{-2}$. Evidently the $Q_j$ are pairwise disjoint and satisfy the estimate of part (B). The problem is to prove that $-\Delta + V$ has at most $CN$ negative eigenvalues. This we do as usual by constructing a space $H \subseteq L^2(R^n)$ of codimension $\leq CN$ so that

$$(\dagger) \qquad \|\nabla u\|^2 + \langle Vu, u \rangle \geq 0 \quad \text{for } u \in H.$$

To build the subspace $H$, we first add to the $Q_1, \ldots, Q_N$ certain other cubes $Q_{N+1}, \ldots, Q_M$, taking care that $M \leq CN$. We then define $H$ to consist of all functions $u$ orthogonal to $H_+^{Q_j}$ for $j = 1, 2, \ldots, M$. Evidently $H$ has codimension at most $CN$. It remains to prove the estimate (†).

Now define sets $E(Q_j) = Q_j \setminus \bigcup_{j'} Q_{j'}$, where $Q_{j'}$ varies over all cubes among $Q_1, \ldots, Q_M$ which are properly contained in $Q_j$. Also set $E(R^n) = R^n \setminus \bigcup_1^M Q_j$. Thus, $E(R^n)$, $E(Q_1), \ldots, E(Q_M)$ partitions $R^n$ into disjoint subsets. We shall prove that

$$\left( \frac{1}{|Q|} \int_{Q \cap E(Q_j)} |V|^p\, dx \right)^{1/p} \leq C\gamma(\operatorname{diam} Q)^{-2}$$

for all $Q$ and each $j$. Also

$$\left( \frac{1}{|Q|} \int_{Q \cap E(R^n)} |V|^p\, dx \right)^{1/p} \leq C\gamma(\operatorname{diam} Q)^{-2}$$

for all $Q$. Therefore, Lemma C yields the estimates

$$\int_{E(Q_j)} |u|^2 |V|\, dx \le C\gamma |||u|||^2 \quad \text{for } u \perp H_+^{Q_j},$$

(††)

$$\int_{E(R^n)} |u|^2 |V|\, dx \le C\gamma |||u|||^2 \quad \text{for all } u \in C_0^\infty.$$

We are in position to prove (†).

Given a function $u = \sum_Q \hat{u}(Q)$, define

$$u_j = \sum_{Q \in \mathfrak{U}_j} \hat{u}(Q), \quad \text{where } Q \in \mathfrak{U}_j \text{ if } Q \subseteq Q_j, \text{ but } Q \not\subseteq Q_{j'}$$

for any $Q_j$, properly contained in $Q_j$.

$$u_0 = \sum_{Q \in \mathfrak{U}_0} \hat{u}(Q), \quad \text{where } Q \in \mathfrak{U}_0 \text{ if, for all } j,\ Q \not\subseteq Q_j.$$

Since each dyadic cube appears in exactly one of the $\mathfrak{U}_j$ $(j = 0, 1, \dots, M)$, we have $|||u|||^2 = \sum_{j=0}^M |||u_j|||^2$ for any $u \in C_0$. Also, for $u \in H$ we note that $u(x) = \sum_{Q \subseteq Q_j} \hat{u}(Q)$ for $x \in Q_j$ (because $u \perp H_+^{Q_j}$). Therefore $u(x) = u_j(x)$ for $x \in E(Q_j)$. Similarly $u(x) = u_0(x)$ for $x \in E(R^n)$. Finally, we apply (††) to the $u_j$ to conclude that

$$\begin{aligned}\int_{R^n} |u|^2 |V|\, dx &= \int_{E(R^n)} |u_0|^2 |V|\, dx + \sum_{j=1}^M \int_{E(Q_j)} |u_j|^2 |V|\, dx \\ &\le C\gamma |||u_0|||^2 + C\gamma \sum_{j=1}^M |||u_j|||^2 = C\gamma |||u|||^2 \quad \text{for } u \in H.\end{aligned}$$

Taking $\gamma$ small enough and applying Lemma B, we get

$$\int_{R^n} |u|^2 |V|\, dx \le \|\nabla u\|^2 \quad \text{for } u \in H,$$

which is (†).

To summarize, the key estimate (†) will follow if we can carry out three steps:

(a) Define the additional cubes $Q_{N+1}, \dots, Q_M$.

(b) Check that $M \le CN$.

(c) Prove the estimates

(i) $$\left( \frac{1}{|Q|} \int_{Q \cap E(Q_j)} |V|^p\, dx \right)^{1/p} \le C\gamma (\operatorname{diam} Q)^{-2}$$

and

(ii) $$\left( \frac{1}{|Q|} \int_{Q \cap E(R^n)} |V|^p\, dx \right)^{1/p} \le C\gamma (\operatorname{diam} Q)^{-2}.$$

First we define the extra cubes $Q_{N+1}, \dots, Q_M$. Let $\mathcal{B}$ be the collection of all $Q$ satisfying $(\mathrm{Av}_Q |V|^p)^{1/p} \ge \gamma (\operatorname{diam} Q)^{-2}$. For $Q \in \mathcal{B}$ define $D(Q)$, the

"descendents of $Q$", to consist of the maximal $Q' \in \mathcal{B}$ which are properly contained in $Q$. We shall say that $Q$ *branches* if there are at least two cubes in $D(Q)$. Now define

$\mathcal{Q}^0 =$ family of minimal $Q$,

$\mathcal{Q}^1 =$ family of maximal $Q$,

$\mathcal{Q}^2 =$ family of branching $Q$,

$\mathcal{Q}^3 = \bigcup_{Q \in \mathcal{Q}^2} D(Q)$, the family of descendents of branching cubes.

In particular, $\mathcal{Q}^0$ consists of the cubes $Q_1, \ldots, Q_N$. The family of extra cubes is defined as $\mathcal{Q}^1 \cup \mathcal{Q}^2 \cup \mathcal{Q}^3$. Step (a) is now complete.

Next we carry out step (b). That is, we check that the number of cubes in $\mathcal{Q}^1 \cup \mathcal{Q}^2 \cup \mathcal{Q}^3$ is at most a constant multiple of the number of cubes in $\mathcal{Q}^0$. It is immediately clear that $|\mathcal{Q}^1| \leq |\mathcal{Q}^0|$. In fact, we just associate to each maximal $Q \in \mathcal{B}$ some minimal $Q' \in \mathcal{B}$ contained in $Q$; note that distinct maximal $Q$ are pairwise disjoint and thus give rise to distinct $Q'$. To show that $|\mathcal{Q}^2 \cup \mathcal{Q}^3| \leq C|\mathcal{Q}^0|$ is an exercise in graph theory. The collection $\mathcal{B}$ of "bad" dyadic cubes has the structure of several trees under inclusion as in Figure 1.

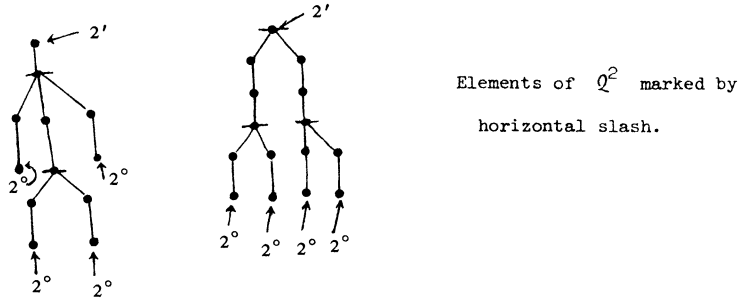



FIGURE 1

For any tree graph we may define $\mathcal{Q}^0 =$ minimal elements, $\mathcal{Q}^2 =$ branching elements, and we have

LEMMA D.

$$(\dagger) \qquad \sum_{Q \in \mathcal{Q}^2} \{1 + |D(Q)|\} \leq 5|\mathcal{Q}^0|.$$

This means $|\mathcal{Q}^2 \cup \mathcal{Q}^3| \leq 5|\mathcal{Q}^0|$, so step (b) is reduced to Lemma D.

PROOF OF LEMMA D. Grow the tree from the top down. To start with, the tree consists of a single point, so $\mathcal{Q}^2$ is empty and $\mathcal{Q}^0$ consists of one point. The tree grows by repeatedly adding the descendents $D(\overline{Q})$ for a point $\overline{Q}$ which is minimal in the part of the tree grown so far. See Figure 2.

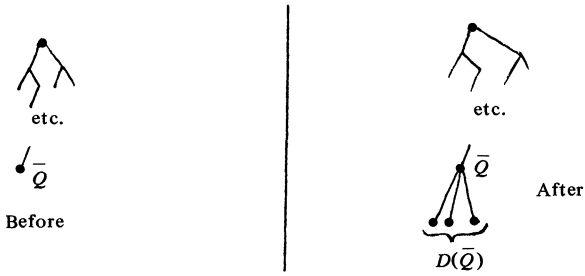



FIGURE 2. How the tree grows

If $|D(\overline{Q})| = 1$, then $\mathcal{Q}^2$ remains unchanged, while to $\mathcal{Q}^0$ we add $D(\overline{Q})$ and remove $\overline{Q}$. Thus $\sum_{Q\in\mathcal{Q}^2}\{1+|D(Q)|\} - 5|\mathcal{Q}^0|$ remains unchanged. On the other hand, if $|D(\overline{Q})| > 1$, then $\overline{Q}$ is added to $\mathcal{Q}^2$, while to $\mathcal{Q}^0$ we add $D(\overline{Q})$ and remove $\overline{Q}$. Thus, $\sum_{Q\in\mathcal{Q}^2}\{1+|D(Q)|\}$ increases by $1+|D(\overline{Q})|$, while $|\mathcal{Q}^0|$ is increased by $|D(\overline{Q})|-1$. The total change in $\sum_{Q\in\mathcal{Q}^2}\{1+|D(Q)|\} - 5|\mathcal{Q}^0|$ is $[1+|D(\overline{Q})|]-5[|D(\overline{Q})|-1] = 6-4|D(\overline{Q})| < 0$, since $|D(\overline{Q})| \geq 2$. So in either case, $\sum_{Q\in\mathcal{Q}^2}\{1+|D(Q)|\} - 5|\mathcal{Q}^0|$ remains constant or decreases as the tree grows. Lemma D follows, and step (b) is complete. Q.E.D.

Now we come to step (c), the proof of estimates (i) and (ii). We start with (ii). Let $Q$ be any dyadic cube. If $Q \notin \mathcal{B}$, then already

$$\left(\frac{1}{|Q|}\int_Q |V|^p\,dx\right)^{1/p} \leq \gamma(\operatorname{diam} Q)^{-2},$$

so (ii) is obvious. On the other hand, if $Q \in \mathcal{B}$, then $Q \subseteq Q'$ for a maximal $Q' \in \mathcal{B}$. Since $Q' \in \mathcal{Q}^1$ is among the extra cubes $Q_{N+1},\ldots,Q_M$, it follows that $Q' \cap E(R^n) = \emptyset$, so $Q \cap E(R^n) = \emptyset$ also, and again (ii) is obvious.

Next we prove (i). Observe that it is enough to prove (i) when $Q$ is a proper subcube of $Q_j$. Indeed, the case $Q = Q_j$ will then follow (with a different constant) from applying (i) to the $2^n$ subcubes obtained from bisecting $Q_j$. Since (i) may be rewritten

$$\int_{Q\cap E(Q_j)} |V|^p\,dx \leq [C\gamma]^p c(\operatorname{diam} Q)^{n-2p},$$

the case $Q \supseteq Q_j$ follows from the case $Q = Q_j$ as long as $p \leq n/2$. (Recall $E(Q_j) \subseteq Q_j$.) Theorem 6 only gets stronger as $p$ decreases, and we have $n \geq 3$; hence we may assume $p \leq n/2$. So if (i) holds for $Q$ proper subcubes of $Q_j$, then we have checked that (i) follows also for $Q \supseteq Q_j$. The only remaining dyadic cubes $Q$ are disjoint from $Q_j$, hence also from $E(Q_j)$, and so (i) holds vacuously.

So our job is to prove (i) for $Q$ a proper subcube of $Q_j$. We may also assume $Q \in \mathcal{B}$, since otherwise (i) is automatic. We consider separately the three cases $Q_j \in \mathcal{Q}^0$, $Q_j \in \mathcal{Q}^2$, $Q_j \notin \mathcal{Q}^0 \cup \mathcal{Q}^2$.

If $Q_j \in \mathcal{Q}^0$, then we cannot have $Q \in \mathcal{B}$ properly contained in $Q_j$, so there is nothing to prove.

If $Q_j \in \mathcal{Q}^2$, then $D(Q_j) \subseteq \mathcal{Q}^3$ so $E(Q_j) \subseteq Q_j \setminus \bigcup_{Q' \in D(Q_j)} Q'$. For $Q \in \mathcal{B}$ properly contained in $Q_j$, we must have $Q \subseteq Q'$ for some $Q' \in D(Q_j)$. Thus $Q \cap E(Q_j) = \emptyset$ and estimate (i) is trivial.

Finally, suppose $Q_j \notin \mathcal{Q}^0 \cup \mathcal{Q}^2$. Then some cube in $\mathcal{Q}^0$ is properly contained in $Q_j$. It follows that we can find a maximal element $Q^\#$ among the cubes of $\mathcal{Q}^0 \cup \mathcal{Q}^2$ properly contained in $Q_j$. Since $Q^\# \in \mathcal{Q}^0 \cup \mathcal{Q}^2$, we have $E(Q_j) \subseteq Q \setminus Q^\#$. Now let $Q \in \mathcal{B}$ be properly contained in $Q_j$. Then $Q$ must intersect $Q^\#$. (Otherwise we obtain a contradiction as follows. We have $Q$, $Q^\# \subseteq Q_j$, so we can find a minimal element $\tilde{Q}$ among the cubes of $\mathcal{B}$ which contain both $Q$ and $Q^\#$. Since $Q$, $Q^\# \subseteq \tilde{Q}$, we know that $\tilde{Q} \notin \mathcal{Q}^0$, and therefore $D(\tilde{Q}) \neq \emptyset$. If $D(\tilde{Q})$ consisted of a single element $\tilde{\tilde{Q}}$, then both $Q$, $Q^\# \subseteq \tilde{\tilde{Q}}$, contradicting the minimal property of $\tilde{Q}$. Hence $D(\tilde{Q})$ consists of at least two cubes, i.e. $\tilde{Q} \in \mathcal{Q}^2$. Since $Q_j \notin \mathcal{Q}^2$, we must have $\tilde{Q}$ properly contained in $Q_j$. Now since $Q$, $Q^\# \subseteq \tilde{Q}$ with $\tilde{Q} \in \mathcal{Q}^2$ properly contained in $Q_j$, we get a contradiction to the maximal property of $Q^\#$.)

Since $Q$ intersects $Q^\#$, we must have $Q \subseteq Q^\#$ or $Q^\# \subseteq Q$. If $Q \subseteq Q^\#$, then $Q \cap E(Q_j) = \emptyset$ and (i) is trivial. So we may assume $Q^\# \subset Q$. The next step is to make a Calderón-Zygmund decomposition of $Q$. We bisect $Q$ repeatedly, stopping at $\tilde{Q}_\alpha$ if either $\tilde{Q}_\alpha = Q^\#$ or $\tilde{Q}_\alpha \cap Q^\# = \emptyset$. Thus $Q$ is cut up as a union of the $\tilde{Q}_\alpha$. One of the $\tilde{Q}_\alpha$ will be $Q^\#$; the other $\tilde{Q}_\alpha$ are disjoint from $Q^\#$ and therefore cannot belong to $\mathcal{B}$. The situation is as in Figure 3. In particular, there are only a bounded number of $\tilde{Q}_\alpha$ of a given diameter ($2^n - 1$ in fact).

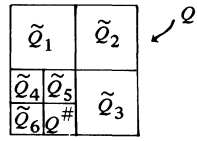



FIGURE 3. Decomposition of $Q$ into $\tilde{Q}_\alpha$

Since the $\tilde{Q}_\alpha \neq Q^\#$ are not in $\mathcal{B}$, we have

$$\int_{Q \cap E(Q_j)} |V|^p \, dx \leq \int_{Q \cap [Q_j \setminus Q^\#]} |V|^p \, dx = \sum_{\tilde{Q}_\alpha \neq Q^\#} \int_{\tilde{Q}_\alpha} |V|^p \, dx$$

$$\leq \sum_{\tilde{Q}_\alpha \neq Q^\#} [\gamma(\operatorname{diam} \tilde{Q}_\alpha)^{-2}]^p |\tilde{Q}_\alpha| = c_n \sum_{\tilde{Q}_\alpha \neq Q^\#} \gamma^p (\operatorname{diam} \tilde{Q}_\alpha)^{n-2p}$$

$$\leq c_n \cdot (2^n - 1) \sum_{2^k \leq \operatorname{diam} Q} \gamma^p \cdot 2^{k(n-2p)}.$$

Again since $n \geq 3$ and Theorem 6 gets stronger for decreasing $p$, we may assume $p < n/2$, so the above geometric series yields

$$\int_{Q \cap E(Q_j)} |V|^p \, dx \leq C\gamma^p (\operatorname{diam} Q)^{n-2p},$$

i.e.

$$\left(\frac{1}{|Q|} \int_{Q \cap E(Q_j)} |V|^p \, dx\right) \leq C'\gamma (\operatorname{diam} Q)^{-2},$$

which is (i). This completes the verification of estimate (i), which was the last step in our proof of Theorem 6. Q.E.D.

The proof of Theorem 6 yields a slightly sharper result which will be important for the application to mathematical physics. Let $V \leq 0$ and $\mathcal{Q}^0$ be as above and define

$N(E,V) =$ number of eigenvalues of $L = -\Delta + V$ which are $< -E$;

$N_{UP}(E,V) =$ number of cubes of $\mathcal{Q}^0$ with diameter at most $CE^{-1/2}$.

Then we have

THEOREM 6′.

$$N(E,V) \leq C N_{UP}(E,V). \tag{†}$$

COROLLARY. *The sum of the absolute values of the negative eigenvalues of* $-\Delta + V$ *is at most* $C \sum_{Q \in \mathcal{Q}^0} (\operatorname{diam} Q)^{-2}$.

The corollary follows from Theorem 6′ just by integrating the estimate (†) over all $E$.

To prove Theorem 6′ we cut $R^n$ into a grid of cubes $Q^\nu$ of diameter $\sim E^{-1/2}$ and apply to each $Q^\nu$ the proof of Theorem 6. We omit the details.

Note that Theorem 6′ immediately implies the nontrivial part of theorem 5. There are probably sharp versions of Theorems 5, 6 and 6′ in which the $L \log L$ norm is used in place of the $L^p$-norm. Such results (if true) would clearly be best possible. A natural way to prove the sharp results would be to prove Lemma A with $p = 1$, possibly with a larger $S$-function such as

$$S^+(u) = \left(\sum_Q \frac{\{\sum_{Q' \subseteq Q} \|\hat{u}(Q')\|^2\}}{|Q|} \chi_Q\right)^{1/2}.$$

This leads to deep questions of Fourier analysis, such as those treated by R. Fefferman [18].

Now we come to mathematical physics. Our starting point is the Thomas-Fermi approximation, which deals with $N$-body problems in quantum mechanics. Say we have $N$ electrons and $N'$ nucleii fixed at locations $y_1, \ldots, y_{N'}$. The state of the system is given by a wave function $\psi(x_1, \ldots, x_N)$ with $\|\psi\| = 1$. We take $\psi$ scalar-valued. (Physically, this neglects the fact that electrons can

have spin up or down; but it makes no important difference in the mathematics to follow.) Electrons satisfy Fermi statistics, which means that $\psi(x_1, \ldots, x_N)$ is antisymmetric in the $x_j$. Now $N$ and $N'$ for macroscopic objects are $\sim 10^{26}$, so $\psi$ is a function of many variables. The main problem is to find $\psi$; but even if $\psi$ is given to us we still have no intuitive sense of what is happening to the electrons. What we want are more understandable quantities, such as the energy

$$E_{QM}(\psi) = \|\nabla\psi\|^2 + \langle V\psi, \psi\rangle$$

with

$$V(x_1, \ldots, x_N) = \sum_{j<k} \frac{1}{|x_j - x_k|} + \sum_{j<k} \frac{z_j z_k}{|y_j - y_k|} - \sum_{j,k} \frac{z_k}{|x_j - y_k|},$$

$$z_k = \text{ charge of } k\text{th nucleus } (1 \le z_k \le 92),$$

or the electron density

$$\rho(x_1) = N \int |\psi(x_1, x_2, \ldots, x_N)|^2 \, dx_2 \cdots dx_N.$$

The interpretation of $\rho$ is that on the average we expect to find $\int_\Omega \rho(x_1)\, dx_1$ electrons inside a subset $\Omega \subseteq R^3$.

Now Thomas-Fermi theory is an attempt to approximate $E_{QM}(\psi)$ in terms of $\rho$, and to guess both $E_{QM}(\psi)$ and $\rho$ and $\psi$ is a ground-state eigenvector (i.e. $\psi$ is picked to minimize $E_{QM}$). If we think of $\rho(x_1)$ as a charge density in classical physics, then the electrostatic potential energy associated to $\rho$ will be

$$\begin{aligned} V(\rho) &= \frac{1}{2} \int_{R^3 \times R^3} \frac{\rho(x)\rho(y)}{|x-y|}\, dx\, dy \\ &\quad + \sum_{j<k} \frac{z_j z_k}{|y_j - y_k|} - \int_{R^3} \left[ \sum_k \frac{z_k}{|x - y_k|} \right] \rho(x)\, dx \\ &\equiv V_1 + V_2 - V_3. \end{aligned}$$

Here $V_2$ and $V_3$ agree exactly with the analogue terms in $\langle V\psi, \psi\rangle$ if we just put in the definition of $\rho$ in terms of $\psi$. On the other hand, $V_1$ is only an approximation to its quantum-mechanical counterpart, and it isn't immediately clear how good the approximation is. At least we have some expression $V(\rho)$ which is a candidate to approximate the potential energy $\langle V\psi, \psi\rangle$. In deciding to use $V(\rho)$, we haven't brought in quantum mechanics.

Next we come to the kinetic energy $\|\nabla\psi\|^2$. The presence of kinetic energy in the lowest-energy state is a purely quantum-mechanical effect, since classically a distribution of charge can stay at rest with kinetic energy zero. To see how a quantum-mechanical ground state can have kinetic energy, we look first at the elementary case of $N$ particles in a box $B \subseteq R^3$ with side $\delta$. The state of the $N$ particles is a function $\phi(x_1, \ldots, x_N)$ defined on $\mathcal{B} = B \times B \times \cdots \times B$ and vanishing at $\partial\mathcal{B}$. For $\|\phi\|_{L^2(\mathcal{B})} = 1$, the kinetic

energy is given by $T = \int_B |\nabla\phi|^2\, dx_1 \cdots dx_N$. The quadratic form is easily diagonalized by separation of variables, the eigenfunctions being products $\phi(x_1 \cdots x_N) = \phi_{j_1}(x_1)\phi_{j_2}(x_2)\cdots\phi_{j_N}(x_N)$, where $\phi_j$ are the eigenfunctions of $-\Delta$ with Dirichlet boundary conditions on $B$. The kinetic energy associated to such an eigenfunction is $\lambda_{j_1} + \lambda_{j_2} + \cdots + \lambda_{j_N}$, where $\lambda_j$ is the eigenvalue corresponding to $\phi_j$.

Now if the $N$ particles are Fermions, we restrict our attention to antisymmetric functions $\phi(x_1, \ldots, x_N)$, so the eigenfunctions for $T$ are antisymmetrized products $\sum_\pi (\operatorname{sgn}\pi)\phi_{j_1}(x_{\pi(1)})\cdots\phi_{j_N}(x_{\pi(N)})$. This time we must take $j_1, j_2, \ldots, j_N$ distinct or else the antisymmetrized product will vanish. Consequently, the smallest possible kinetic energy of the $N$ Fermions in the box is equal to the sum $\lambda_1 + \lambda_2 + \cdots + \lambda_N$ of the $N$ smallest eigenvalues of $-\Delta$ on $B$. Recalling that $B$ has side $\delta$, we may use volume-counting to estimate the eigenvalues for large $N$:

$$\left\{\begin{matrix}\text{Lowest kinetic energy of} \\ N \text{ Fermions in } B\end{matrix}\right\} = c_n \int_{\substack{x\in B \\ |\xi|^2<E}} |\xi|^2\, dx\, d\xi,$$

where $E$ is picked so that

$$\left\{\begin{matrix}\text{Number of eigenvalues} \\ \text{of } -\Delta_B < E\end{matrix}\right\} = c_n \int_{\substack{x\in B \\ |\xi|^2<E}} dx\, d\xi = N.$$

The result is that the lowest possible kinetic energy of the $N$ Fermions in the box $B$ is asymptotic for large $N$ to $(\text{const})(N/\delta^3)^{5/3}\delta^3$. Note that $N/\delta^3$ may be interpreted as a density $\rho$ of particles per unit volume.

Let us now return to Thomas-Fermi theory and ask again how to guess the kinetic energy $\int |\nabla\psi|^2\, dx_1 \cdots dx_N$ in terms of $\rho$. Imagine we cut $R^3$ into boxes $B_j$ which are small, but not too small. If the boxes are small, then $\rho$ will remain roughly constant $= \rho_j$ in each $B_j$. If the boxes are not too small, then the expected number of electrons in $B_j$, which is $N_j = \rho_j \operatorname{vol}(B_j)$, will be large. We computed that the ground-state energy of $N_j$ Fermions in a box $B_j$ is given by

$$c(N_j/\operatorname{vol}(B_j))^{5/3} \operatorname{vol}(B_j) = c\rho_j^{5/3} \operatorname{vol}(B_j),$$

so it is natural to guess that the total ground-state energy of all $N = \sum_j N_j$ Fermions should be approximately

$$\sum_j c\rho_j^{5/3} \operatorname{vol}(B_j) \sim c \int_{R^3} \rho^{5/3}(x)\, dx.$$

This is the Thomas-Fermi kinetic energy. Altogether, the energy associated to $\rho$ is

$$\mathcal{E}_{\mathrm{TF}}(\rho) = c\int_{R^3} \rho^{5/3}\, dx + \frac{1}{2}\int_{R^3\times R^3} \frac{\rho(x)\rho(y)}{|x-y|}\, dx\, dy$$
$$+ \sum_{j<k} \frac{z_j z_k}{|y_1 - y_k|} - \int_{R^3} \left[\sum_k \frac{z_k}{|x-y_k|}\right] \rho(x)\, dx.$$

We hope that this gives a good approximation to $E_{QM}(\psi)$. Moreover, it is natural to try approximating the electron-density $\rho$ arising from a ground-state $\psi$ (which is picked to minimize $E_{QM}$) by a Thomas-Fermi density $\rho_{\text{TF}}$. One simply defines $\rho_{\text{TF}}$ as the density which minimizes $\mathcal{E}_{\text{TF}}(\rho)$ subject to the obvious constraints $\rho \geq 0$, $\int_{R^3} \rho(x)\,dx = N$.

How well does Thomas-Fermi theory work? In nature the electron density of atoms (with $z \geq 20$) is concentrated in rather small clouds about the nucleii, and these clouds are modeled very well by Thomas-Fermi theory. However, a small part of the total charge distribution goes to make up an outer shell of valence electrons around each atom. It is, of course, precisely these shells which account for chemical bonding, and Thomas-Fermi theory fails to predict their existence. Thus, atoms do not bind together in Thomas-Fermi theory. Remarkably, this can be used to advantage, as we shall see in a moment.

Next we describe a striking application of Thomas-Fermi theory, the so-called stability of matter. We expect that in the ground-state $\psi$ our electrons form themselves into atoms with a definite size and binding energy which stay essentially unchanged as the numbers $N$, $N'$ of particles tend to infinity. This is confirmed by the following result.

THEOREM 8 (STABILITY OF MATTER; DYSON AND LENARD [**9a**], LIEB AND THIRRING [**30**]). *Let $\psi(x_1, \ldots, x_N)$ be antisymmetric with $\|\psi\| = 1$. Then*

(A) *$E_{QM}(\psi) \geq -CN$.*

(B) *Assume $E_{QM}(\psi) \leq +C_1 N$. Then any set $\Omega \subseteq R^3$ with $\int_\Omega \rho\,dx \geq \frac{1}{2}N$ has volume at least $c_2 N$. Here $c_2$ depends only on $C_1$.*

Thus it takes a lot of energy per particle to squeeze the $N$ electrons into a volume small compared to $N$.

We shall sketch the proof of Theorem 8 given by Lieb-Thirring. The main idea is to make a rigorous comparison between $E_{QM}(\psi)$ and $E_{\text{TF}}(\rho)$.

LEMMA A (LIEB AND THIRRING). *For $\psi(x_1, \ldots, x_N)$ antisymmetric with norm 1, we have $\|\nabla\psi\|^2 \geq \mathcal{K} \int_{R^3} \rho^{5/3}\,dx$. Here $\mathcal{K}$ is a positive constant smaller than the constant $c$ in the Thomas-Fermi energy.*

Naturally, it would be very interesting to find the best possible $\mathcal{K}$. Maybe one can take $\mathcal{K}$ equal to the Thomas-Fermi value.

PROOF OF LEMMA A. By a trick one reduces matters to Corollary 2 of Theorem 1, which in fact was first proved to establish Lemma A. Set $H = \sum_{j=1}^N (-\Delta_{x_j} - \gamma\rho^{2/3}(x_j))$, where $\gamma \ll 1$ will be picked later.

Separation of variables lets us diagonalize $H$ in terms of the eigenfunctions of $H_0 = -\Delta - \gamma\rho^{2/3}$ on $R^3$. It follows that the lowest eigenvalue for an antisymmetric eigenfunction of $H$ is equal to the sum of the $N$ lowest eigenvalues of $H_0$. (If $H_0$ has $< N$ negative eigenvalues, this must be modified slightly, since above zero energy $H_0$ has continuous spectrum.) In particular, the lowest antisymmetric eigenvalue of $H$ is greater than or equal to the sum of all the negative eigenvalues of $H_0$. So Corollary 2 of Theorem 1 yields

$$\langle H\phi, \phi\rangle + E\|\phi\|^2 \geq 0 \quad \text{for } E = C\int_{R^3} [\gamma\rho^{2/3}]^{5/2}\,dx = C\gamma^{5/2}\int_{R^3} \rho^{5/3}\,dx$$

and $\phi(x_1,\dots,x_N)$ antisymmetric. However, plugging in the definitions of $\rho$ in terms of $\psi$ yields

$$\langle H\psi,\psi\rangle = \|\nabla\psi\|^2 - \int_{R^3} \gamma\rho^{2/3}(x_1)N\int |\psi(x_1,x')|^2\,dx'\,dx_1$$

$$= \|\nabla\psi\|^2 - \int_{R^3} \gamma\rho^{2/3}(x_1)\rho(x_1)\,dx_1.$$

Therefore

$$0 \le \langle H\psi,\psi\rangle + E\|\psi\|^2 = \|\nabla\psi\|^2 - \gamma\int_{R^3}\rho^{5/3}\,dx + C\gamma^{5/2}\int_{R^3}\rho^{5/3}\,dx.$$

For $\gamma$ small enough we have $C\gamma^{5/2} < \gamma$, so the last estimate proves that $\|\nabla\psi\|^2 \ge \mathcal{K}\int_{R^3}\rho^{5/3}\,dx$ with $\mathcal{K} = \gamma - C\gamma^{5/2}$. Q.E.D.

For simplicity we take $Z_k = 1$ and $N = N'$.

LEMMA B. *If $\rho \ge 0$ and $\int_{R^3}\rho\,dx = N$, then $\mathcal{E}_{\mathrm{TF}}(\rho) \ge -CN$.*

REMARKS ON THE PROOF. Set $\mathcal{E}_{\mathrm{TF}}\{y_1,\dots,y_N\} = \inf\{\mathcal{E}_{\mathrm{TF}}(\rho) \mid \rho \ge 0$ and $\int_{R^3}\rho\,dx = N\}$. The fact that atoms do not bind in Thomas-Fermi theory means that

$$(\dagger)\qquad \mathcal{E}_{\mathrm{TF}}\{y_1,\dots,y_N\} \ge \mathcal{E}_{\mathrm{TF}}\{y_1\} + \mathcal{E}_{\mathrm{TF}}\{y_2\} + \cdots + \mathcal{E}_{\mathrm{TF}}\{y_N\}.$$

That is, separating the nucleii to infinity lowers the Thomas-Fermi energy. Since $\mathcal{E}_{\mathrm{TF}}\{y_k\}$ is simply a fixed negative constant (the T-F binding energy of a hydrogen atom), (†) yields trivially $\mathcal{E}_{\mathrm{TF}}\{y_1,\dots,y_N\} \ge -CN$, which is the content of Lemma B.

The proof of (†), due to E. Teller, appears in [**30**]. One applies the methods of potential theory to the Euler-Lagrange equation for $\mathcal{E}_{\mathrm{TF}}(\rho)$, which is a quasi-linear variant of Laplace's equation. We omit the argument.

LEMMA C.

$$\left\langle \sum_{j<k} |x_j - x_k|^{-1}\psi,\psi\right\rangle \ge \frac{1}{2}\int_{R^3\times R^3}\frac{\rho(x)\rho(y)}{|x-y|}\,dx\,dy$$

$$-\beta\int_{R^3}\rho^{5/3}\,dx - C(\beta)N \quad \text{for } \beta > 0.$$

PROOF. Lemma B shows that

$$(\dagger)\qquad \beta\int_{R^3}\rho^{5/3}\,dx + \frac{1}{2}\int_{R^3\times R^3}\frac{\rho(x)\rho(y)}{|x-y|}\,dx\,dy + \sum_{j<k}|y_j - y_k|^{-1}$$

$$-\int_{R^3}\left[\sum_k |x-y_k|^{-1}\right]\rho(x)\,dx \ge -C(\beta)N.$$

Multiply both sides by $|\psi(y_1,\dots,y_N)|^2\,dy_1\cdots dy_N$ and integrate. In view of the definition of $\rho$ in terms of $\psi$, the result is

$$(\dagger\dagger)\qquad \begin{aligned}&\beta\int_{R^3}\rho^{5/3}\,dx+\frac{1}{2}\int_{R^3\times R^3}\frac{\rho(x)\rho(y)\,dx\,dy}{|x-y|}\\&+\int_{R^{3N}}\sum_{j<k}|y_j-y_k|^{-1}|\psi(y)|^2\,dy-\int_{R^3\times R^3}\frac{\rho(y)\rho(x)\,dx\,dy}{|x-y|}\geq -C(\beta)N,\end{aligned}$$

which is the statement of Lemma C. Q.E.D.

It is remarkable that the $y_k$ are interpreted as the positions of the nucleii in (†), while in (††) they are the positions of the electrons. From Lemmas A, B, C the proof of stability of matter is easy. Picking $\beta=\mathcal{H}/2$ in Lemma C, we find that

$$\begin{aligned}E_{QM}(\psi)&=\|\nabla\psi\|^2+\left\langle\sum_{j<k}|x_j-x_k|^{-1}+\sum_{j<k}|y_j-y_k|^{-1}-\sum_{j,k}|x_j-y_k|^{-1}\psi,\psi\right\rangle\\&\geq\left[\mathcal{H}\int_{R^3}\rho^{5/3}\,dx\right]\\&\quad+\left[\frac{1}{2}\int_{R^3\times R^3}\frac{\rho(x)\rho(y)}{|x-y|}\,dx\,dy-\frac{\mathcal{H}}{2}\int_{R^3}\rho^{5/3}\,dx-C\left(\frac{\mathcal{H}}{2}\right)N\right]\\&\quad+\sum_{j<k}|y_j-y_k|^{-1}-\int_{R^3}\left[\sum_k|x-y_k|^{-1}\right]\rho(x)\,dx\\&=\left[\frac{\mathcal{H}}{4}\int_{R^3}\rho^{5/3}\,dx+\frac{1}{2}\int_{R^3\times R^3}\frac{\rho(x)\rho(y)\,dx\,dy}{|x-y|}+\sum_{j<k}|y_j-y_k|^{-1}\right.\\&\qquad\left.-\int_{R^3}\left(\sum_k|x-y_k|^{-1}\right)\rho(x)\,dx\right]\\&\quad+\frac{\mathcal{H}}{4}\int\rho^{5/3}\,dx-C\left(\frac{\mathcal{H}}{2}\right)N.\end{aligned}$$

The term in brackets is $\geq -C(\mathcal{H})N$ by Lemma B, so

$$(*)\qquad E_{QM}(\psi)\geq\frac{\mathcal{H}}{4}\int_{R^3}\rho^{5/3}\,dx-\left[C(\mathcal{H})+C\left(\frac{\mathcal{H}}{2}\right)\right]N.$$

Conclusion (A), the lower bound for the total energy, is now obvious with $C = C(\mathcal{H}) + C(\mathcal{H}/2)$. To prove (B) suppose $E_{QM}(\psi) \leq C_1 N$. Then $\int_{R^3} \rho^{5/3}\,dx \leq C_2 N$ by (*). If $\int_\Omega \rho(x)\,dx \geq \frac{1}{2}N$, Hölder's inequality yields

$$\frac{1}{2}N \geq \int_\Omega \rho\,dx \leq |\Omega|^{2/5}\left(\int_\Omega \rho^{5/3}\,dx\right)^{3/5} \leq C_3|\Omega|^{2/5}N^{3/5}.$$

That is, $|\Omega| \geq C_3^{-5/2}N$, which proves Theorem 8(B). Q.E.D.

Ultimately, the proof of stability of matter rests on the volume-counting of Corollary 2 of Theorem 1. The much sharper Theorem 6′ suggests that stability of matter can be improved upon in the spirit of the *SAK* principle. Our goal is to prove that matter does not merely take up a definite volume, but is actually made of atoms of a fixed size. We may look at matter on an atomic scale by restricting the electron density $\rho(x)$ to a cube $Q$ of diameter $\sim 1$. We then expect to see the following:

(1) $\int_Q \rho \sim 1$;

(2) $\rho \geq c$ on a fixed portion of the volume of $Q$;

(3) $\int_Q \rho^{5/3} \leq C$;

(4) At least one nucleus, and at most $C$ nucleii, belongs to $Q$. Although lacking in fine detail, these properties capture the idea of an atom as a ball of definite size, carrying a nucleus and some electrons, and possibly sharing its electron cloud with its neighbors.

THEOREM 9 (ATOMIC STRUCTURE OF MATTER).[4] *Let $\psi(x_1, \ldots, x_N)$ be a Fermion wave function with $E_{QM}(\psi) < -\epsilon N$. Then there is a collection of at least $cN$ pairwise disjoint cubes $Q_\gamma$ of size $c \leq \operatorname{diam} Q_\gamma \leq C$ on which the electron density satisfies* (1)–(4) *above. In fact,* (3) *holds in the stronger form*

(3)′ $\int_{Q_\gamma} \rho^3 \leq C$.

The various constants $c$, $C$ in Theorem 9 depend on $\epsilon$ but are independent of $N$. The $\epsilon$-dependence is unavoidable, as one sees already from the excited states of a single hydrogen atom. Theorem 9 would be nonsense without the hypothesis $E_{QM}(\psi) < -\epsilon N$, for then we could disperse the nucleii to infinity and take $\rho$ spread thinly through a large volume. The resulting $\psi$ has no cubes of diameter $\sim 1$ with atomic structure (1)–(4), while $0 < E_{QM}(\psi) < +\epsilon N$, where we may take $\epsilon > 0$ arbitrarily small.

PROOF OF THEOREM 9. Set $H = -\Delta + V(x)$ where

$$V(x) = \sum_{j<k} |x_j - x_k|^{-1} + \sum_{j<k} |y_j - y_k|^{-1} - \sum_{j,k} |x_j - y_k|^{-1}.$$

So again we are taking $\psi$ scalar-valued and supposing the nucleii have charge 1. As before, these restrictions may be trivially removed.

1. *Geometry of cubes.* Start with a grid of cubes of side $R$ (large const. to be picked later). Subdivide the cubes, stopping at $Q_\nu$ of side $\delta_\nu$ if $3Q_\nu$ contains at most $K$ nucleii ($K \gg 1$ to be picked later). Two $Q_\nu$ which touch

[4] In the version of this article distributed at the Denver meeting, Theorem 9 is given as a conjecture. The author subsequently found the proof given here.

have comparable $\delta_\nu$, and the $\delta_\nu$ are bounded below. Note that $9Q_\nu$ contains at least $K$ nucleii unless $\delta_\nu = R$. Define sets of indices as follows:

$\nu \in B$ if $9Q_\nu$ contains at least one nucleus;

$\nu \in G$ if $\nu \in B$ and $\delta_\mu \sim \delta_\nu$ whenever $Q_\mu \cap 9Q_\nu \neq \emptyset$;

$\nu \in K$ if $Q_\nu$ contains at least $cK$ nucleii.

LEMMA 1.

$$\sum_k \delta_\nu^{-1} + \frac{N}{R} \sim \sum_G \delta_\nu^{-1} + \frac{N}{R} \sim \sum_B \delta_\nu^{-1} + \frac{N}{R}.$$

PROOF. If $\nu \in G$ then either $\delta_\nu = R$ or else for some $\mu$ we have $(\mu, \nu) \in H = \{(\mu,\nu) \mid \nu \in G,\ \mu \in K,\ Q_\mu \cap 9Q_\nu \neq \emptyset\}$. Since $|B| \leq CN$, and since for each $\mu$ there are only $\leq C$ different $\nu$ with $(\mu,\nu) \in H$ [for, $\delta_\mu \sim \delta_\nu$ and $Q_\nu \subseteq 100Q_\mu$], we can write

$$\text{(a)} \qquad \sum_G \delta_\nu^{-1} \leq \frac{N}{R} + \sum_{(\mu,\nu)\in H} \delta_\nu^{-1} \leq \frac{N}{R} + C\sum_H \delta_\mu^{-1} \leq \frac{N}{R} + C'\sum_K \delta_\mu'.$$

Now let $\nu \in B$. Either $\nu \in G$ or else $\delta_\nu \not\sim \delta_\mu$ for some $Q_\mu$ meeting $9Q_\nu$. We cannot have $\delta_\mu \gg \delta_\nu$, since $9Q_\nu \subseteq 3Q_\mu$, so the cutting procedure would have stopped before reaching $Q_\nu$. Hence $\delta_\mu \leq \delta_\nu/100$. Note that $\delta_\mu < R$, so $\mu \in B$. Also note that $100Q_\mu \subseteq 100Q_\nu$.

Now set $\nu_1 = \mu$ and repeat the above argument for $Q_{\nu_1}$. Continuing in this way, we get a sequence $\nu = \nu_0, \nu_1, \nu_2, \ldots$ so that $\delta_{\nu_{k+1}} \leq \delta_{\nu_k}/100$, $100Q_{\nu_{k+1}} \subseteq 100Q_{\nu_k}$. The sequence stops at $\nu_s$ when $\nu_s \in G$. Note that the sequence must stop eventually, since the $\delta_\nu$ are bounded below. Hence for every $\nu \in B$ there is a $\mu \in G$ so that $(\mu,\nu) \in H' = \{(\mu,\nu) \mid \mu \in G,\ \nu \in B,\ 100Q_\mu \subseteq 100Q_\nu\}$. So

$$\sum_B \delta_\nu^{-1} \leq \sum_{(\mu,\nu)\in H'} \delta_\nu^{-1} \leq C\sum_G \delta_\mu^{-1};$$

this last estimate holds because

$$\sum (\text{side } Q)^{-1} \leq C\delta_\mu^{-1},$$

where the sum is taken over all dyadic cubes $Q$ with $100Q_\mu \subseteq 100Q$. Combining the above with (a) yields

$$\sum_B \delta_\nu^{-1} + \frac{N}{R} \leq C\sum_G \delta_\nu^{-1} + \frac{N}{R} \leq C'\sum_K \delta_\nu^{-1} + \frac{CN}{R}.$$

Since $K \subseteq B$ we also have

$$\sum_K \delta_\nu^{-1} + \frac{N}{R} \leq \sum_B \delta_\nu^{-1} + \frac{N}{R},$$

which completes the proof of Lemma 1. Q.E.D.

2. *Estimates for potential energy.* Suppose the electrons are in position $x = (x_1, \ldots, x_N) \in (R^3)^N$. Set $L(\nu, x) =$ number of electrons in $Q_\nu = \sum_k \chi_{Q_\nu}(x_k)$.

Define

$$V(x) = \sum_{j<k} |x_j - x_k|^{-1} + \sum_{j<k} |y_j - y_k|^{-1} - \sum_{j,k} |x_j - y_k|^{-1},$$

$$V_L(x) = \sum_{j<k} |x_j - x_k|^{-1} + \sum_{j<k} |y_j - y_k|^{-1}$$

$$- \sum_{\nu} \sum_{j,k} |x_j - y_k|^{-1} \chi_{Q_\nu}(x_j) \chi_{(3Q_\nu)^c}(y_k),$$

$$V_M(z) = \sum_{\delta_\nu > \delta_0} \sum_{k} |z - y_k|^{-1} \chi_{3Q_\nu}(y_k) \chi_{Q_\nu}(z), \qquad z \in R^3,$$

$$V_S(z) = \sum_{\delta_\nu \leq \delta_0} \sum_{k} |z - y_k|^{-1} \chi_{3Q_\nu}(y_k) \chi_{Q_\nu}(z), \qquad z \in R^3.$$

Here $\delta_0$ is a small const. to be picked later. We have

$$V(x) = V_L(x) - \sum_k V_M(x_k) - \sum_k V_S(x_k).$$

LEMMA 2.

$$V_L(x) \geq cK^2 \sum_B \delta_\nu^{-1} + c \sum_{L(\nu) \geq L_0} L^2(\nu, x) \delta_\nu^{-1} - \frac{CN}{R}.$$

PROOF. First note that for any complex function $\varphi(x)$ on $R^3$ we have

$$\frac{1}{2} \int \frac{\varphi(x)\overline{\varphi(y)}}{|x-y|}\, dx\, dy = c \int \frac{|\hat{\varphi}(\xi)|^2}{|\xi|^2}\, d\xi \geq 0.$$

Apply this[5] to $\varphi(x) = \sum_j \rho_j(x) - \sum_k \eta_k(x)$, where:

$\rho_j(x)$ is spherically symmetric about $x_j$, has total integral 1, is supported in $|x - x_j| < \delta_j' = \delta_\nu/10$ for the $Q_\nu$ containing $x_j$, and is bounded by $C(\delta_j')^{-3}$.

$\eta_k(x)$ is spherically symmetric about $y_k$, has total integral 1, is supported in $|x - y_k| < \delta_k'' = \delta_\nu/10$ for the $Q_\nu$ containing $y_k$, and is bounded by $C(\delta_k'')^{-3}$.

[5]The key idea of replacing point charges by continuous distributions appears already in Dyson-Lenard [**9a**].

Thus,

$$0 \le \frac{1}{2}\sum_j \int \frac{\rho_j(x)\rho_j(y)}{|x-y|}\,dx\,dy + \frac{1}{2}\sum_k \int \frac{\eta_k(x)\eta_k(y)}{|x-y|}\,dx\,dy$$
$$+ \sum_{j<k} \int \frac{\rho_j(x)\rho_k(y)}{|x-y|}\,dx\,dy + \sum_{j<k} \int \frac{\eta_j(x)\eta_k(y)}{|x-y|}\,dx\,dy$$
$$- \sum_{j,l} \int \frac{\rho_j(x)\eta_l(y)}{|x-y|}\,dx\,dy$$
$$\equiv \mathrm{I} + \mathrm{II} + \mathrm{III} + \mathrm{IV} - \mathrm{V}.$$

Now

$$\mathrm{I} \le \sum_\nu CL(\nu,x)\delta_\nu^{-1}, \qquad \mathrm{II} \le \sum_{\nu\in B} CK\delta_\nu^{-1}$$

since

$$Q_\nu \text{ contains } \begin{cases} \text{at most } K \text{ nucleii always,} \\ \text{no nucleii unless } \nu \in B. \end{cases}$$

Also,

$$\mathrm{III} \le \sum_{j<k} \left( \frac{1}{|x_j - x_k|} - c(\delta_j' + \delta_k')^{-1} \chi_{|x_j - x_k| < (\delta_j' + \delta_k')/10} \right)$$

by subharmonicity of $z \mapsto |z|^{-1}$. Similarly,

$$\mathrm{IV} \le \sum_{j<k} \left( \frac{1}{|y_j - y_k|} - c(\delta_j'' + \delta_k'')^{-1} \chi_{|y_1 - y_k| < (\delta_j'' + \delta_k'')/10} \right).$$

Finally,

$$\mathrm{V} \ge \sum_{j,k} |x_j - y_k|^{-1} \chi_{|x_j - y_k| \ge \delta_j' + \delta_k''}$$
$$\ge \sum_\nu \sum_{j,k} |x_j - y_k|^{-1} \chi_{Q_\nu}(x_j) \chi_{(3Q_\nu)^c}(y_k).$$

Combining our estimates for I–V, we get

$$0 \le V_L(x) + C\sum_B K\delta_\nu^{-1} + C\sum_\nu L(\nu,x)\delta_\nu^{-1}$$
$$- c\sum_\nu \delta_\nu^{-1} \left( \#(j,k) \mid x_j, x_k \in Q_\nu, |x_j - x_k| < \frac{\delta_\nu}{10^8},\ j<k \right)$$
$$- c\sum_\nu \delta_\nu^{-1} \left( \#(j,k) \mid y_j, y_k \in Q_\nu, |y_j - y_k| < \frac{\delta_\nu}{10^8},\ j<k \right).$$

Thus

$$V_L(x) \ge c\sum_\nu \delta_\nu^{-1}[\#(E_\nu)] + c\sum_\nu \delta_\nu^{-1}[\#(F_\nu)] - C\sum_B K\delta_\nu^{-1}$$
$$= C\sum_\nu L(\nu,x)\delta_\nu^{-1}$$

with

$$E_\nu = \{(j<k) \mid x_j, x_k \in Q_\nu, |x_j - x_k| < \delta_\nu/10^8\},$$

$$F_\nu = \{(j<k) \mid y_j, y_k \in Q_\nu, |y_j - y_k| < \delta_\nu/10^8\}.$$

Observe that

$$\#(E_\nu) \geq cL^2(\nu,x) \quad \text{if } L(\nu,x) \geq L_0,$$

$$\#(F_\nu) \geq cK^2 \qquad \text{if } \nu \in K.$$

This follows from the pigeonhole principle. So

$$V_L(x) \geq \sum_K cK^2\delta_\nu^{-1} + \sum_{L(\nu,x)\geq L_0} cL^2(\nu,x)\delta_\nu^{-1} - \sum_B CK\delta_\nu^{-1}$$

(+)

$$- \sum_{\nu\in B} CL(\nu,x)\delta_\nu^{-1} - \sum_{\nu\notin B} CL(\nu,x)\delta_\nu^{-1}.$$

Now if we pick $K$ large enough (depending on $L_0$ only), then

$$\sum_B CK\delta_\nu^{-1} \leq \frac{1}{20}\sum_K cK^2\delta_\nu^{-1} + \frac{CN}{R} \quad \text{by Lemma 1,}$$

$$\sum_{\nu\in B} CL(\nu,x)\delta_\nu^{-1} \leq \frac{1}{20}\sum_{L(\nu,x)\geq L_0} cL^2(\nu,x)\delta_\nu^{-1} + \frac{1}{20}\sum_B CK\delta_\nu^{-1},$$

and we know that the right-hand side is

$$\leq \frac{1}{20}\sum_{L(\nu,x)\geq L_0} cL^2(\nu,x)\delta_\nu^{-1} + \frac{1}{20}\sum_K cK^2\delta_\nu^{-1} + \frac{CN}{R},$$

$$\sum_{\nu\notin B} CL(\nu,x)\delta_\nu^{-1} = \frac{C}{R}\sum_{\nu\notin B} L(\nu,x) \leq \frac{CN}{R}.$$

Substituting these into (+), we get

$$V_L(x) \geq c'\sum_K K^2\delta_\nu^{-1} + c'\sum_{L(\nu,x)\geq L_0} L^2(\nu,x)\delta_\nu^{-1} - \frac{CN}{R}.$$

Another application of Lemma 1 yields the conclusion of Lemma 2. Q.E.D.

3. *The exclusion principle.*

LEMMA 3. *Let $\psi(x_1,\dots,x_N)$ be antisymmetric. Then*

$$\|\nabla\psi\|^2 \geq \langle V^+\psi,\psi\rangle \quad \text{where } V^+(x) = \sum_{L(\nu,x)\geq 2} cL^{5/3}(\nu,x)\delta_\nu^{-2}.$$

PROOF. Let $u(x_1,\dots,x_L)$ be antisymmetric on $Q^L$, $L \geq 2$. Then

$$\|\nabla u\|^2_{L^2(Q^L)} \geq cL^{5/3}(\text{side } Q)^{-2}\|u\|^2.$$

This follows by expanding $u(x_1,\dots,x_L)$ in $\psi_{\lambda_1}(x_1),\dots,\psi_{\lambda_L}(x_L)$ with $\psi_\lambda$ eigenfunctions for the Neumann problem on $Q$. We require $L \geq 2$ since the lowest eigenvalue of the Neumann problem is zero.

Consequently, if $\psi$ is antisymmetric on $I = \prod_\nu (Q_\nu)^{L_\nu}$, then

$$\|\nabla\psi\|^2_{L^2(I)} \geq \left( c \sum_{L_\nu \geq 2} L_\nu^{5/3} \delta_\nu^{-2} \right) \|\psi\|^2_{L^2(I)}.$$

Cut up $(R^3)^N$ into boxes $I = Q_{\nu_1} \times \cdots \times Q_{\nu_N}$. Apply the above estimate on each I and sum over I to get

$$\|\nabla\psi\|^2 \geq \langle V^+\psi, \psi\rangle \quad \text{with } V^+ = \sum_I c \sum_{L_\nu(I) \geq 2} L_\nu^{5/3}(I) \delta_\nu^{-2} \chi_I,$$

$L_\nu(I) =$ [number of $k$ with $\nu_k = \nu$]. Since $L(\nu, x) = L_\nu(I)$ for $x \in I$, we see that $V^+$ is the same potential that appears in the statement of Lemma 3. Q.E.D.

4. *The uncertainty principle.*

LEMMA 4. *Suppose we pick $\delta_0$ small enough, depending on $K$ but not on $N$. Then*

$$\frac{1}{4}\|\nabla\psi\|^2 - \sum_k \langle V_S(x_k)\psi, \psi\rangle \geq -\langle V_{CR}\psi, \psi\rangle,$$

*where*

$$V_{CR}(x) = C \sum_{\delta_\nu \leq \delta_0} KL(\nu, x)\delta_\nu^{-1}.$$

PROOF. Set $V_\nu(z) = \sum_{y_l \in 3Q_\nu} |z - y_l|^{-1} \chi_{Q_\nu}(z)$. Thus, $V_S(z) \leq \sum_{\delta_\nu \leq \delta_0} V_\nu(z)$. We shall prove the following estimate for functions $u(z)$ on $Q_\nu$:

$$\text{(A)} \qquad \tfrac{1}{4}\|\nabla u\|^2_{L^2(Q_\nu)} - \langle V_\nu u, u\rangle \geq -(CK/\delta_\nu)\|u\|^2_{L^2(Q_\nu)}.$$

To prove (A) write $u = (u - \overline{u}) + \overline{u}$ with $\overline{u} = \mathrm{Av}_{Q_\nu} u$. Then

$$\begin{aligned} \langle V_\nu u, u\rangle &\leq 2\langle V_\nu \overline{u}, \overline{u}\rangle + 2\langle V_\nu(u - \overline{u}), (u - \overline{u})\rangle \\ &\leq 2|\overline{u}|^2 \|V_\nu\|_{L^1(Q_\nu)} + 2\|V_\nu\|_{L^{3/2}} \|u - \overline{u}\|^2_{L^6(Q_\nu)} \qquad (+) \\ &\leq 2|\overline{u}|^2 \|V_\nu\|_{L^1(Q_\nu)} + C\|V_\nu\|_{L^{3/2}} \|\nabla u\|^2_{L^2(Q_\nu)} \end{aligned}$$

by Hölder and Sobolev. Now $\|V_\nu\|_{L^1(Q_\nu)} \leq CK/\delta_\nu \cdot \delta_\nu^3$ since the sum defining $V_\nu$ has at most $K$ terms. So the first term on the right of (+) is at most $(CK/\delta_\nu)\|u\|^2_{L^2(Q_\nu)}$. Also

$$\|V_\nu\|_{L^{3/2}(Q_\nu)} \leq \sum_{y_l \in 3Q_\nu} \left( \int_{Q_\nu} |z - y_l|^{-3/2}\, dz \right)^{2/3} \leq CK\delta_0$$

since $\delta_\nu \le \delta_0$ and the sum contains at most $K$ terms. Picking $\delta_0 < 1/4CK$, we dominate the 2nd term on the right of (+) by $\frac{1}{4}\|\nabla u\|^2_{L^2(Q_\nu)}$, completing the proof of (A).

To deduce Lemma 4 from (A) we write

$$\frac{1}{4}\|\nabla\psi\|^2 - \sum_k \langle V_S(x_k)\psi,\psi\rangle \ge \sum_{\substack{k,\nu \\ (\delta_\nu \le \delta_0)}} \frac{1}{4}\|\nabla_{x_k}\psi\|^2_{L^2(Q_\nu)} - \langle V_\nu(x_k)\psi,\psi\rangle,$$

as follows from

$$V_S(z) \le \sum_{\substack{\nu \\ (\delta_\nu \le \delta_0)}} V_\nu(z).$$

Applying estimate (A), we get

$$\begin{aligned}\frac{1}{4}\|\nabla\psi\|^2 - \sum_k \langle V_S(x_k)\psi,\psi\rangle &\ge -\sum_{\substack{k,\nu \\ (\delta_\nu \le \delta_0)}} CK\delta_\nu^{-1}\int_{\substack{x_k \in Q_\nu \\ x' \text{ arb}}} |\psi(x)|^2\,dx \\ &= -\sum_{\delta_\nu \le \delta_0}\sum_k CK\delta_\nu^{-1}\langle \chi_{Q_\nu}(x_k)\psi,\psi\rangle \\ &= -\sum_{\delta_\nu \le \delta_0} CK\delta_\nu^{-1}\langle L(\nu,x)\psi,\psi\rangle = -\langle V_{CR}(x)\psi,\psi\rangle. \quad \text{Q.E.D.}\end{aligned}$$

LEMMA 4′. *For $\delta_\nu \ge \delta_0$, $\nu \in B$, we have*

$$\tfrac{1}{4}\|\nabla u\|^2_{L^2(Q_\nu)} - \langle V_\nu u, u\rangle = -E(\delta_0, K)\|u\|^2_{L^2(Q_\nu)}.$$

PROOF. Subdivide $Q_\nu$ into cubes of side $\sim \delta_0$ and use the proof of estimate (A) above on each of the small cubes. Q.E.D.

This will be used to control $V_M$, since

$$V_M(z) \le \sum_{\substack{\delta_\nu > \delta_0 \\ \nu \in B}} V_\nu(z).$$

5. *Stability of matter.* Our method gives an alternate proof of Theorem 8(A), which we now present. As operators on antisymmetric functions, we have

$$\begin{aligned} H &= -\Delta + V(x) = -\Delta + V_L(x) - \sum_k V_S(x_k) - \sum_k V_M(x_k) \\ &= -\frac{1}{4}\Delta + V_L(x) - \frac{1}{4}\Delta + \left(-\frac{1}{4}\Delta - \sum_k V_S(x_k)\right) + \left(-\frac{1}{4}\Delta - \sum_k V_M(x_k)\right) \\ &\geq -\frac{1}{4}\Delta + \left(cK^2 \sum_B \delta_\nu^{-1} + c \sum_{L(\nu,x)\geq L_0} L^2(\nu,x)\delta_\nu^{-1} - \frac{CN}{R}\right) \\ &\quad + \left(c \sum_{L(\nu,x)\geq 2} L^{5/3}(\nu,x)\delta_\nu^{-2}\right) - \left(C \sum_{\delta_\nu \leq \delta_0} KL(\nu,x)\delta_\nu^{-1}\right) \\ &\quad + \left(-\frac{1}{4}\Delta - \sum_k V_M(x_k)\right) \end{aligned}$$

by virtue of Lemmas 2, 3 and 4. Now $\delta_\nu \leq \delta_0$ implies $\nu \in B$, so

$$C \sum_{\delta_\nu \leq \delta_0} KL(\nu,x)\delta_\nu^{-1} \leq \frac{c}{2}K^2 \sum_B \delta_\nu^{-1} + \frac{c}{2} \sum_{L(\nu,x)\geq 2} L^{5/3}(\nu,x)\delta_\nu^{-2}$$

as long as $K \geq C/\frac{c}{2}$ to take care of the case $L(\nu,x) \leq 1$, and $\delta_0 < \frac{c}{2}(CK)^{-1}$ to take care of the case $L(\nu,x) \geq 2$. Hence as operators on antisymmetric functions

$$\begin{aligned} H \geq &-\frac{1}{4}\Delta + c'K^2 \sum_B \delta_\nu^{-1} + c' \sum_{L(\nu,x)\geq L_0} L^2(\nu,x)\delta_\nu^{-1} - \frac{CN}{R} \\ &+ \left(-\frac{1}{4}\Delta - \sum_k V_M(x_k)\right). \end{aligned}$$

If we pick $K \geq L_0$, then

$$\sum_B L^2(\nu,x)\delta_\nu^{-1} \leq K^2 \sum_B \delta_\nu^{-1} + \sum_{L(\nu,x)\geq L_0} L^2(\nu,x)\delta_\nu^{-1},$$

so

$$H \geq -\frac{1}{4}\Delta + c\sum_B L^2(\nu,x)\delta_\nu^{-1} - \frac{CN}{R} + \left(-\frac{1}{4}\Delta - \sum_k V_M(x_k)\right)$$

on antisymmetric functions. Lemma 4′ yields

$$-\frac{1}{4}\Delta_{x_k} - V_M(x_k) \geq -E(\delta_0, K) \sum_{\substack{\delta_\nu > \delta_0 \\ \nu \in B}} \chi_{Q_\nu}(x_k),$$

so

$$-\frac{1}{4}\Delta - \sum_k V_M(x_k) \geq -E(\delta_0, K) \sum_{\substack{\delta_\nu > \delta_0 \\ \nu \in B}} L(\nu,x).$$

Thus

$$(**) \qquad H \geq -\frac{1}{4}\Delta + c\sum_B L^2(\nu,x)\delta_\nu^{-1} - E(\delta_0,K)\sum_{\substack{\delta_\nu>\delta_0\\ \nu\in B}} L(\nu,x) - \frac{CN}{R}$$

as operators on antisymmetric functions. Since $\sum_\nu L(\nu,x) = N$, it follows at once that $H \geq -[E(\delta_0,K)+C/R]N$ on antisymmetric functions. Recall that $\delta_0$, $K$, $R$ are fixed consts. independent of $N$, so we have proved stability of matter.

6. *Bound states.* We now make use of the critical hypothesis $E_{QM}(\psi) < -\epsilon N$. So suppose $\psi$ is antisymmetric, $\|\psi\| = 1$, $\langle H\psi,\psi\rangle < -\epsilon N$. Then $(**)$ yields

$$\begin{aligned}(\#) \qquad -\epsilon N &\geq \frac{1}{4}\|\nabla\psi\|^2 + c\sum_B \langle L^2(\nu,x)\psi,\psi\rangle\delta_\nu^{-1}\\ &\quad - E(\delta_0,K)\sum_{\substack{\delta_\nu>\delta_0\\ \nu\in B}}\langle L(\nu,x)\psi,\psi\rangle - \frac{CN}{R}\\ &\geq \frac{1}{4}\|\nabla\psi\|^2 + c\sum_B\left(\int_{Q_\nu}\rho\right)^2\delta_\nu^{-1} - E(\delta_0,K)\sum_{\substack{\delta_\nu>\delta_0\\ \nu\in B}}\left(\int_{Q_\nu}\rho\right) - \frac{CN}{R}.\end{aligned}$$

Here we have used the identity $\langle L(\nu,x)\psi,\psi\rangle = \int_{Q_\nu}\rho$ and the Cauchy-Schwartz inequality $\langle L^2(\nu,x)\psi,\psi\rangle \geq \langle L(\nu,x)\psi,\psi\rangle^2$. (Recall $|\psi(x)|^2\,dx$ is a probability measure.) Picking $R \gg 1$ so that $CN/R < \epsilon N/2$, we conclude that

$$E(\delta_0,K)\sum_{\substack{\delta_\nu>\delta_0\\ \nu\in B}}\left(\int_{Q_\nu}\rho\right) \geq \frac{\epsilon}{2}N,$$

i.e.

$$\sum_{\substack{\delta_\nu>\delta_0\\ \nu\in B}}\left(\int_{Q_\nu}\rho\right) \geq \mathcal{K}N, \qquad \mathcal{K} = \frac{\epsilon/2}{E(\delta_0,K)}.$$

Also

$$E(\delta_0,K)\sum_{\substack{\delta_\nu>\delta_0\\ \nu\in B}}\left(\int_{Q_\nu}\rho\right) \leq E(\delta_0,K)\int_{R^3}\rho = N\cdot E(\delta_0,K),$$

so $(\#)$ implies

$$\sum_B \delta_\nu^{-1}\left(\int_{Q_\nu}\rho\right)^2 \leq C'N \quad \text{and} \quad \|\nabla\psi\|^2 \leq C'N.$$

To exploit this last estimate, take arbitrary potentials $W_\nu(z)$ supported in $Q_\nu$ ($\nu\in B$, $\delta_\nu > \delta_0$) with $\|W_\nu\|_{L^{3/2}} \leq \bar{c} \ll 1$. Sobolev yields

$$-\Delta_z - \sum_\nu W_\nu(z) \geq -1$$

if $\bar{c}$ is taken small enough. Hence,

$$\|\nabla\psi\|^2 - \sum_{\nu,k}\langle W_\nu(x_k)\psi,\psi\rangle \geq -N,$$

i.e.

$$N + \|\nabla\psi\|^2 \geq \sum_{\substack{\delta_\nu > \delta_0 \\ \nu \in B}} \left( \int W_\nu \rho \right).$$

Hence

$$\sum_{\substack{\delta_\nu > \delta_0 \\ \nu \in B}} \left( \int W_\nu \rho \right) \leq C'' N$$

whenever $W_\nu$ supported in $Q_\nu$ has $\|W_\nu\|_{L^{3/2}} \leq \bar{c}$. This means

$$\sum_{\substack{\delta_\nu > \delta_0 \\ \nu \in B}} \left( \int_{Q_\nu} \rho^3 \right)^{1/3} \leq C'' N.$$

Set $B_1 = \{\nu \in B \mid \delta_\nu > \delta_0\}$. We know that $\delta_0 < \delta_\nu \leq R$ for $\nu \in B_1$, i.e. all the cubes in $B_1$ are roughly the same size. We know from the previous estimates that

$$(+) \qquad \begin{cases} \sum_{\nu \in B_1} \left( \int_{Q_\nu} \rho \right) \geq \mathcal{H} N, \\ \sum_{\nu \in B_1} \left( \int_{Q_\nu} \rho \right)^2 + \left( \int_{Q_\nu} \rho^3 \right)^{1/3} \leq C N, \end{cases}$$

where $\mathcal{H}, C$ depend on $\epsilon, \delta_0, K, R$ but not on $N$. Set

$$B_2 = \left\{ \nu \in B_1 \mid \left( \int_{Q_\nu} \rho \right) \geq \frac{\mathcal{H}}{2C} \left[ \left( \int_{Q_\nu} \rho \right)^2 + \left( \int_{Q_\nu} \rho^3 \right)^{1/3} \right] \right\}.$$

We have

$$\sum_{\nu \in B_1 \setminus B_2} \left( \int_{Q_\nu} \rho \right) \leq \frac{\mathcal{H}}{2C} \sum_{\nu \in B_1 \setminus B_2} \left[ \left( \int_{Q_\nu} \rho \right)^2 + \left( \int_{Q_\nu} \rho^3 \right)^{1/3} \right] \leq \frac{\mathcal{H}}{2} N$$

by (+); hence

$$(\mathrm{I}) \qquad \sum_{\nu \in B_2} \left( \int_{Q_\nu} \rho \right) \geq \frac{\mathcal{H}}{2} N.$$

Also for $\nu \in B_2$ we have

$$(\mathrm{II}) \qquad \left( \int_{Q_\nu} \rho \right)^2 \leq \frac{2C}{\mathcal{H}} \left( \int_{Q_\nu} \rho \right), \quad \text{i.e.} \quad \int_{Q_\nu} \rho \leq C',$$

and

$$(\mathrm{III}) \qquad \left( \int_{Q_\nu} \rho^3 \right)^{1/3} \leq \frac{2C}{\mathcal{H}} \left( \int_{Q_\nu} \rho \right), \quad \text{i.e.} \quad \int_{Q_\nu} \rho^3 \leq C''.$$

In view of (I), (II), and the fact that $B_2$ contains at most $CN$ indices (recall $\nu \in B_2$ implies $9Q_\nu$ contains at least one nucleus, and the $Q_\nu$ have sides $\delta_\nu$ of roughly fixed size, $\delta_0 \le \delta_\nu \le R$), we conclude that

$$\left(\int_{Q_\nu} \rho\right) \ge \mathcal{H}' \quad \text{for at least a fixed portion } \mathcal{H}' \text{ of the } \nu \in B_2.$$

Otherwise,

$$\begin{aligned}\sum_{\nu \in B_2} \left(\int_{Q_\nu} \rho\right) &\le \mathcal{H}' \cdot (\text{Number of } \nu \in B_2) \\ &\quad + C'\left(\text{Number of } \nu \in B_2 \text{ with } \int_{Q_\nu} \rho \ge \mathcal{H}'\right) \\ &\le C\mathcal{H}'N + C'\mathcal{H}'N < \mathcal{H}N/2.\end{aligned}$$

Thus there is a subset $B_3 \subseteq B_2$, so that

$$\mathcal{H}' \le \int_{Q_\nu} \rho \le C' \quad \text{for } \nu \in B_3,$$

$$\int_{Q_\nu} \rho^3 \le C'' \quad \text{for } \nu \in B_3.$$

For each $\nu \in B_3$ there is a nucleus in $9Q_\nu$. There are at least $\mathcal{H}'N$ elements of $B_3$. The $Q_\nu$, $\nu \in B_3$, are pairwise disjoint and have sides bounded above and below, independently of $N$.

This is almost the conclusion of the theorem. To complete the proof we discard all the $Q_\nu$, $\nu \in B_3$, with more than $\overline{C}$ nucleii lying within distance $15 \operatorname{diam} Q_\nu$ of $Q_\nu$. If $\overline{C}$ is taken large enough, then we have discarded at most $\mathcal{H}'N/2$ cubes $Q_\nu$, since the total number of nucleii is $N$ and the cubes $Q_\nu$ ($\nu \in B_3$) are all roughly the same size.

Now the cubes $Q_\nu$ ($\nu \in B_3$, not discarded) are pairwise disjoint, at least $cN$ in number, and satisfy the conclusions of Theorem 9. Q.E.D.

Our next result indicates that atoms bind together chemically. We look only at hydrogen, so it is now essential to take the nucleii of charge 1 and the electrons to have a two-component spin. (This means the wave function takes the form $\psi(x_1, \sigma_1, \ldots, x_N, \sigma_N)$ with $\sigma_k = \pm 1$; see [**30**].) We have $E_{QM}(\psi) \ge -E_1 N$, where the best constant $E_1$ represents the binding energy per nucleus. Simple numerical calculations for a single $H_2$-molecule prove that $E_1$ is strictly greater than the binding energy $E_0$ for a single hydrogen atom.

THEOREM 10 (CHEMICAL BONDING). *Let $\psi(x_1, \sigma_1, \ldots, x_N, \sigma_N)$ be a Fermion wave function with energy $E_{QM}(\psi) < -(E_0 + \epsilon)N$. Then there are at least $cN$ pairs $\{y_{m_1}, y_{n_1}\}, \ldots, \{y_{m_s}, y_{n_s}\}$ of nucleii, all distinct, with $|y_{m_1} - y_{n_1}| \le C$.*

Unless we put in enough energy to make isolated atoms, there must be many pairs of nucleii which stay close together. It would be interesting to show that under suitable conditions, hydrogen forms a diatomic gas. This is, of course, a problem of quantum statistical mechanics. Another interesting open problem is to sharpen our estimates enough to make the binding energy

$E_1$ computable in principle. (The author has recently solved this; details will appear elsewhere.) Theorem 10 is established by combining the proof of Theorem 9 with simple estimates for a big $Q_\nu$ containing a single nucleus.

We close this section by noting a connection between Theorem 6′, Theorem 9, and Thomas-Fermi theory. We ask how it can happen that the Thomas-Fermi kinetic energy is comparable to the true quantum-mechanical kinetic energy, instead of being much too small. There is an obvious example of a Fermion wave function $\psi(x, \dots, x_N)$ whose quantum and T-F kinetic energies are comparable. Given disjoint cubes $Q_1, \dots, Q_N \subseteq R^3$, we define "bump" functions $\phi_j(x)$ supported on $Q_j$, with $\|\phi_j\| = 1$ and $\|\nabla\phi_j\| \sim (\operatorname{diam} Q_j)^{-1}$. Then we define $\psi(x_1, \dots, x_N)$ as the antisymmetrized product of the $\phi_j$:

$$\psi(x_1, \dots, x_N) = \frac{1}{\sqrt{N!}} \sum_\pi \phi_1(x_{\pi(1)}) \cdots \phi_N(x_{\pi(N)}).$$

Both the quantum and T-F kinetic energies are $\sim \sum_j (\operatorname{diam} Q_j)^{-2}$.

We have in this example $\rho(x) = \sum_j \phi_j^2$. Both $\psi$ and $\rho$ represent a situation in which we expect to find one particle in each of the $Q_j$, and that particle is (roughly) equally likely to be found anywhere in $Q_j$. The $Q_j$ are crudely analogous to atoms. The next theorem shows that whenever the Thomas-Fermi and quantum kinetic energies are comparable, the situation is much like our simple example.

THEOREM 11. *Let $\rho$ be the electron density for a Fermion wave function $\psi$ whose T-F and quantum kinetic energies are comparable:*

$$(\dagger) \qquad \|\nabla\psi\|^2 \le C \int_{R^3} \rho^{5/3}\, dx.$$

*Then there is a family $\{Q_j\}$ of pairwise disjoint cubes in $R^3$ with the following properties.*

(i) $\int_{Q_j} \rho\, dx \sim 1$;
(ii) $\int_{Q_j} \rho^{5/3}\, dx \sim (\operatorname{diam} Q_j)^{-2}$;
(iii) $\rho \sim (\operatorname{diam} Q_j)^{-3}$ on a subset $E_j \subseteq Q_j$ with $|E_j|/|Q_j| \ge c$;
(iv) $\int_{R^3} \rho^{5/3}\, dx \le C' \sum_j \int_{Q_j} \rho^{5/3} dx$.

PROOF OF THEOREM 11. Again we look at $\sum_{j=1}^N (-\Delta_{x_j} - \gamma\rho^{2/3}(x_j))$ and $H_0 = -\Delta - \gamma\rho^{2/3}(x)$, but this time take $\gamma \gg 1$ and use Theorem 6′ in place of Theorem 1. We obtain a collection $\{Q_j\}$ of disjoint cubes in $R^3$ with the properties:

$$(a) \qquad (\operatorname{Av}_{Q_j} \gamma^r \rho^{2/3r})^{1/r} \ge c(\operatorname{diam} Q_j)^{-2}, \qquad r = 1 + \epsilon.$$

(b) The sum of the negative eigenvalues of $H_0$ is dominated in absolute value by $E = C' \sum_j (\operatorname{diam} Q_j)^{-2}$.

It follows as in the proof of Lieb-Thirring that $\langle H\phi, \phi\rangle \ge -E\|\phi\|^2$ for $\phi$ antisymmetric; taking $\phi = \psi$ and substituting the definition of $\rho$ in terms of

$\psi$, we obtain

$$\|\nabla\psi\|^2 - \gamma\int_{R^3}\rho^{5/3}\,dx \geq -C\sum_j(\operatorname{diam} Q_j)^{-2}.$$

Picking $\gamma$ twice as large as the constant $C$ in (†), we now have

$$\text{(c)}\qquad \int_{R^3}\rho^{5/3}\,dx \leq C''\sum_j(\operatorname{diam} Q_j)^{-2}.$$

Separate the $Q_j$ into

$$\text{TYPE I}: \quad \int_{Q_j}\rho^{5/3}\,dx \geq \mathcal{H}(\operatorname{diam} Q_j)^{-2},$$

$$\text{TYPE II}: \quad \int_{Q_j}\rho^{5/3}\,dx \leq \mathcal{H}(\operatorname{diam} Q_j)^{-2}.$$

By (c) we have

$$C''\sum_j(\operatorname{diam} Q_j)^{-2} \geq \sum_{\text{TYPE I}}\int_{Q_j}\rho^{5/3}\,dx \geq \mathcal{H}\sum_{\text{TYPE I}}(\operatorname{diam} Q_j)^{-2}.$$

Taking $\mathcal{H} > 2C''$, we get

$$\sum_{\text{TYPE II}}(\operatorname{diam} Q_j)^{-2} \geq \frac{1}{2}\sum_j(\operatorname{diam} Q_j)^{-2}.$$

Therefore we may simply drop all the TYPE I cubes from our collection $\{Q_j\}$, and properties (a), (b), (c) still hold. We now also have

$$\text{(d)}\qquad \int_{Q_j}\rho^{5/3}\,dx \leq c''(\operatorname{diam} Q_j)^{-2}.$$

Now (a) and Hölder's inequality, together with (d), yields

$$\begin{aligned} c(\operatorname{diam} Q_j)^{-2} &\leq (\operatorname{Av}_{Q_j}\rho^{2r/3})^{1/r} \leq (\operatorname{Av}_{Q_j}\rho)^{2/3} \\ &\leq (\operatorname{Av}_{Q_j}\rho^{5/3})^{2/5} \leq C(\operatorname{diam} Q_j)^{-2}. \end{aligned}$$

Therefore

$$(\operatorname{Av}_{Q_j}\rho^{5/3})^{2/5} \sim (\operatorname{Av}_{Q_j}\rho)^{2/3} \sim (\operatorname{diam} Q_j)^{-2},$$

which amounts to (i) and (ii). Also, (ii) and (c) yield (iv). It remains to check (iii). Set

$$E_j = \{x \in Q_j \mid \beta(\operatorname{diam} Q_j)^{-3} \leq \rho(x) \leq \mathcal{H}(\operatorname{diam} Q_j)^{-3}\},$$

$$F_j = \{x \in Q_j \mid \rho(x) > \mathcal{H}(\operatorname{diam} Q_j)^{-3}\},$$

with $\beta \ll 1 \ll \mathcal{H}$ to be picked later. Property (ii) yields

$$C(\operatorname{diam} Q_j)^{-2} \geq \int_{F_j}[(\operatorname{diam} Q_j)^{-3}]^{2/3}\rho(x)\,dx,$$

so $\int_{F_j} \rho\,dx \le C/\mathcal{H}^{2/3}$. Therefore (i) implies

$$c \le \int_{Q_j} \rho\,dx \le \int_{Q_j} [\beta(\operatorname{diam} Q_j)^{-3} + \mathcal{H}(\operatorname{diam} Q_j)^{-3}\chi_{E_j} + \rho\chi_{F_j}]\,dx$$

$$\le c_n\beta + c_n\mathcal{H}|E_j|/|Q_j| + C/\mathcal{H}^{2/3}.$$

For $\beta$ small enough and $\mathcal{H}$ large enough, this gives $c_n\mathcal{H}|E_j|/|Q_j| \ge c/2$, which is (iii). Q.E.D.

Theorem 11 is related to atomic structure of matter by virtue of the following observation.

LEMMA D. *If $E_{QM}(\psi) \le -\epsilon N$, then $\|\nabla\psi\|^2 \sim \int_{R^3} \rho^{5/3}\,dx \sim N$.*

PROOF. We know already that $\|\nabla\psi\|^2 \ge c\int_{R^3} \rho^{5/3}\,dx$. Let us show $\|\nabla\psi\|^2 \le \gamma\int_{R^3} \rho^{5/3}\,dx$ for a large constant $\gamma$. If this were false, Lemmas B and C in the proof of Theorem 8, with $\mathcal{H} = \beta = \gamma/2$, yield

$$E_{QM}(\psi) \ge \left[\gamma\int_{R^3} \rho^{5/3}\,dx\right] + \left[\frac{1}{2}\int \frac{\rho(x)\rho(y)}{|x-y|}\,dx\,dy - \frac{\gamma}{2}\int \rho^{5/3}\,dx - C\left(\frac{\gamma}{2}\right)N\right]$$

$$+ \sum_{j<k} \frac{z_j z_k}{|y_j - y_k|} - \int_{R^3} \rho(x)\sum_k \frac{z_k}{|x-y_k|}\,dx$$

$$\ge -2C(\gamma/2)N.$$

Recall that $C(\gamma)$ is the T-F binding energy of an atom; so $C(\gamma) \to 0$ as $\gamma \to \infty$. Hence for large $\gamma$ we contradict $E_{QM}(\psi) < -\epsilon N$. So $\|\nabla\psi\|^2 \sim \int \rho^{5/3}\,dx$.

A similar argument shows that $\int \rho^{5/3}\,dx \sim N$. Already the proof of Theorem 8(B) gives $\int \rho^{5/3}\,dx \le CN$. If $\int \rho^{5/3}dx \le \delta N$ for a small constant $\delta$, then Lemmas B and C with $\beta = \mathcal{H} \gg 1$ imply

$$E_{QM}(\psi) \ge \left[\frac{1}{2}\int \frac{\rho(x)\rho(y)}{|x-y|}\,dx\,dy - \mathcal{H}\int \rho^{5/3}\,dx - C(\mathcal{H})N\right]$$

$$+ \sum_{j<k} \frac{z_j z_k}{|y_j - y_k|} - \int \rho(x)\sum_k \frac{z_k}{|x-y_k|}\,dx$$

$$= \left[\mathcal{H}\int \rho^{5/3}\,dx + \frac{1}{2}\int \frac{\rho(x)\rho(y)}{|x-y|}\,dx\,dy\right.$$

$$\left. + \sum_{j<k} \frac{z_j z_k}{|y_j - y_k|} - \int \rho(x)\sum_k \frac{z_j}{|x-y_k|}\,dx\right]$$

$$- 2\mathcal{H}\int \rho^{5/3}\,dx - C(\mathcal{H})N$$

$$\ge -2\mathcal{H}\int \rho^{5/3}\,dx - 2C(\mathcal{H})N \ge -2[\mathcal{H}\delta + C(\mathcal{H})]N.$$

Picking first $\mathcal{H} \gg 1$, then $\delta \ll \mathcal{H}^{-1}$, we again contradict $E_{QM}(\psi) < -\epsilon N$. So $\|\nabla\psi\|^2 \sim \int \rho^{5/3}\,dx \sim N$. Q.E.D.

Now Lemma D and Theorem 11 yield a collection of pairwise disjoint cubes $\{Q_j\}$ with the properties:

($\alpha$) $\int_{Q_j} \rho\, dx \sim 1$;

($\beta$) $\int_{Q_j} \rho^{5/3}\, dx \sim (\operatorname{diam} Q_j)^{-2}$;

($\gamma$) $\rho \sim (\operatorname{diam} Q_j)^{-3}$ on a fixed part of the measure of $Q_j$;

($\delta$) $\sum_j (\operatorname{diam} Q_j)^{-2} \sim N$.

If all the cubes $Q_j$ had diameter $\sim 1$, then ($\delta$) would show there are at least $c_1 N$ cubes, while ($\alpha$), ($\beta$), ($\gamma$) would become the conditions defining atomic structure (except for the nucleii). Now it is easy to get all the diameters of the $Q_j \leq C_1$. In fact, we simply throw out all the $Q_j$ with diameters larger than $C_1$. Properties ($\alpha$)–($\gamma$) are of course preserved, so we need only check that ($\delta$) still holds. From ($\alpha$) we see that there are at most $CN$ cubes altogether, so the contribution to the sum in ($\delta$) arising from the discarded cubes is at most $CN \cdot (C_1)^{-2}$. Taking $C_1$ large enough, we see that ($\delta$) still holds for the remaining cubes. So the diameters of the $Q_j$ are not too big. However, they might easily be too small. Let us look at a simple example: Imagine $k \ll N$ protons and $k$ electrons packed together inside a small cube. Their total quantum-mechanical energy can be taken to be $\sim +cN$ for a small constant $c$. To these particles we now add N-K hydrogen atoms very far from one another and from the original $2k$ particles. The result is a system satisfying $E_{QM}(\psi) < -c'N$, yet we can construct cubes $Q_j$ satisfying ($\alpha$)–($\gamma$) in terms of the $2k$ tightly packed particles, ignoring the real atoms. So ($\alpha$)–($\gamma$) do not automatically yield $\operatorname{diam} Q_j \geq c$.

## CHAPTER III: DECOMPOSITIONS OF PHASE SPACE

So far we know how to cut phase space into blocks $\{B_j\}$ as in Chapter I, Figure 3. Now we shall look inside the blocks $B_j$. To get a deep understanding of a symbol $A(x, \xi)$ on $B_j$, we shall make a Calderón-Zygmund decomposition of $B_j$ into pieces $\{Q_\nu\}$ as in Figure 1.

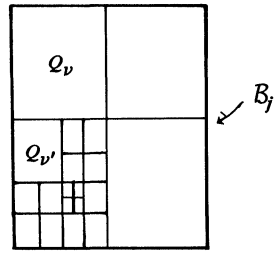



FIGURE 1

The $Q_\nu$ have large volume, so we aren't yet close to diagonalizing $A(x, D)$, but the technique is sharp enough to give interesting applications. In particular, we can determine the lowest eigenvalue of a pseudodifferential operator, prove sufficiency of the Nirenberg-Trèves condition ($\mathcal{P}$) for local solvability, and prove hypoellipticity with a sharp estimate for sums of squares of vector fields. Let us begin by stating these applications precisely.

To find the lowest eigenvalue, we will work with a nonnegative symbol $A(x,\xi)$ satisfying the second-order estimates

$$|\partial_x^\alpha \partial_\xi^\beta A| \le C_{\alpha\beta} M^{2-|\beta|} \quad (\text{we say } A \in S^2(1 \times M)).$$

This is the estimate satisfied by a classical second-order symbol on a block $|x| \le 1$, $|\xi| \sim M$ in Chapter I, Figure 3. According to the *SAK* principle, the eigenvalues of $A(x,D)$ should be determined by a covering $\{\mathcal{B}_\nu\}$ of phase space by distorted unit boxes $\mathcal{B}_\nu$. In fact, we expect that the number of eigenvalues $< K$ for $A(x,D)$ will be comparable to the number of $\mathcal{B}_\nu$ contained in $S(A,K) = \{(x,\xi) \in R^{2n} \mid A(x,\xi) \le K\}$. This leads us to predict that the lowest eigenvalue $\lambda_1(A)$ for $A(x,D)$ is comparable to $\min_\nu \max_{(x,\xi)\in\mathcal{B}_\nu} A(x,\xi)$. We should be a little more specific about what $\mathcal{B}_\nu$ looks like.

DEFINITION. Let $\Phi\colon (z,\zeta) \to (x,\xi)$ be a canonical transformation mapping $|z|, |\zeta| \le M^\epsilon$ into $R^{2n}$ and satisfying the estimates

$$|\partial_{z,\zeta}^\alpha x| \le M^{-\epsilon|\alpha|}, \quad |\partial_{z,\zeta}^\alpha \xi| \le M^{1-\epsilon|\alpha|} \qquad (|\alpha| \ge 1).$$

Then the image $\mathcal{B} = \Phi(Q^0)$ of the unit cube $Q^0$ is called a *testing box*.

THEOREM 1. *To a symbol $A \in S^2(1 \times M)$, $A \ge 0$, we can associate a family of testing boxes $\{\mathcal{B}_\nu\}$ so that*

$$\lambda_1(A) = \textit{lowest eigenvalue of } A(x,D) \quad \textit{and}$$
$$\mu_1(A) = \min_\nu \max_{(x,\xi)\in\mathcal{B}_\nu} A(x,\xi)$$

*are related by*

$$\lambda_1(A) \le C_\epsilon \mu_1(A) + C_\epsilon M^{2\epsilon}, \qquad \mu_1(A) \le C_\epsilon \lambda_1(A) + C_\epsilon M^{2\epsilon}.$$

*Thus $\lambda_1(A)$ and $\mu_1(A)$ are comparable unless both are $\le M^\epsilon$.*

Similar results hold for the $N$th eigenvalue; see [**15**]. From Theorem 1 we can very easily deduce hypoellipticity of sums of squares of vector fields. In fact, we have the following sharp estimate.

THEOREM 2. *Suppose vector fields $X_1, \dots, X_N$ and their commutators of order $\le m$ span the tangent space at every point. Then for $L = \sum_j X_j^2$ we have*

$$C\|u\|^2 + C\,\mathrm{Re}\langle Lu, u\rangle \ge \|u\|^2_{1/(m+1)}. \tag{1}$$

This implies easily that $Lu \in H^s_{\mathrm{loc}}(\Omega)$ implies $u \in H^{s+(2/(m+1))}_{\mathrm{loc}}(\Omega)$, so $L$ is hypoelliptic. The estimate (1) with $1/(m+1)$ replaced by $1/(m+1) - \epsilon$ goes back to Hörmander [**22**]; the present sharp form is much harder and is due to Rothschild and Stein [**37**].

Regarding local solvability, we work with a partial differential operator $L(x,D)$ whose principal symbol $p + iq$ is of *principal type*, i.e., $p + iq$ and $\mathrm{grad}(p+iq)$ do not vanish together at any point $(x,\xi)$, $\xi \ne 0$. This hypothesis

ensures that lower-order terms in $L(x,D)$ can be regarded as insignificant perturbations. Thus the Laplace equation and wave equation are of principal type, but the heat equation and Schrödinger equation are not. At a point $(x^0,\xi^0)$ where $p+iq=0$ and (say) $\nabla p\neq 0$, we form the *null bicharacteristic of* $p$ through $(x^0,\xi^0)$, which is the orbit of $(x^0,\xi^0)$ under Hamilton's equations

$$dx_k/dt=\partial p/\partial\xi_k,\qquad d\xi_k/dt=-\partial p/\partial x_k.$$

*Condition* $(\mathcal{P})$ asserts that $q$ has constant sign on this null bicharacteristic. That is, either $q\geq 0$ on the whole curve, or $q\leq 0$ on the whole curve.

THEOREM 3. *A partial differential equation of principal type is locally solvable if and only if its symbol satisfies condition* $(\mathcal{P})$.

Calderón-Zygmund decompositions of phase space play a role in proving the sufficiency of $(\mathcal{P})$.

Now let us study the cutting of symbols. To see the ideas we start with the elementary case of a real vector field $X=\sum_{k\geq 1}a_k(x)(\partial/\partial x_k)$ in $|x|\leq 1$. If $|a_1(x)|\geq 1$ for all $x$, then $X$ can be straightened out by a change of variable $y=\phi(x)$: In $y$-coordinates we have $X=\partial/\partial y_1$. A general vector field can be localized into pieces that look like $\partial/\partial y_1$ by virtue of the following simple result.

LEMMA 1. *Let $X$ be a real $C^\infty$ vector field on $|x|\leq 1$. There is a Calderón-Zygmund decomposition of $|x|\leq 1$ into cubes $\{Q_\nu\}$ with the properties:*

(a) *There is a coordinate change $\phi_\nu\colon Q_\nu\to$* unit cube *so that in the new coordinates $X|_{Q_\nu}$ becomes $\partial/\partial z_1$ on the unit cube. The map $\phi_\nu$ and its derivatives satisfy natural estimates.*

(b) *If $1=\sum_\nu\theta_\nu$ is a partition of unity with $\theta_\nu$ supported essentially in $Q_\nu$, then $X$ commutes with the decomposition operator $u\to(u_\nu)=(\theta_\nu u)$ modulo an error bound on $L^2$.*

So the equation $Xu=f$ is still reduced to $\partial u/\partial z_1=f$, even without assuming $X$ can be globally straightened out.

PROOF. Bisect $\{|x|\leq 1\}$ repeatedly into subcubes, stopping at a cube $Q$ if $\max_{x\in Q}\max_k|a_k(x)|\geq 2C(\operatorname{diam}Q)$.

Here the large constant $C$ is picked so that $|\nabla a_k(x)|\leq C$ for $|x|\leq 1$. Now the unit cube is cut up into subcubes $\{Q_\nu\}$. On each $Q_\nu$ we have

$$|a_{k_0}(x_0)|\geq 2C(\operatorname{diam}Q)\quad\text{for an } x_0\in Q_\nu \text{ and } 1\leq k_0\leq n. \tag{2}$$

Also,

$$|a_k(x)|\leq 4C(\operatorname{diam}Q),\qquad x\in Q_\nu, \tag{3}$$

for otherwise the cutting would have stopped before we reached $Q_\nu$.

In each $Q_\nu$ we make the change of scale $y=(x-y_0)/(\operatorname{diam}Q_\nu)$. Thus $Q_\nu$ goes over to the unit cube, while $X=\sum_k\tilde{a}_k(y)(\partial/\partial y_k)$ with $\tilde{a}_k(y)=(\operatorname{diam}Q_\nu)^{-1}a_k(x_0+(\operatorname{diam}Q_\nu)y)$. Now $\tilde{a}_k\in C^\infty$(unit cube) with a priori bounds on $\tilde{a}_k$ and all its derivatives. In fact, $|\partial_y^\alpha\tilde{a}_k|=(\operatorname{diam}Q_\nu)^{|\alpha|-1}|\partial^\alpha a_k|$, which takes care of all cases except $\alpha=0$, i.e., boundedness of $\tilde{a}_k$ itself. For this

we use (3), which immediately gives $|\tilde{a}_k| \leq 4C$. So $X|_{Q_\nu}$ is a nice vector field on the new scale. Moreover, $|\tilde{a}_{k_0}| \geq C$ throughout the unit cube. The reason is that (2) and $|\nabla a_{k_0}| \leq C$ yield $|a_{k_0}(x)| \geq C(\operatorname{diam} Q_\nu)$ for $x \in Q_\nu$, i.e., $|\tilde{a}_{k_0}(x)| \geq C$ for $|x| \leq 1$. Consequently, $X = \sum_k \tilde{a}_k(y)(\partial/\partial y_k)$ can be brought to the form $X = \partial/\partial z_1$ by a change of coordinates $z = \psi(y)$ on the unit cube, so part (a) of the lemma is proved.

Checking part (b) amounts to showing that $X\theta_\nu$ is uniformly bounded, which follows at once from (3) and the natural estimates $|\nabla\theta_\nu| \leq C(\operatorname{diam} Q_\nu)^{-1}$. Q.E.D.

Notice that it was essential to have $\{Q_\nu\}$ fine exactly where $X$ looks degenerate.

The next simplest problem is to understand a second-order equation $L = -\sum_{jk} a_{jk}(x)(\partial^2/\partial x_j \partial x_k)$, lower-order terms, $(a_{jk}(x)) \geq 0$ real. To make the discussion easy we will ignore the lower-order errors. We call $L$ *nondegenerate* if $a_{11}(x) \geq 1$ for $|x| \leq 1$.

LEMMA 2. *A nondegenerate equation may be placed in the following special form by a change of variable*:

$$L = -\left(\frac{\partial}{\partial y_1}\right)^2 - \sum_{j,k\geq 2} b_{jk}(y_1, y')\frac{\partial^2}{\partial y_j \partial y_k} + \text{lower-order terms.} \tag{4}$$

This may be thought of as a Schrödinger operator $-(\partial/\partial y_1)^2 + V(y_1)$, where $V(y_1)$ is an operator in fewer variables.

PROOF. For simplicity we will show only a slightly weaker result, namely that $eL$ can be written in the form (4) for a smooth nonvanishing factor $e$. Take $e = 1/a_{11}(x)$ and write

$$-eL = \sum_{jk} \frac{a_{jk}(x)}{a_{11}(x)} \frac{\partial^2}{\partial x_j \partial x_k} = \left(\frac{\partial}{\partial x_1} - \sum_{k\geq 2} \frac{a_{1k}}{a_{11}} \frac{\partial}{\partial x_k}\right)^2 + \sum_{j,k\geq 2} b_{jk} \frac{\partial^2}{\partial x_j \partial x_k}$$

$$= Y^2 + \sum_{j,k\geq 2} b_{jk}(x) \frac{\partial^2}{\partial x_j \partial x_k}$$

with

$$Y = \frac{\partial}{\partial x_1} - \sum_{k\geq 2} \frac{a_{jk}}{a_{11}} \frac{\partial}{\partial x_k}.$$

Define $\Phi\colon (x_1, x') \to (y_1, y')$ by $y_1 = x_1$, $(0, y') = \exp(-x_1 Y)(x_1, x')$. In $y$-coordinates we have

$$Y = \frac{\partial}{\partial y_1}, \qquad \sum_{j,k\geq 2} b_{jk}(x) \frac{\partial^2}{\partial x_j \partial x_k} + \cdots = \sum_{j,k\geq 2} \tilde{b}_{jk}(y) \frac{\partial}{\partial y_j \partial y_k} + \cdots,$$

so $eL$ has been placed in the form (4). Q.E.D.

Now a general second-order equation can be cut into nondegenerate pieces as in the case of vector fields.

LEMMA 3. *Given* $L = -\sum_{jk} a_{jk}(x)\partial^2/\partial x_j \partial x_k + \cdots$ *with* $(a_{jk}) \geq 0$, *we can cut the unit cube* $|x| \leq 1$ *into Calderón-Zygmund cubes* $\{Q_\nu\}$ *so that:*

(a) *On each* $Q_\nu$ *there is a change of variable which carries* $Q_\nu$ *to the unit cube and places* $L$ *in the form* (4).

(b) *If* $1 = \sum_\nu \theta_\nu^2$ *is a partition of unity with* $\theta_\nu$ *supported essentially in* $Q_\nu$, *then* $L = \sum_\nu \theta_\nu L \theta_\nu +$ (*bounded error*).

So the study of general $L$ in $n$ variables is reduced by cutting and bending to the case of a Schrödinger operator whose potential is a differential operator in $(n-1)$ variables.

PROOF. Bisect $|x| \leq 1$ repeatedly into subcubes, stopping at $Q$ as soon as

$$\max_{j,k} \max_{x \in Q} |a_{jk}(x)| \geq 20C(\operatorname{diam} Q)^2. \tag{5}$$

Here $C$ is a large constant depending on

$$\max_{j,k} \max_{|\beta| \leq 2} \max_{|x| \leq 1} |\partial^\beta a_{jk}(x)|.$$

Thus, $|x| \leq 1$ is cut into cubes $\{Q_\nu\}$, and we know

$$|a_{jk}(x)| \leq 80C(\operatorname{diam} Q_\nu)^2 \quad \text{for } x \in Q_\nu, \tag{6}$$

otherwise we would have stopped cutting before reaching $Q_\nu$. Also for some $l$ we have

$$a_{ll}(x) \geq C(\operatorname{diam} Q_\nu)^2 \quad \text{for } x \in Q_\nu. \tag{7}$$

This is slightly harder than the analogue for vector fields. We need the standard

REMARK. Assume $f \geq 0$ and $|f''| \leq 1$ in $|x| \leq 1$, while $f(x^0) = \delta^2 \ll 1$. Then $f(x) \geq \delta^2/8$ for $|x - x^0| < \delta/2$.

Let us check the Remark. If $f(x) < \delta^2/8$, $|x - x^0| < \delta/2$, then $f(x+y) \leq f(x+y)+f(x-y) = F(y)$. We have $F(0) < \delta^2/4$, $F'(0) = 0$, $|f''| \leq 2$ everywhere, so $F(y) \leq \delta^2/4 + |y|^2$. Taking $y = x^0 - x$, we find that

$$f(x^0) \leq F(x^0 - x) \leq \delta^2/4 + |x^0 - x|^2 \leq \delta^2/4 + \delta^2/4 < \delta^2,$$

contradicting the hypothesis.

Returning to (7), we know by definition (5) of $Q_\nu$ that

$$a_{ll}(x^0) \geq 20C(\operatorname{diam} Q_\nu)^2 \quad \text{for some } x \in Q^0.$$

The reason is that $(a_{jk}(x^0)) \geq 0$, hence $\max_{jk}|a_{jk}(x^0)| = \max_l a_{ll}(x^0)$. Now taking $f = a_{ll}$ and $\delta = (20C)^{1/2} \operatorname{diam} Q_\nu$, we obtain (7) as a consequence of the Remark.

We make the change of scale $y = (x - x_0)/(\operatorname{diam} Q_\nu)$, which carries $Q_\nu$ to the unit cube and puts $L$ in the form

$$\sum_{jk} \tilde{a}_{jk}(y) \frac{\partial^2}{\partial y_j \partial y_k} + \cdots \quad \text{with } \tilde{a}_{jk}(y) = (\operatorname{diam} Q_\nu)^{-2} \cdot a_{jk}(x).$$

We can check that $\tilde{a}_{jk} \in C^\infty$(unit cube) with a priori bounds. In fact

$$|\partial_y^\alpha \tilde{a}_{jk}| = (\operatorname{diam} Q_\nu)^{|\alpha|-2} |\partial_x^\alpha a_{jk}|,$$

which takes care of all $\alpha$ except $|\alpha| = 0, 1$. The estimate $|\tilde{a}_{jk}| \leq$ Const. corresponding to $\alpha = 0$ follows at once from (6). To handle the case $|\alpha| = 1$, we first write $\tilde{a}_{jk}(y) - \tilde{a}_{jk}(0) = y \cdot \nabla \tilde{a}_{jk}(0) + O(1)$ for $|y| \leq 1$, by virtue of our estimates for the second derivatives of $\tilde{a}_{jk}$. Since the left-hand side is also $O(1)$, we have $|\nabla \tilde{a}_{jk}(0)| \leq$ Const. Again using our estimates on the second derivatives, we get $|\nabla \tilde{a}_{jk}(y)| \leq$ Const., all $|y| \leq 1$, which is the missing case $|\alpha| = 1$. Equation (7) now shows that $L = \sum_{jk} \tilde{a}_{jk}(y)(\partial^2/\partial y_j \partial y_k) + \cdots$ satisfies the hypotheses of Lemma 2. So part (a) of Lemma 3 is proved.

To check (b) we just note that $L - \sum_\nu \theta_\nu L \theta_\nu$ has as its worst term

$$\sum_\nu \sum_{jk} \theta_\nu a_{jk} \left( \frac{\partial^2 \theta_\nu}{\partial x_j \partial x_k} \right).$$

Since $|a_{jk}| \leq C(\operatorname{diam} Q_\nu)^2$ and $|\partial^2 \theta / \partial x_j \partial x_k| \leq C(\operatorname{diam} Q_\nu)^{-2}$ on $Q_\nu$, this term is bounded. Q.E.D.

Again there is an essential balance between the fineness of $\{Q_\nu\}$ and the degeneracy of $L$.

Now we shall analyze general pseudodifferential operators by methods analogous to Lemmas 1–3. We start with a symbol $A(x, \xi)$ defined on a block $|x| \leq 1$, $|\xi| \leq M$. The symbol is said to belong to $S^m\{|x| \leq 1, |\xi| \leq M\} = S^m(1 \times M)$ if we have $|\partial_x^\alpha \partial_\xi^\beta A| \leq C_{\alpha\beta} M^{m-|\beta|}$.

These are the estimates satisfied by a classical $m$th order symbol on the blocks of Chapter I, Figure 3. A symbol $A \in S^m(1 \times M)$ is called *$m$th order elliptic* if $|A(x, \xi)| \geq cM^m$. We can use standard $\psi d0$ calculus, including Egorov's theorem, to manipulate symbols in $S^m(1 \times M)$.

First we have to understand what symbols look like in the nondegenerate case.

LEMMA 4. *Let $A(x, \xi) \in S^1(1 \times M)$ be a real symbol satisfying*

$$\max_{|\alpha|+|\beta| \leq 1} \max_{|x|, M^{-1}|\xi| \leq 1} \frac{|\partial^\alpha \partial_\xi^\beta A|}{M^{1-|\beta|}} \geq C. \tag{8}$$

*Then either $A$ is first-order elliptic or, by a canonical transformation $\Phi\colon (x, \xi) \to (y, \eta)$, as in Egorov's theorem, $A$ can be put in the form $A \circ \Phi^{-1}(y, \eta) = \eta_1$.*

This corresponds to straightening a vector field.

LEMMA 5. *Let $A(x, \xi) \geq 0$ belong to $S^2(1 \times M)$ and assume*

$$\max_{|\alpha|+|\beta|=2} \max_{|x|, M^{-1}|\xi| \leq 1} \frac{|\partial_x^\alpha \partial_\xi^\beta A|}{M^{2-|\beta|}} \geq C. \tag{9}$$

*Then either $A$ is second-order elliptic or, by a canonical transformation $\Phi\colon (x, \xi) \to (y, \eta)$, as in Egorov's theorem, we may bring $A$ to the form $A \circ \Phi^{-1}(y, \eta) = \eta_1^2 + V(y_1, y', \eta')$. Thus, $A \circ \Phi^{-1}(y, D) = -(\partial/\partial y_1)^2 + V(y_1)$, where $V(y_1)$ is a second-order pseudodifferential operator in fewer variables.*

The proofs of these lemmas are essentially no harder than their analogues for vector fields and second order PDE. The ideas are on the level of the implicit function theorem.

Our plan is to cut phase space into blocks $\{Q_\nu\}$ on which Lemmas 4 and 5 apply. To do this we have to answer a simple question. Given a symbol $A \in S^m(1 \times M)$ and a block $Q$ of sides $\delta \times M\delta$ as in Figure 2, when is it reasonable to localize $A$ to $Q$?

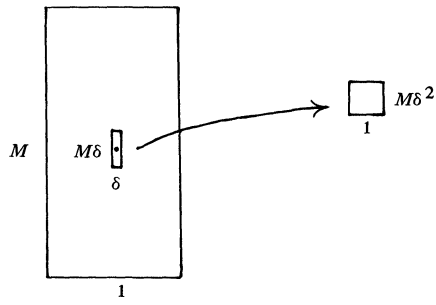



FIGURE 2

If $Q$ is centered at $(x^0, \xi^0)$, then a natural change of scale $\iota\colon (x, \xi) \to (y, \eta)$ is $y = (x - x_0)/\delta$, $\eta = (\xi - \xi^0)\delta$. Thus $Q$ goes over to the block $|y| \leq 1$, $|\eta| \leq M\delta^2$. It makes sense to study $A$ localized to $Q$ if $A \circ \iota^{-1} \in S^m(1 \times \tilde{M})$ on $|y| \leq 1$, $|\eta| \leq \tilde{M} = M\delta^2$. This means $A|_Q$ satisfies estimates

$$|\partial_x^\alpha \partial_\xi^\beta A| \leq C_{\alpha\beta}(M\delta)^{m-|\beta|}(|\delta|)^{m-|\alpha|}. \tag{10}$$

If (10) holds for $(x, \xi) \in Q$ of sides $\delta \times M\delta$, then we say $A$ belongs to $S^m(Q)$.

When do the estimates (10) hold on $Q$? For $|\alpha| + |\beta| \geq 2m$, they already follow from the fact that $A \in S^m(1 \times M)$. For $|\alpha| + |\beta| < 2m$, this is no longer the case, and they hold for small $Q$ only if $A$ looks rather degenerate around $Q$.

These remarks suggest how to cut up $\{|x| \leq 1, |\xi| \leq M\}$. We bisect this block repeatedly into smaller blocks, stopping at $Q$ of sides $\delta \times M\delta$ if either

$$\max_Q \max_{|\alpha|+|\beta|<2m} \frac{|\partial_x^\alpha \partial_\xi^\beta A|}{(M\delta)^{m-|\beta|}\delta^{m-|\alpha|}} \geq C \tag{11}$$

or $\mathrm{Vol}(Q) \sim 1$. (Certainly if $\mathrm{Vol}(Q) \sim 1$ we had better stop cutting, in view of the uncertainty principle.) Thus $\{|x|, M^{-1}|\xi| \leq 1\}$ is cut into blocks $\{Q_\nu\}$ of sides $\delta_\nu \times M\delta_\nu$. We have $A \in S^m(Q_\nu)$ since otherwise we would have stopped cutting before reaching $Q_\nu$. In particular, $A$ is bounded on the $Q_\nu$ of volume $\sim 1$. On the $Q_\nu$ of volume $\gg 1$, we make the change of scale $\iota$, and (11) then gives a kind of nondegeneracy for $A \circ \iota^{-1}$.

Specializing to the cases $m = 1, 2$ and invoking Lemmas 4 and 5 now yield

LEMMA 6. *Let $A \in S^1(1 \times M)$ and let $Q_\nu$ be one of the Calderón-Zygmund blocks with volume $\gg 1$. Either $A$ is elliptic on $Q_\nu$ or else a canonical transformation $\Phi$ brings $A|_{Q_\nu}$ to the form $A \circ \Phi^{-1}(y, \eta) = \eta_1$.*

LEMMA 7. *Let $A \in S^2(1 \times M)$ be nonnegative and let $Q_\nu$ be one of the Calderón-Zygmund blocks with volume $\gg 1$. Either $A$ is elliptic on $Q_\nu$ or else a canonical transformation $\Phi$ brings $A|_{Q_\nu}$ to the special form $A \circ \Phi^{-1}(y, \eta) = \eta_1^2 + V(y_1, y', \eta')$.*

Although we omit details, it is worth mentioning one point in the proof of Lemma 7. The definition (11) of the $Q_\nu$ shows that $\partial_x^\alpha \partial_\xi^\beta A|_{Q_\nu}$ must be rather large for some $|\alpha| + |\beta| \leq 3$. On the other hand, $A$ is nonnegative, so the even-order terms in its Taylor expansion must dominate the odd-order terms. Hence $\partial_x^\alpha \partial_\xi^\beta A|_{Q_\nu}$ will be rather large either for $|\alpha| + |\beta| = 0$ or 2. This leads to the dichotomy between the elliptic case and the case where we apply Lemma 5.

In Figures 3 and 4 we have shown how first and second order symbols break up according to Lemmas 6 and 7.

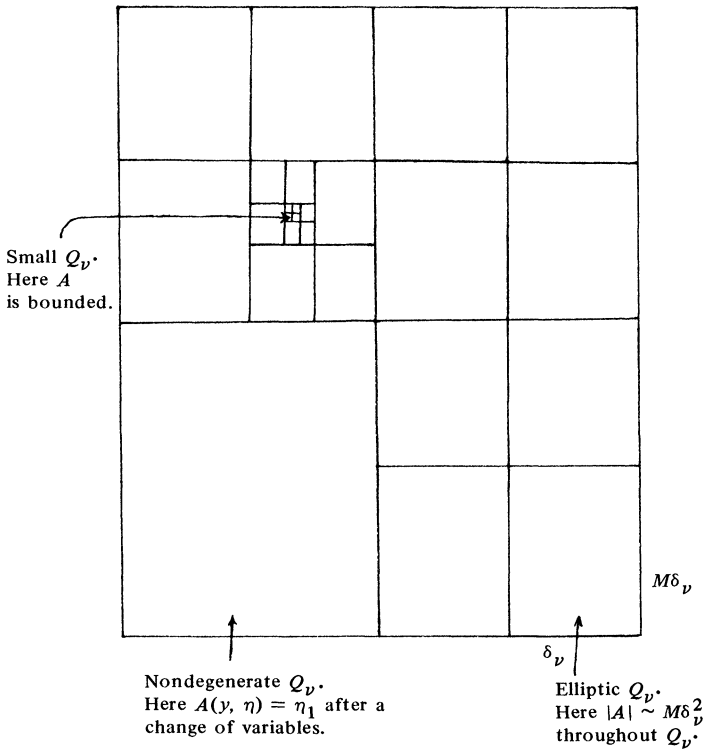



FIGURE 3. Analysis of a first-order symbol

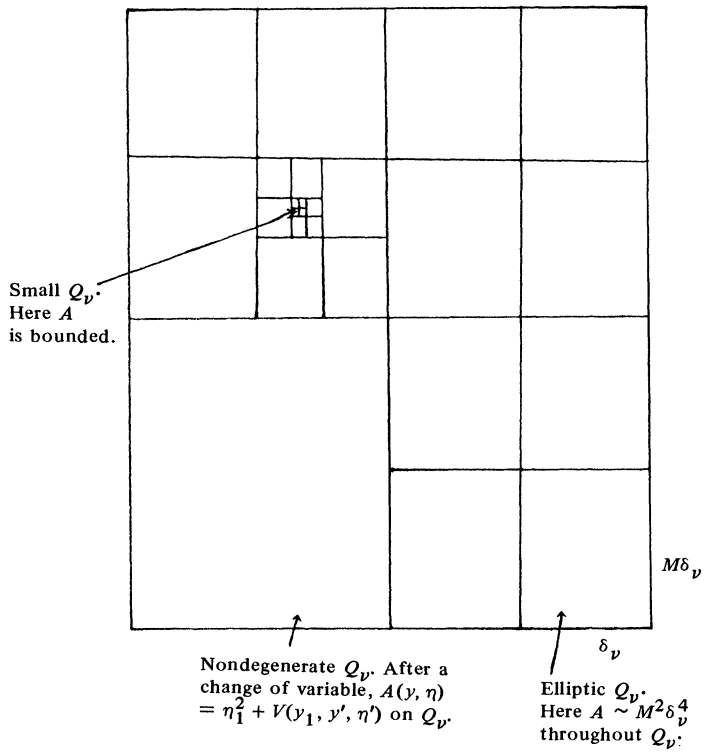



FIGURE 4. Analysis of a second-order operator

Now we can break up symbols $A(x, \xi)$ as in Figures 3 and 4. For this to be useful, we have to show that the operator $A(x, D)$ breaks up correspondingly as an approximate direct sum. The best way to do this is to introduce the

*Beals calculus of $\psi d0$.* Let $\{Q_\nu\}$ be the blocks arising from a symbol $A \in S^m(1 \times M)$. We say that a symbol $p(x, \xi)$ belongs to $S^\mu\{Q_\nu\}$ if $p|_{Q_\nu} \in S^\mu(Q_\nu)$ with seminorms bounded uniformly in $\nu$. In particular, we saw that $A \in S^m\{Q_\nu\}$. By analogy with ordinary $\psi d0$ calculus, we have

THEOREM 4 (SEE [2]). *If $p \in S^\mu\{Q_\nu\}$ and $q \in S^{\mu'}\{Q_\nu\}$, then $p(x, D)q(x, D) = p \circ q(x, D)$*

(a) *with*

$$p \circ q = \sum_{|\alpha|<N} \frac{1}{\alpha!}\left(\frac{1}{i}\frac{\partial}{\partial \xi}\right)^\alpha p\left(\frac{\partial}{\partial x}\right)^\alpha q \mod S^{\mu+\mu'-N}\{Q_\nu\}.$$

*Also*

(b) $[p(x,D)]^* = p^{\#}(x,D)$ *with*

$$p^{\#} = \sum_{|\alpha|<N} \frac{1}{\alpha!}\left(\frac{1}{i}\frac{\partial}{\partial\xi}\right)^{\alpha}\left(\frac{\partial}{\partial x}\right)\overline{p} \mod S^{\mu-N}\{Q_\nu\}.$$

(c) *If* $p \in S^0\{Q_\nu\}$, *then* $p(x,D)$ *is bounded on* $L^2$.

The proofs of (a) and (b) basically just repeat the familiar arguments for classical symbols, but (c) is deeper than its classical analogue. We can restate it in more familiar language as

THEOREM 5 (CALDERÓN AND VAILLANCOURT [4]). *If* $p(x,\xi)$ *satisfies* $|\partial_x^\alpha \partial_\xi^\beta p(x,\xi)| \le C_{\alpha\beta} M^{(|\alpha|-|\beta|)/2}$, *then* $p(x,D)$ *is bounded on* $L^2$.

This is proved by cutting $p(x,D) = \sum_\nu (\phi_\nu p)(x,D) \equiv \sum_\nu A_\nu$, where $1 = \sum_\nu \phi_\nu$ and $\phi_\nu$ is supported in a box of sides $M^{-1/2} \times M^{1/2}$. Each fixed $A_\nu$ is "almost orthogonal" to the other $A_{\nu'}$ in the sense that

$$\begin{aligned} &\sum_{\nu'} \|A_{\nu'}^* A_\nu\|^{1/2} \le C, \quad \sum_{\nu'} \|A_\nu^* A_{\nu'}\|^{1/2} \le C, \\ &\sum_{\nu'} \|A_{\nu'} A_\nu^*\|^{1/2} \le C, \quad \sum_{\nu'} \|A_\nu A_{\nu'}^*\|^{1/2} \le C. \end{aligned} \tag{12}$$

These estimates hold because $A_\nu f(x) = \int K_\nu(x,y) f(y)\, dy$ with distinct $K_\nu$ essentially having either disjoint support or Fourier transforms with disjoint support.

A beautiful lemma by Cotlar and Stein says that abstract operators $A_\nu$ which satisfy (12) have a bounded sum. The model case is $A_\nu$ =projection from $H = \bigoplus_\alpha H_\alpha$ into $H_\nu$. Details may be found in [**11**]. The story is by now well known.

Taking Theorems 4 and 5 for granted, we can use Beals calculus to pass from analysis of symbols (Lemmas 6 and 7) to decompositions of operators. The point is that the functions $\phi_\nu$ from a partition of unity $1 = \sum_\nu \phi_\nu^2$, $\phi_\nu$ supported essentially in $Q_\nu$, belong uniformly to $S^0\{Q_\nu\}$. Hence we can write, e.g., for $A \in S^2$,

$$A(x,D) = \sum_\nu \phi_\nu(x,D)^*[\chi_\nu A(x,D)]\phi_\nu(x,D) + \mathcal{E}, \qquad \chi_\nu = 1 \text{ on } \operatorname{supp}\phi_\nu. \tag{13}$$

It is easiest to think of $(x,\xi) \to (\phi_\nu(x,\xi))$ as a vector-valued symbol in $S^0\{Q_\nu\}$. Beals calculus shows at once that $\mathcal{E} = e(x,D)$ with $e \in S^0\{Q_\nu\}$; in particular, $\mathcal{E}$ is bounded on $L^2$.

So $A(x,D)$ is effectively broken into pieces microlocalized to the $Q_\nu$. Each piece is either trivial (bounded, or else elliptic $\sim M^2\delta_\nu^4$) or else looks like $-(\partial/\partial y_1)^2 + V(y_1)$, with $V(y_1)$ a $\psi d0$ in fewer variables. Similarly, first order $\psi d0$ are cut into simple pieces (bounded, elliptic $\sim \pm M\delta_\nu^2$, or $\partial/\partial y_1$). At least we are in a position to study $\psi d0$ using induction on the dimension.

Returning now to the applications, we first apply our cutting technique to prove Theorem 1. The first order of business is to check that $A(x,D)$ is positive.

PROPOSITION. *If $A \geq 0$ is a second-order symbol, then $A(x,D) \geq -C$.*

This is proved by induction on the dimension. Given a symbol $A$ in $n$ dimensions, we apply Lemma 7 and formula (13). From (13) we see that it is enough to show that $(\chi_\nu A)(x,D) \geq -C$. However, on supp $\chi_\nu$ we know what $A(x,\xi)$ looks like. Either it is bounded or elliptic $\sim M^2\delta_\nu^4$ (in which case positivity of $\chi_\nu A(x,D)$ is obvious) or else $A$ looks like $-(\partial/\partial y_1)^2 + V(y_1)$, with $V(y_1)$ a second-order $\psi d0$ with nonnegative symbol. Assuming our Proposition holds in $(n-1)$ dimensions, we get $V(y_1) \geq -C$ for each $y_1$, so obviously $-(\partial/\partial y_1)^2 + V(y_1) \geq -C$. The Proposition is proved. Q.E.D.

So far we have estimated $-(\partial/\partial y_1)^2 + V(y_1)$ from below simply by dropping the second derivative. We can do much better. The Main Lemma of Chapter II suggests the following.

MAIN LEMMA. *Let $L = -(\partial/\partial y_1)^2 + V(y_1)$, where $V(y_1) = V(y_1, y', D_{y'})$, and $0 \leq V(y_1, y', \eta') \in S^2(1 \times M)$. Then $L \geq cK \geq M^\epsilon$ if and only if $\mathrm{Av}_{y_1 \in I}\, V(y_1) \geq cK$ for every interval $I$ of length $K^{-1/2}$.*

We certainly know this if $V(y_1)$ is a polynomial scalar potential. The condition $K \geq M^\epsilon$ lets us treat $V(y_1)$ as a polynomial in $y_1$. In fact, Taylor-expanding the symbol $V(y_1, y', \eta')$ on $y_1 \in I$ up to order $d$ leads to an error $O(M^2 K^{-d/2})$. If $K \geq M^\epsilon$ we can pick $d > 2/\epsilon$ and the error will be bounded. So in effect $V(y_1)$ is a polynomial on each $I$ of length $\sim K^{-1/2}$. The fact remains that $V(y_1)$ is not a scalar but a pseudodifferential operator. Our Main Lemma is a hard theorem, not an elementary exercise. Its proof uses induction on the dimension, Lemma 7, and Beals calculus. Details can be found in [13].

Finally we can explain the ideas in the proof of Theorem 1. The main ideas will already be clear in the following easier result, with which we content ourselves here.

COROLLARY 1. *Let $A(x,\xi) \geq 0$ be a symbol in $S^2(1 \times M)$. Then the lowest eigenvalue of $A(x,D)$ satisfies*

$$\lambda_1(A) \geq c_\epsilon \inf_{\mathcal{B}} \max_{(x,\xi)\in\mathcal{B}} A(x,\xi) - C_\epsilon M^{2\epsilon}, \tag{14}$$

*where $\mathcal{B}$ runs over all possible testing boxes.*

PROOF. Set $K = \inf_{\mathcal{B}} \max_{(x,\xi)\in\mathcal{B}} A(x,\xi)$. We have to prove $A(x,D) \geq cK$. We may assume $K \geq M^\epsilon$, since otherwise (14) already follows from the Proposition.

Let us apply Lemma 7 and formula (13). These tools reduce matters to three cases: $A(x,D)$ bounded, $A(x,D)$ elliptic, $A(x,D) = -(\partial/\partial x_1)^2 + V(x_1)$ with $V$ a $\psi d0$-valued potential. The first two cases are trivial, and the third may be analyzed by the Main Lemma. To show that $A(x,D) \geq cK$, it is enough to prove that $\mathrm{Av}_{y_1 \in I} V(y_1) \geq cK$ for any interval $I$ of length $\sim K^{-1/2}$.

To prove this in turn, we assume Corollary 1 holds in $(n-1)$ variables. Therefore it is enough to show that the symbol $\overline{V}(y',\eta') = \mathrm{Av}_{y_1 \in I} V(y_1, y', \eta')$ satisfies

$$\max_{(y',\eta') \in \mathcal{B}'} \overline{V}(y',\eta') \geq c'K \quad \text{for each testing box } \mathcal{B}' \in R^{2(n-1)}. \tag{15}$$

Once this is proved, Corollary 1 follows by induction on the dimension.

However, (15) is quite easy to see. Suppose instead

$$\max_{(y',\eta') \in \mathcal{B}'} V(y',\eta') < c'K \quad \text{for a testing box } \mathcal{B}'.$$

For fixed $(y',\eta') \in \mathcal{B}'$, this means that $y_1 \to V(y_1, y', \eta')$ has average at most $c'K$ on the interval $I$. Since $V$ is essentially a polynomial of bounded degree in $y_1$, it follows that $V(y_1, y', \eta') \leq K/4$ for $y_1 \in I^*$, the double of $I$. Note that $|I^*| = 2K^{-1/2}$. Now we can set up a testing box $\mathcal{B} = \{(y_1, y', \eta_1, \eta') \in R^{2n} \mid y_1 \in I^*, |\eta_1| < K^{1/2}/2, (y',\eta') \in \mathcal{B}\}$ in $R^{2n}$. Evidently

$$\begin{aligned} \max_{\mathcal{B}} A(y,\eta) &= \max_{\mathcal{B}} \{\eta_1^2 + V(y_1, y', \eta')\} = K/4 + \max_{\substack{y_1 \in I^* \\ (y',\eta') \in B}} V(y_1, y', \eta') \\ &\leq K/4 + K/4 = K/2. \end{aligned}$$

This contradicts the definition of $K$ as $\inf_{\text{all } \mathcal{B}} \max_{\mathcal{B}} A$. So (15) must hold, and Corollary 1 is proved. Q.E.D.

Next we come to Theorem 2 on squares of vector fields. As promised, this follows easily from Theorem 1. To prove Theorem 2 we first microlocalize to a basic box $|x| \leq 1$, $|\xi| \sim M$ as in Chapter I, Figure 3. The theorem asserts that $L \geq cM^{2/(m+1)}$. Corollary 1 above reduces this to an estimate on the symbol

$$\max_{(x,\xi) \in \mathcal{B}} L(x,\xi) \geq cM^{2/(m+1)} \quad \text{for any testing box } \mathcal{B}.$$

Now $L = \sum_j p_j^2$, where $p_j$ is the symbol of the vector field $X_j$, and the hypothesis on commutators says that some repeated Poisson bracket is elliptic:

$$|\{p_{j_1}, \{p_{j_2}, \ldots, \{p_{j_{m'-1}}, p_{j_{m'}}\} \ldots\}\}| \geq cM \quad \text{on } |x| \leq 1, |\xi| \sim M, m' \leq m.$$

Also, the testing box arises as the image $\Phi(Q^0)$ of the unit cube under a canonical transformation satisfying "good bounds".

Setting $\overline{p}_j = p_j \circ \Phi$, we can rewrite the hypothesis as

$$|\{\overline{p}_{j_1}, \{\overline{p}_{j_2}, \ldots, \{\overline{p}_{j_{m'-1}}, \overline{p}_{j_{m'}}\} \ldots\}\}| \geq cM \quad \text{on the unit cube.} \tag{16}$$

The repeated Poisson bracket is an $(m'+1)$rst degree polynomial in the derivatives $(\partial^\alpha \overline{p}_j)_{|\alpha| \leq m'}$. So (16) implies

$$\max_j \|\overline{p}_j\|_{C^{m'}(Q^0)} \geq c' M^{1/(m+1)}.$$

This in turn yields the seemingly stronger estimate

$$\max_j \max_{(x,\varsigma) \in Q^0} |\overline{p}_j(z,\varsigma)| \geq c'' M^{1/(m+1)}. \tag{17}$$

The reason is that the estimates for $\Phi$ and $S^1(1\times M)$ for $p_j$ together show that $\overline{p}_j = p_j \circ \Phi$ has its high-order derivatives very small. Hence $\|\overline{p}_j - P_j\|_{C^m(Q^0)} \leq 1$ for polynomials $P_j$ of bounded degree. Since $\|P_j\|_{C^m(Q^0)} \leq C \max_{Q^0}|P_j|$ for polynomials of bounded degree, (17) follows from (16). Recalling that $\overline{p}_j = p_j \circ \Phi$, we rewrite (17) in the form $\max_j \max_B |p_j| \geq cM^{1/(m+1)}$, which is what we had to prove. Q.E.D.

We have seen in the proof of Theorem 2 that it can be easy to show that $S(A,K) = \{A(x,\xi) < K\}$ contains no testing box. The reason is that the canonical transformations $\Phi$ by which we distort the unit cube are restricted by certain estimates. If we drop those estimates and think about imbeddings by arbitrary canonical transformations, we immediately arrive at some interesting and possibly quite difficult questions.

*Question* I. Can the unit cube in $R^4$ be imbedded by a canonical transformation into $T_\delta = \{(x,y,\xi,\eta)|\ |x|,|\xi| < \delta; y,\eta \text{ arbitrary}\}$? Here $\delta \ll 1$.

There is a natural motivation for this question in terms of the uncertainty principle.

*Question* II. Can $T_\delta$ above be imbedded by a canonical transformation into $T_{\delta'}$ with $\delta' < \delta$?

Before leaving Theorems 1 and 2, we set down an important consequence of the Main Lemma which will be used in Chapter IV.

COROLLARY TO MAIN LEMMA. *Let* $L = -(\partial/\partial y_1)^2 + V(y_1)$, *where* $V(y_1) = V(y_1, y', D_{y'})$ *is a polynomial of degree* $\leq d$ *in* $y_1$, *and* $0 \leq V(y_1,y',\eta') \in S^2(1\times M)$. *If* $\overline{V} = \mathrm{Av}_{|y_1|\leq 1} V(y_1) \geq cM^\epsilon$, *then*

$$\langle Lu, u\rangle \geq cM^{\epsilon'}\|u\|^2 \quad \textit{for } u(y_1,y') \textit{ supported in } |y_1| \leq 1.$$

PROOF. According to the Main Lemma, it is enough to show that $\overline{V}_I = \mathrm{Av}_{y_1\in I} V(y_1) \geq cM^{\epsilon'}$ for intervals $I \subset \{|y_1| \leq 1\}$ of length $M^{-\epsilon'/2}$.

However, $\overline{V}_I \geq c(M^{-\epsilon'/2})^d \overline{V}$ as symbols, since $V$ is a polynomial on $y_1$. So the Proposition on positivity implies the operator estimate

$$\overline{V}_I \geq cM^{-(\epsilon' d)/2}\overline{V} - C \geq cM^{-(\epsilon' d)/2}M^\epsilon - C.$$

Picking $\epsilon' \leq \epsilon/(d+1)$, we get $\overline{V}_I \geq cM^{\epsilon'}$ as needed. Q.E.D.

Next we sketch the proof of sufficiency of $(\mathcal{P})$ in Theorem 3. For necessity, see Moyer [**31**]. First we cut up the problem by standard microlocal analysis, and restrict attention to a block $|x| \sim 1$, $|\xi| \sim M$ from Chapter I, Figure 3. On that block we can straighten out (say) the imaginary part of the symbol. So $L = i\tau + a(t,x,\xi)$, with $a \in S^1(1\times M)$. Condition $(\mathcal{P})$ says that for fixed $(x,\xi)$, $t \to a(t,x,\xi)$ never changes sign. Nirenberg and Trèves [**34**] had the idea to regard

$$[\partial/\partial t + a(t,x,D_x)]u = f, \qquad -T \leq t \leq +T, \tag{18}$$

as an evolution equation. The equation can be solved in three easy cases.

*Case* I. Suppose $a \geq 0$. Then (18) can be solved with time flowing forward, with initial condition $u|_{t=-T} = 0$. The analogue of this fact can easily be

checked for scalar evolution equations $[d/dt + a(t)]u = f$, $a \geq 0$. To make a proof in the $\psi d0$ context, we can exploit the simple estimate

$$\frac{d}{dt}||u(t,\cdot)||^2 = 2\,\mathrm{Re}\langle u_t, u\rangle = 2\,\mathrm{Re}\langle f - a(t,x,D_x)u, u\rangle$$

$$\leq 2\,\mathrm{Re}\langle f(t,\cdot), u(t,\cdot)\rangle + C||u(t,\cdot)||^2 \quad \text{(by the Proposition)}$$

$$\leq ||f(t,\cdot)||^2 + C'||u(t,\cdot)||^2.$$

If $u(-T,\cdot) = 0$, we get

$$||u(t,\cdot)||^2 < \int_{-T}^{t} e^{c'(t-s)}||f(s,\cdot)||^2\,dx,$$

so solutions of (18) are well behaved. It would be a diastrous mistake to try instead to impose the initial condition on $u(+T,\cdot)$, as one sees already for scalar evolution equations.

Suppose now $a \leq 0$. Then (18) can be solved just as simply, but this time we must impose the initial condition at $t = +T$.

*Case* II. Suppose $a(t,x,\xi) \geq 0$ for $x_1 > 0$, $a(t,x,\xi) \leq 0$ for $x_1 < 0$. Now we have trouble whether we put an initial condition at $+T$ or at $-T$. However, we could regard $L^2(R^n) = L^2\{x_1 > 0\} \oplus L^2\{x_1 < 0\} = H_+ \oplus H_-$, and hope that $a(t,x,D_x)$ breaks up modulo bounded errors as $A_+ \oplus A_-$ with $A_+ \geq 0$, $A_- \leq 0$. In that case, (18) breaks up into two uncoupled evolution equations, each of which can be treated as in Case I. This remarkable idea was carried out in Nirenberg and Trèves [**34**].

*Case* III. If $A(t,x,D_x)$ is bounded, then (18) is a trivial ODE in Hilbert space.

To prove sufficiency of $(\mathcal{P})$, Beals and Fefferman [**1, 2**] cut up the problem into microlocal equations which fall into Cases I–III. We make a Calderón-Zygmund decomposition of the initial block $|x| \sim 1$, $|\xi| \sim M$ by repeated bisection, stopping at $Q$ of sides $\delta \times M\delta$ if either $\mathrm{Vol}(Q) \sim 1$, or

$$\max_{|\alpha|+|\beta|\leq 1} \frac{|\partial_x^\alpha \partial_\xi^\beta a(t,x,\xi)|}{(M\delta)^{1-|\beta|}\delta^{-|\alpha|}} \geq C \quad \text{for some } t.$$

Thus $|x|$, $M^{-1}|\xi| \sim 1$ is cut into blocks $\{Q_\nu\}$. This is a slight variant of the cutting of Lemma 6 because of the extra parameter $t$. We still get $a(t,x,\xi) \in S^1\{Q_\nu\}$, so Beals calculus reduces (18) to a family of microlocal problems

$$(19) \qquad [\partial/\partial t + \chi_\nu a(t,x,D_x)]u_\nu = f_\nu, \quad \text{with } \chi_\nu \text{ essentially supported in } Q_\nu.$$

However, these problems all come under the easy Cases I–III. For the blocks $Q_\nu$ of volume $\sim 1$ we have $||\chi_\nu a(t,x,D_x)|| \leq C$, so we are in Case III. For the blocks $Q_\nu$ of volume $\gg 1$, a canonical transformation brings about either

(a) $a(t_0,x,\xi)\,|_{Q_\nu}$ is first-order elliptic for some $t_0 \in [-T,+T]$, or

(b) $a(t_0,x,\xi)\,|_{Q_\nu} = (M\delta_\nu^2)x_1$ for some $t_0$.

In case (a), the symbol $a(t_0,\cdot,\cdot)$ is always positive (or always negative) in $\mathrm{supp}(\chi_\nu)$. Condition $(\mathcal{P})$ shows that $\chi_\nu a(t,x,\xi) \geq 0$ for all $t,x,\xi$ (or else $\leq 0$ for all $t,x,\xi$), so we are in Case I.

Finally, in case (b) we know that

$$\begin{aligned} a(t_0, x, \xi) > 0 \quad &\text{for } x_1 > 0, \\ a(t_0, x, \xi) < 0 \quad &\text{for } x_1 < 0, \end{aligned} \qquad (x, \xi) \in Q_\nu.$$

Condition $(\mathcal{P})$ implies

$$\begin{aligned} \chi_\nu a(t, x, \xi) \geq 0 \quad &\text{for } x_1 > 0, \\ \chi_\nu a(t, x, \xi) \leq 0 \quad &\text{for } x_1 < 0, \end{aligned} \qquad \text{all } t, x, \xi.$$

So this time we are in Case II.

Hence the microlocalized problems are all solvable. The proof of sufficiency of $(\mathcal{P})$ is complete.

We close this Chapter by discussing the implications of Theorem 1 for second-order equations

$$L = \sum_{jk} a_{jk}(x) \frac{\partial^2}{\partial x_j \partial x_k} + \cdots, \qquad (a_{jk}(x)) \geq 0.$$

First let us suppose $L$ is selfadjoint with real coefficients. All the properties of $L$ as a PDE are determined by the geometry of a family of "balls" $B_L(x, \rho)$ associated to $L$. This fundamental idea was brought to light by Stein and his collaborators in [**19, 20, 32, 37**]. To define $B_L(x, \rho)$ we can suppose first that $(a_{jk}(x)) > 0$. Thus $L$ is the Laplacian in a metric, and we know what the ball $B_L(x, \rho)$ means. Now suppose $L$ is degenerate, and we perturb $L$ to make it positive definite: Say, $L_\epsilon = L + \epsilon\Delta$. As $\epsilon \to 0+$, much of the geometry of the $L_\epsilon$ will change wildly. For instance, the length of a typical curve $\gamma$ measured in the $\epsilon$-metric will tend to infinity as $\epsilon \to 0$. However, the ball $B_{L_\epsilon}(x, \rho)$ shrinks to a definite limit $B_L(x, \rho)$ as $\epsilon \to 0$. Usually $B_L$ is an open neighborhood of $x$, even though $L$ is degenerate at $+$ near $x$. It is interesting to see in examples how the shortest path from $x$ to $y$ changes as $\epsilon \to 0$.

In all but pathological examples, $Lu = f$ is hypoelliptic because it satisfies a subelliptic estimate

$$\langle -Lu, u \rangle + C\|u\|^2 \geq c\|u\|^2_{(\epsilon)}. \tag{20}$$

THEOREM 6. *Estimate* (20) *holds if and only if* $B_\Delta(x, \rho) \subseteq B_L(x, C\rho^\epsilon)$ *for all* $x$ *and some constant* $C$.

Compare with Oleinik and Radkevitch [**35**], who give necessary and sufficient conditions for the existence of an $\epsilon > 0$ with (20). We will give much sharper results in Chapter IV.

THEOREM 7. *Suppose* $L$ *is subelliptic and selfadjoint on a compact manifold* $M$ *with smooth measure* $\mu$. *Then the number* $N(\lambda, L)$ *of eigenvalues of* $-L$ *which are* $< \lambda$ *is given by*

$$N(\lambda, L) \approx \int_M \frac{d\mu(x)}{\mu(B_L(x, \lambda^{-1/2}))}. \tag{21}$$

For some purposes this formula is to be preferred to the obvious phase space volume even for the Laplacian. Already in the elementary example of

flat tori, (21) holds uniformly for all $\lambda$ and all tori, while the phase space volume is asymptotically correct for large $\lambda$ for each fixed torus. It would be interesting to prove (21) for the Laplacian under weak assumptions on $M$.

To prove Theorems 6 and 7, we combine the Main Lemma with some control on the geometry of $B_L$. The key result may be stated roughly as follows. (See [**16**] for the technical details.)

GEOMETRIC LEMMA. *Suppose* $L = (\partial/\partial y_1)^2 + \tilde{L}(y_1, y', \partial_{y'})$. *Then* $B_L(0,\rho)$ *looks essentially like a product* $\{|y_1| \le \rho\} \times B_{\overline{L}}(0,\rho)$, *where*

$$\overline{L} = \mathrm{Av}_{|y_1|<\rho} \tilde{L}(y_1, y', \partial_{y'}).$$

Together with Lemma 3, this lets us compute what $B_L(x,\rho)$ looks like. In particular, we see that $B_L(x,\rho)$ is essentially the image of a rectangle $\{|x_k| < \delta_k\}$ under a map $\phi$ with $|\phi|, |(\phi')^{-1}| \le C$.

The proof of the Geometric Lemma is a curious repetition of the proof of the Main Lemma, with $L^\infty$-norms replacing the $L^2$-norms.

Geometric ideas closely related to the $B_L(x,\rho)$ wil play a role in the ultrafine cutting in Chapter IV.

## CHAPTER IV: APPROXIMATE DIAGONALIZATION

At last we present the approximate diagonalization of pseudodifferential operators. Let us begin by explaining what kind of approximate diagonalization we are trying to achieve. We want to write a $\psi d0$ $L(x,\xi)$ as a sum $A + \mathcal{E}$, where

$$A = \begin{pmatrix} \boxed{\Lambda_1 A_1} & & & \mathbf{0} \\ & \boxed{\Lambda_2 A_2} & & \\ & & \boxed{\Lambda_3 A_3} & \\ \mathbf{0} & & & \boxed{\text{etc.}} \end{pmatrix}.$$

Here $\Lambda_\alpha$ is a scalar, the blocks are bounded in size, and $\|A_\alpha\| \le \text{Const}$. The error $\mathcal{E}$ is small in the sense that $\|\mathcal{E}u\|^2 \le \delta \sum_\alpha \Lambda_\alpha^2 \|u_\alpha\|^2$, where $\delta \ll 1$ and $u_\alpha$ is the component of the vector $u$ belonging to the block $\Lambda_\alpha A_\alpha$. By comparison, the main term $A$ satisfies $\|Au\|^2 \le C \sum_\alpha \Lambda_\alpha^2 \|u_\alpha\|^2$. If the matrix $A_\alpha$ satisfies $\|A_\alpha u_\alpha\|^2 \ge c\|u_\alpha\|^2$ with $c \gg \delta$, then $\|\mathcal{E}u\| \ll \|Au\|$, so $\|L(x,D)u\|^2 \sim \sum_\alpha \Lambda_\alpha^2 \|u_\alpha\|^2$. Thus, we understand $\|L(x,D)u\|$. Also, $L(x,D)^{-1}$ is well approximated by $A^{-1}$, which again breaks up into blocks of bounded size. So we understand $L(x,D)^{-1}$ also. In effect we have "approximately diagonalized" $L(x,D)$. The numbers $\Lambda_\alpha$ play the role of eigenvalues. Note that the error term is as bad as a small constant times the main term.

Next we fill in a little detail about the way $L(x,D)$ can be brought into the above form. We will cut phase space into boxes $\mathcal{B}_\alpha$ of bounded volume. Each $\mathcal{B}_\alpha$ may be straightened out by a canonical transormation $\Phi_\alpha$: $I \to \mathcal{B}_\alpha$, with $I = \{|x| \le 1, |\xi| \le B\}$, $B$ a large constant. Using a partition of unity $1 = \sum_\alpha \phi_\alpha$ with $\phi_\alpha$ supported in $\mathcal{B}_\alpha$, we can cut the symbol $L(x,\xi)$ into a sum

of symbols $L_\alpha(x,\xi) = \phi_\alpha L(x,\xi)$. Formally, we can apply Egorov's theorem to write

$$L(x,D) \simeq \sum_\alpha U_\alpha^*(L_\alpha \circ \Phi_\alpha)(x,D)U_\alpha \tag{1}$$

for suitable Fourier integral operators $U_\alpha$ associated to $\Phi_\alpha$. The symbol $\tilde{L}_\alpha = L_\alpha \circ \Phi_\alpha$ is localized to a straightened block $I$; but unfortunately $\tilde{L}_\alpha$ is much too big to belong to a good symbol class $S^m(I)$. To get around this, we simply introduce a large constant $\Lambda_\alpha$ comparable to $\max_I|\tilde{L}_\alpha| \sim \max_{B_\alpha}|L(x,\xi)|$, so that $\tilde{L}_\alpha^\dagger = L_\alpha(x,\xi)/\Lambda_\alpha = \phi_\alpha L/\Lambda_\alpha \circ \Phi_\alpha$ is a good symbol on $I$. Thus, (1) goes over to

$$L(x,D) \approx \sum_\alpha \Lambda_\alpha U_\alpha^* L_\alpha^\dagger(x,D)U_\alpha \tag{2}$$

with $L_\alpha^\dagger$ a good symbol on $I$. Now $L_\alpha^\dagger$ is a good symbol localized to a block $I$ of bounded volume $\sim B^n$. So in effect $L_\alpha^\dagger$ acts on a finite-dimensional vector space of functions microlocalized to $I$. If we use the Fourier integral operators $U_\alpha$ to split $u$ into its "components" $u_\alpha = U_\alpha u$, then we are in the situation of the opening paragraphs with $A_\alpha = L_\alpha^\dagger(x,D)$. The error term $\mathcal{E}$ arising from (2) will satisfy $\|\mathcal{E}u\|^2 \le \delta \sum_\alpha \Lambda_\alpha \|u_\alpha\|^2$ with $\delta$ a negative power of $B$ because symbolic calculus on the block $I$ works modulo errors dominated by lower powers of $B$.

A few points should be clarified. First of all, the symbols $\phi_\alpha$ and canonical transformations $\Phi_\alpha$ don't belong to any standard classes. Therefore it is not trivial to define the Fourier integral operators $U_\alpha$ or to prove (2); this is actually the main technical problem. On the other hand, we saw that $L_\alpha^\dagger(x,\xi)$ is a good symbol in a standard class $S^m(I)$. We shall prove theorems about $L(x,D)$ by studying the $L_\alpha^\dagger(x,D)$. We saw that an estimate $\|A_\alpha u_\alpha\| \ge cB^{m-2+\epsilon}\|u_\alpha\|$ is needed to make the appropriate diagonalization work. This amounts to a subelliptic estimate of the form

$$\|L_\alpha^\dagger(x,D)u\| \ge cB^{m-2+\epsilon}\|u\| \quad \text{microlocally in} \quad |x| \le 1, |\xi| \le B, \tag{3}$$
$$L_\alpha^\dagger \in S^m(1 \times B).$$

Once (3) holds, our machine will grind out the approximate diagonalization and approximate inverse for $L(x,D)$, a priori estimates, eigenvalues, etc. To explain how (3) is proved, we have to be more specific about what the symbol $L_\alpha^\dagger(x,\xi)$ looks like. There are two important cases.

I. Consider a second-order differential operator $L(x,D)$ with the nonnegative symbol

$$L(x,\xi) = \sum_{jk} a_{jk}(x)\xi_j\xi_k + \sum_k b_k(x)\xi_k + V(x)$$

microlocalized to $|x| \le 1$, $|\xi| \le B$.

Here $a_{jk}(x), b_k(x), V(x)$ are polynomials of degree at most $d$, and $|a_{jk}(x)| \le C$, $|b_k(x)| \le CB$, $|V(x)| \le CB^2$ for $|x| \le 1$ so that $L \in S^2(I)$.

Note that $b_k$ and $V$ are allowed to be rather large, so they cannot be neglected. We define the notion of $L$ in *normal form* by induction on the dimension $n$. In zero dimensions the block $I$ is a point and the symbol $L(x,\xi)$

is a number $V$. Then the symbol is in normal form if $V \geq cB^2$. In $n$ dimensions $L$ is in normal form if either

(a) $L$ is elliptic, i.e., $L(x,\xi) \geq cB^2$ in $|x| \leq 1$, $|\xi| \leq B$, or

(b) $L(x,\xi) = \xi_1^2 + \tilde{L}(x_1, x', \xi')$, with $\tilde{L}(x',\xi') = \int_{|x_1|\leq 1} \tilde{L}(x_1, x', \xi')\, dx_1$ in normal form in $(n-1)$ dimensions.

Here we have $L \in S^2(1 \times B)$, while the Main Lemma of Chapter III shows that $L(x,D) \geq cB^\epsilon$ microlocally in $|x| \leq 1$, $|\xi| \leq B$. So the needed subelliptic estimate (3) holds for second-order real symbols in normal form.

II. Consider the complex symbol $L(t,x,\tau,\xi) = i\tau + a(t)\xi + V(t,x)$ microlocalized to $|t|, |x| \leq 1$, $|\tau|, |\xi| \leq B$. Here $a, V$ are polynomials of degree at most $d$, and we suppose:

$$B^{-2} \leq a(t) \leq 1 \quad \text{for } |t| \leq 1, \qquad |V(t,x)| \leq B \quad \text{for } |t|, |x| \leq 1, \tag{4}$$

$$\frac{V(t,x)}{a(t)} \text{ is an increasing function of } t \text{ for each fixed } x. \tag{5}$$

Estimates (4) ensure that $L \in S^1(1 \times B)$, while (5) says that $L$ satisfies condition $(\Psi)$.

Egorov and Hörmander studied subelliptiticy of general $\psi d0$ by reducing matters to the special case $L(t,x,\tau,\xi)$ considered here. Sharp subelliptic estimates are not easy even in the special case, while the reduction of the general case to the example is very hard. However, as Egorov already saw, it is trivial to get a crude subelliptic estimate for $L$. In fact the change of variable $(t,x) \to (s,x)$ with $ds = a(t)\,dt$ reduces the problem to a slight variant of the Cauchy-Riemann equations in one complex variable.

For $u \in C_0^\infty$, we argue as follows. Let

$$\left[\frac{\partial}{\partial t} + ia(t)\frac{\partial}{\partial x} + V(t,x)\right]u = f.$$

Then

$$\left[\frac{\partial}{a(t)\partial t} + i\frac{\partial}{\partial x} + \frac{V(t,x)}{a(t)}\right]u = \frac{f}{a}, \quad \text{or} \quad \left[\frac{\partial}{\partial s} + i\frac{\partial}{\partial x} + G(s,x)\right]u = \frac{f}{a},$$

with $G(s,x) = V(t,x)/a(t)$. Note that $\partial G/\partial s \geq 0$ by (5). Now

$$\int |f(t,x)|^2 \frac{dt\,dx}{a(t)} = \int \left|\frac{f}{a}\right|^2 ds\,dx = \int \left|\frac{\partial u}{\partial s} + i\frac{\partial u}{\partial x} + G(s,x)u\right|^2 ds\,dx$$

$$= \int \left\{\left|\frac{\partial u}{\partial s}\right|^2 + \left|i\frac{\partial u}{\partial x} + Gu\right|^2 + \frac{\partial G}{\partial s}|u|^2\right\} ds\,dx.$$

This last step follows from the elementary computation

$$\left(\frac{\partial u}{\partial s} + i\frac{\partial u}{\partial x} + G\right)^*\left(\frac{\partial}{\partial s} + i\frac{\partial}{\partial x} + G\right) = \left(\frac{\partial u}{\partial s}\right)^*\left(\frac{\partial u}{\partial s}\right)$$

$$+ \left(i\frac{\partial}{\partial x} + G\right)^*\left(i\frac{\partial}{\partial x} + G\right) + \frac{\partial G}{\partial s}.$$

Since $\partial G/\partial s \geq 0$, we get

$$\int |f(t,x)|^2 \frac{dt\,dx}{a(t)} \geq \int \left|\frac{\partial u}{\partial s}\right|^2 ds\,dx + \int \left|i\frac{\partial u}{\partial x} + Gu\right|^2 ds\,dx$$

$$= \int \left|\frac{\partial u}{\partial t}\right|^2 \frac{dt\,dx}{a(t)} + \int \left|ia(t)\frac{\partial u}{\partial x} + Vu\right|^2 \frac{dt\,dx}{a(t)}. \tag{6}$$

So we have an energy estimate in a weighted $L^2$-norm. Unfortunately, what we need for Egorov's theorem is the analogous estimate without the weight factor $1/a(t)$. It is not easy to prove this.

However, we can get a cheap estimate if we just replace $1/a(t)$ in (6) by its maximum on the left and by its minimum on the right. The result is

$$\|L(t,x,D_t,D_x)u\|^2 \geq B^{-2}\left\|\frac{\partial u}{\partial t}\right\|^2 + B^{-2}\left\|ia\frac{\partial u}{\partial x} + Vu\right\|^2, \qquad u \in C_0^\infty. \tag{7}$$

The sharp result would have no $B^{-2}$ factors on the right, so (7) is very crude. Nevertheless, a subelliptic estimate

$$\|\partial u/\partial t\|^2 + \|ia(\partial u/\partial x) + Vu\|^2 \geq B^\epsilon \|u\|^2 \tag{8}$$

will imply

$$\|L(t,x,D_t,D_x)u\|^2 \geq B^{\epsilon-2}\|u\|^2,$$

i.e.,

$$\|L(t,x,D_t,D_x)u\| \geq B^{\epsilon/2-1}\|u\|, \qquad u \in C_0^\infty,\ L \in S^1(1 \times B),$$

which is (3). In other words, the crude estimate (7), although far from sharp, is strong enough to power our machine. Approximate diagonalization will prove Egorov's theorem in the sharp form

$$\|p(x,D)u\| + \|q(x,D)u\| \leq C\|(p+iq)(x,D)u\| + C\|u\|_{(\epsilon)}$$

for $p+iq \in S^1$ satisfying $(\Psi)$. In the special case $p+iq = \tau + ia(t)\xi + V(t,x)$, we recover the sharp form of (7) without the $B^{-2}$ factors. Here we see a curious feature of our machine: To get good estimates we microlocalize the symbol into pieces of a special form. It is important to carry out the microlocalization, even when the initial problem is already in the special form.

Let us review the situation. To give an approximate diagonalization of $L(x,D)$ we cover phase space by boxes $\mathcal{B}_\alpha$ which arise as images of $I = \{|x| \leq 1, |\xi| \leq B\}$ by canonical transformations $\Phi_\alpha$. Setting $\Lambda_\alpha \sim \max_{\mathcal{B}_\alpha}|L(x,\xi)|$, we hope to write $L(x,D) \approx \sum_\alpha \Lambda_\alpha U^*_\alpha L^\dagger_\alpha(x,D)U_\alpha$, where $L^\dagger_\alpha(x,\xi) = \phi_\alpha L/\Lambda_\alpha \circ \Phi_\alpha$ living on $\mathcal{B}_\alpha$, and $U_\alpha$ is a Fourier integral operator associated to $\Phi_\alpha$. Each $L^\dagger_\alpha(x,D)$ is, in effect, a matrix of bounded rank $\sim B^n$, so the decomposition operator $u \to (u_\alpha) = (U_\alpha u)$ splits $L(x,D)$ as a "diagonal" part plus a small error. The error is dominated by a small constant times the main term, provided each of the $\psi d0$'s $L^\dagger_\alpha(x,D)$ satisfy a subelliptic estimate microlocally in $|x| \leq 1$, $|\xi| \leq B$. It is important that the subelliptic estimate need not be

anywhere near best possible. We can get good enough subelliptic estimates in two important cases:

I. $L_\alpha^\dagger(x,D)$ a selfadjoint second-order PDE in a special form. Here the necessary estimate comes from methods of Chapter III.

II. $L_\alpha^\dagger(x,D)$ a special nonselfadjoint first order PDE satisfying $(\Psi)$. General symbols satisfying $(\Psi)$ will be microlocalized to this case. A crude estimate can be derived by elementary integration by parts.

To make our machine work we have to produce the boxes $\mathcal{B}_\alpha$, the canonical transformations $\Phi_\alpha$, and the Fourier integral operators $U_\alpha$.

It is useful to make a pseudodifferential operator calculus adapted to the boxes $\mathcal{B}_\alpha$ and weights $\Lambda_\alpha$. We say that a symbol $p(x,\xi)$ belongs to $S^{m\mu} = S^{m\mu}(\mathcal{B}_\alpha,\Lambda_\alpha)$ if $(p\circ\Phi_\alpha)/\Lambda_\alpha^{m/2} \in S^\mu(I)$ with uniform bounds on the $S^\mu$-seminorms.

Thus, $p\in S^{m\mu}$, $q\in S^{n\nu}$ imply $pq\in S^{m+n\,\mu+\nu}$, $\{p,q\}\in S^{m+n\,\mu+\nu-1}$. A basic elliptic symbol which plays the role of the Laplacian is $\Lambda(x,\xi)\sim\Lambda_\alpha$ for $(x,\xi)\in\mathcal{B}_\alpha$. Thus $\Lambda\in S^{1,0}$.

To $p\in S^{m\mu}$ we associate the "pseudodifferential operator"

$$p(x,D)=\sum_\alpha \Lambda_\alpha^{m/2}U_\alpha^*\left[\frac{\phi_\alpha p}{\Lambda^{m/2}}\circ\Phi_\alpha(x,D)\right]U_\alpha.$$

The basic facts about our $\psi d0$ are as follows.

THEOREM. *Fix $L(x,\xi)\in S^2(1\times M)$ with $L\geq M^\epsilon$ and take a large constant $B$ with $1\ll B\leq M^\epsilon$. From $L$ and $B$ we get a family of boxes $\{\mathcal{B}_\alpha\}$ in phase space and weights $\Lambda_\alpha=\max_{\mathcal{B}_\alpha}L$. These give rise to a $\psi d0$ calculus with the following properties:*

(A) Composition law. *If $p\in S^{m\mu}$ and $q\in S^{n\nu}$, then $p(x,D)q(x,D)=p\circ q(x,D)$, with $p\circ q=pq+i/2\{p,q\}+\mathcal{E}$ and $\mathcal{E}\in S^{m+n\,\mu+\nu-2}$.*

(B) Adjoints. *If $p\in S^{m\mu}$, then $[p(x,D)]^*=p^\#(x,D)$, with $p^\#=\overline{p}+\mathcal{E}$, $\mathcal{E}\in S^{m\mu-2}$.*

(C) Boundedness on Sobolev spaces. *If $p\in S^{m\mu}$, then*

$$||p(x,D)u||\leq CB^\mu||\Lambda^{m/2}(x,D)u||.$$

*In particular, $p\in S^{0\mu}$ implies $||p(x,D)u||\leq CB^\mu||u||$.*

(D) Elliptic and subelliptic estimates. *Let $p\in S^{m\mu}$ and suppose the symbols $p_\alpha=(p\circ\Phi_\alpha)/\Lambda_\alpha^{m/2}|_I$ satisfy subelliptic estimates $||p_\alpha(x,D)u||\geq cB^{\mu-2+\epsilon}||u||$ microlocally in $I$. Then*

$$||p(x,D)u||\geq cB^{\mu-2+\epsilon}||\Lambda^{m/2}(x,D)u||.$$

Relation to classical $\psi d0$.

(E) *If $p\in S^{m\mu}$ is a classical symbol, then our $\psi d0$ $p(x,D)$ differs from the operator $p(x,D)$ in the classical Weyl calculus by an error $\mathcal{E}(x,D)$ with $\mathcal{E}\in S^{m\mu-2}$.*

(F) *A classical symbol $p(x,\xi)$ belongs to $S^{m\mu}$ if and only if $\max_{\mathcal{B}_\alpha}|p|\leq C\Lambda_\alpha^{m/2}B^\mu$. In particular, $L(x,\xi)\in S^{2,0}$.*

(G) Special behavior of $L(x,\xi)$ in its own calculus.

$$B^2/\Lambda_\alpha L \circ \Phi_\alpha(x,\xi)|_I = \sum_{j,k\geq 1} a_{jk}(x)\xi_j\xi_k + \sum_k b_k(x)\xi_k + V(x) + \mathcal{E}(x,\xi),$$

*where* $|\mathcal{E}(x,\xi)| \leq CM^{-\epsilon}$ *and* $\sum_{jk} a_{jk}\xi_j\xi_k + \sum_k b_k\xi_k + V$ *is in normal form. Also* $a_{jk}(x_1,\dots,x_n)$ *depends only on the* $x_l$ *with* $l < \min(j,k)$.

Applications of the $\psi d0$ calculus are as follows.

*SAK principle.* Start with $L \in S^2(1 \times M)$, $L \geq M^\epsilon$. We have $B^2L(x,\xi) \in S^{2,2}$ by (F), while $B^2L/\Lambda_\alpha \circ \Phi_\alpha(x,D) \geq cB^\epsilon$ microlocally in $I$ by (G) and the subellipticity of PDE in normal form. Therefore (E) implies

$$(*) \qquad ||B^2L(x,D)u|| \geq CB^\epsilon||\Lambda(x,D)u||.$$

On the other hand, suppose $p(x,\xi)$ is a classical symbol satisfying $\max_{\mathcal{B}_\alpha}|p| \leq \max_{\mathcal{B}} L = \Lambda_\alpha$. Then $p \in S^{2,0}$ by (F), so

$$(**) \qquad ||p(x,D)u|| \leq C||\Lambda(x,D)u|| \quad \text{by (C)}.$$

So far $L(x,D)$ and $p(x,D)$ are defined by our $\psi d0$ calculus, but we can switch over to the usual $\psi d0$ $L(x,D)$, $p(x,D)$ with errors of the form $\mathcal{E}(x,D)$, $\mathcal{E} \in S^{2,-2}$ by virtue of (E). Since $||\mathcal{E}(x,D)u|| \leq CB^{-2}||\Lambda(x,D)u||$ by (C), estimates $(*)$ and $(**)$ are unaffected by the switch. Finally, $(*)$ and $(**)$ yield $||p(x,D)u|| \leq CB^{2-\epsilon}||L(x,D)u||$. This is the SAK principle for a priori estimates since $B$ is just a fixed large constant. Of course we could start with $L \geq 0$ instead of assuming $L \geq M^\epsilon$ and simply work with $L + M^\epsilon$ in place of $L$. So we have proved

THEOREM (*SAK*). *Assume* $p, L$ *are classical second-order symbols with* $L \geq 0$. *The a priori estimate* $||p(x,D)u|| \leq C||L(x,D)u|| + C||u||_{(\epsilon)}$ *holds if and only if* $\max_{\mathcal{B}_\alpha}|p| \leq C\max_{\mathcal{B}_\alpha}(L + [1+|\xi|]^\epsilon)$ *for suitable boxes* $\mathcal{B}_\alpha$ *associated to* $L$.

COROLLARY. *If* $p, L$ *are classical second-order symbols with* $L \geq |p|$ *pointwise, then* $||p(x,D)u|| \leq C||L(x,D)u|| + C||u||_{(\epsilon)}$.

The corollary yields highly nontrivial estimates, such as

$$||X_j^2 u|| \leq C\left\|\sum_j X_j^2 u\right\| + C||u||_{(\epsilon)},$$

which requires the full force of Rothschild's and Stein's machine. To obtain $||X_jX_ku|| \leq C||\sum_j X_j^2 u|| + C||u||_{(\epsilon)}$, we need the full *SAK* theorem, since $X_jX_k$ has symbol $p_jp_k + i/2\{p_j,p_k\}$ in the Weyl calculus; here, $p_j$ is the symbol of the vector field $X_j$.

Now $|p_jp_k| \leq \sum_j p_j^2$ pointwise, but regarding $\{p_j,p_k\}$, we can say only that $\max_{\mathcal{B}_\alpha}|\{p_j,p_k\}| \leq C\max_{\mathcal{B}_\alpha}\sum_j p_j^2$, as in the proof of Hörmander's theorem in Chapter III.

Next we pass to

*Egorov's theorem.* For $p+iq \in S^1$ of principal type and satisfying condition $(\Psi)$, we shall prove that

$$||p(x,D)u|| + ||q(x,D)u|| \leq C||(p+iq)(x,D)u|| + C||u||_{(\epsilon)},$$

which immediately yields subellipticity. To prove this sharp estimate, we may first suppose $q = \tau$, $p = p(t, x, \xi)$. (This comes already out of classical microlocal analysis.) Now we set $L = \tau^2 + p^2 + M^\epsilon$ and apply our $\psi d0$ calculus. Since $\tau^2, p^2 \leq L$ pointwise, it follows from (F) that $\tau, q \in S^{1,0}$. A repetition of the proof of the $SAK$ theorem above yields the desired estimate, provided we can establish a subelliptic estimate for $A_\alpha(y, \eta) = (\tau + iq)/\Lambda^{1/2} \circ \Phi_\alpha$, namely $||A_\alpha(y, D)u|| \geq cB^{\epsilon-1}||u||$ microlocally in $I$.

It isn't hard to follow the construction of $\Phi_\alpha$ in this case, and one finds that $A_\alpha$ must be a first-order differential operator of the form

$$\frac{\partial}{\partial t} + i\sum_k a_k(t, x)\frac{\partial}{\partial x_k} + V(t, x)$$

with $\max_{|x|,|t|\leq 1}|a_k(t, x)|$ decreasing in $k$ and $a_k(t, x)$ independent of $x_l$, $l \geq k$. Such an operator can satisfy $(\Psi)$ only if it has the special form[6]

$$\partial/\partial t + ia(t)(\partial/\partial x_1) + V(t, x_1, x').$$

Here $x'$ appears only as an irrelevant parameter, and our subelliptic estimate is reduced to II above. So Egorov's theorem may be read off from our $\psi d0$ calculus by using the estimate II.

*Explicit solutions of PDE.* Let $L$ be a second-order symbol as in the statement of the theorem. We shall look for an approximate solution of $L(x, D)u = f$. Introduce a partition of unity $1 = \sum_\alpha \psi_\alpha(x, \xi)$ with $\psi_\alpha$ supported in $\mathcal{B}_\alpha$. Formally, we can try to write $u = \sum_\alpha U^*_\alpha v_\alpha$, where $v_\alpha$ solves

$$\Lambda_\alpha L_\alpha(x, D)v_\alpha = \tilde{\psi}_\alpha(x, D)U_\alpha f \qquad \text{with } L_\alpha(x, \xi) = \frac{L \circ \Phi_\alpha}{\Lambda_\alpha}, \quad \tilde{\psi}_\alpha = \psi_\alpha \circ \Phi_\alpha.$$

Since $L_\alpha(x, D)$ is subelliptic on $I$, we can write $T_\alpha = [L_\alpha(x, D)]^{-1}$ with $||T_\alpha|| \leq CB^{-\epsilon}$, and

$$u = \sum_\alpha \Lambda_\alpha^{-1} U^*_\alpha T_\alpha \tilde{\psi}_\alpha(x, D)U_\alpha f.$$

Our $\psi d0$ calculus shows that $Lu = f + \mathcal{E}(x, D)f$ with $\mathcal{E} \in S^{0,-2}$. In particular, the error $\mathcal{E}(x, D)f$ has $L^2$-norm small compared to that of $f$, so we can get an exact solution by successive approximation.

Of course our solution is only as explicit as our knowledge of $U_\alpha$ and $T_\alpha$. In at least one case, namely $L$ a second-order differential operator, $U_\alpha$ is easy to understand: Here $f \to U_\alpha f$ is induced by a change of variable $y = \phi_\alpha(x)$ which carries a small ball $B_L(x^\alpha, \rho_\alpha)$ to the unit cube.

For more general symbols, $U_\alpha$ is given as an oscillatory integral with degenerate phase function, so our understanding of $U_\alpha$ is less explicit.

At first sight it appears that we have to calculate $T_\alpha = [L_\alpha(x, D)]^{-1}$. We don't have to! The reason is that the problem is localized to a block $I$ of bounded volume in phase space. In effect, $L_\alpha(x, D)$ acts on the space

$$H_B = [\text{direct sum of eigenfunctions of } -\Delta + B^2|x|^2 \text{ with eigenvalue} < 4B^2].$$

[6]Here we exclude the trivial case in which $p$ always has the same sign.

Letting $\psi_1, \dots, \psi_s$ be an orthonormal basis of Hermite functions in $H_B$, we can write $L_\alpha(x,D)|_I$ in the form of an $s \times s$ matrix $(A_{jk})$. The subelliptic estimate yields a bound for the inverse matrix $(A^{jk})$. All we need is $|A^{jk}| \le C$, which is weaker than the known bound $\|(A^{jk})\| \le CB^{-\epsilon}$. Consequently $T_\alpha$ can be written as an integral operator

$$T_\alpha f(x) = \int \left\{ \sum_{jk} A^{jk} \psi_j(x) \overline{\psi_k(y)} \right\} f(y)\, dy.$$

Each $\psi_j, \psi_k$ is a nice Schwartz function, so $T_\alpha f(x) = \int K_\alpha(x,y) f(y)\, dy$ for a kernel $K_\alpha(\cdot,\cdot) \in \mathcal{S}(R^{2n})$. Of course the seminorms of $K_\alpha$ in $\mathcal{S}(R^{2n})$ grow like high powers of $B$. This does not bother us, since in the end $B$ is just a fixed large constant.

Now we return to our formula

$$u = \sum_\alpha \Lambda_\alpha^{-1} U^*{}_\alpha T_\alpha \tilde{\psi}_\alpha(x,D) U_\alpha f = \sum_\alpha \Lambda_\alpha^{-1} U^*{}_\alpha \tilde{T}_\alpha U_\alpha f, \tag{9}$$

where $\tilde{T}_\alpha = T_\alpha \tilde{\psi}_\alpha(x,D)$ is again given by a Schwartz kernel. Since each $\tilde{T}_\alpha$ is a harmless operator acting on functions on the unit cube, we now have a reasonable chance to write $u = \int K(x,y) f(y)\, dy$ with some control on the properties of $K$.

For instance, say $L$ is a second-order differential operator as in Oleinik and Radkevitch. Then we recall $U_\alpha f(x) = |\det \phi'(x)|^{1/2} f(\phi(x))$ for a change of variable $y = \phi(x)$ carrying the unit cube to a small $B_L(x_\alpha, \rho_\alpha)$. Putting this back into (9), we get $u(x) = \int K(x,y) f(y)\, dy$ with $K(x,y)$ estimated in terms of the constants $\Lambda_\alpha$ and the Jacobian factors $|\det \phi'|^{1/2}$ in $U_\alpha$. The result one expects is as follows. (We exclude the one- and two-dimensional cases.)

CONJECTURE. For $L = -\sum_{jk} a_{jk}(x) \partial_{jk}^2 + \sum_k b_k(x) \partial_k + c(x)$ subelliptic with real coefficients, the solution to $Lu = f$ is given by $u(x) = \int K(x,y) f(y)\, dy$ with

$$|K(x,y)| \le C \frac{\delta^2(x,y)}{\mathrm{Vol}(x,y)}, \qquad \delta(x,y) = \min\{\rho \mid y \in B_L(x,\rho)\},$$

$$\mathrm{Vol}(x,y) = \mathrm{Vol}\, B_L(x,\rho) \quad \text{with } \rho = \delta(x,y).$$

One has also natural estimates for the derivatives of $K$.

If $L$ is selfadjoint, then also $K(x,y) \ge c\delta^2(x,y)/\mathrm{Vol}(x,y)$, so the order of magnitude of $K(x,y)$ is known. The case $L = \sum_j X_j^2 + X_0$ was solved earlier by Nagel, Stein and Wainger [**32**] and independently by Sanchez [**38**] using the Rothschild-Stein lifting. The above methods should make possible a proof in the general case. This is now being carried out by Sanchez.

It would be interesting to write explicit inverses of some equations satisfying $(\Psi)$, such as $\partial/\partial t + ia(t)(\partial/\partial x) + V(t,x)$ microlocalized to $|t|\,|x| \le 1$, $|\tau|$, $|\xi| \le M$.

At one time there were two different ways to understand PDE: Prove a priori estimates or construct approximate solutions. Now the distinction between the two methods is starting to blur. If we can prove sharp enough estimates, then we have understood the geometry of the equation well enough to give an

approximate diagonalization, hence an approximate solution. Construction of the solution rests on a priori estimates for localized problems.

It would be very interesting to understand the analogues of these ideas for systems, especially for $\overline{\partial}$. Although a lot is known about strongly pseudoconvex domains (Bergman, Szegö, Neumann kernels [**3, 36**]; Poincaré metric [**7, 29**]), the class of weakly pseudoconvex domains is much more mysterious. Kohn [**27**] and his students made a deep study of subellipticity of $\overline{\partial}$. In particular, Catlin has given evidence that a subelliptic estimate of order $\epsilon$ holds for $\overline{\partial}$ on (0,1)-forms on $D$ if and only if the following geometric condition holds:

(*) Given a point $p \in D$ with distance $r$ to the boundary, no analytic disc of radius $cr^{\epsilon}$ can be imbedded in $D$ centered at $p$.

The condition is necessary for subellipticity [**6**].

Catlin conjectures other closely related sharp estimates for $\overline{\partial}$, which he can verify in many examples. Now geometric conditions like (*) must come from an *SAK* principle for systems, for they say that the nonexistence of an imbedded geometrical object is equivalent to a subelliptic estimate. If we could carry over our machinery to systems, we would have some hope of understanding the Bergman and Szegö kernels for weakly pseudoconvex domains. We would already be delighted to understand a few nontrivial examples. Right now there is essentially only one weakly pseudoconvex domain on which the kernels can be computed, namely $\{\sum_k |z_k|^{2m_k} < 1\}$ (see D'Angelo [**9**]). The geometry of weakly pseudoconvex domains is very subtle.

We close the chapter by explaining how to associate boxes $\mathcal{B}_\alpha$, canonical transformations $\Phi_\alpha$, and Fourier integral operators $U_\alpha$ to a given symbol $L \in S^2(1 \times M)$. The boxes $\mathcal{B}_\alpha$ are analogous to the non-Euclidean balls associated to a second-order PDE.

$\mathcal{B}_\alpha$ is defined by a construction using induction on the dimension. To understand it, imagine we don't know the correct definition of $B_L(x,\rho)$, but we try to construct it so that the Geometric Lemma of Chapter III holds. We should also understand how $L(x,D)$ behaves on $B_L(x,\rho)$, and this leads us to the following.

LEMMA 1. *After a change of coordinates $y = \psi(x)$ with $|\psi'|, |(\psi')^{-1}| \leq C$, the operator $L$ and the ball $B_L(x,\rho)$ may be placed in the following special form:*

$$B_L(x,\rho) \sim \{|y_1| < \delta_1, |y_2| < \delta_2, \ldots, |y_n| < \delta_n\} \quad \text{with } \rho \geq \delta_1 \geq \delta_2 \geq \cdots \geq \delta_n.$$

$$\rho^2 L = \delta_1^2 \frac{\partial^2}{\partial y_1^2} + \tilde{L}_1(y_1, y', D_{y'});$$

$$\text{define } \overline{L}_1(y', D_{y'}) = \mathrm{Av}_{|y_1|<\delta_1} \tilde{L}_1(y_1, y', D_{y'}).$$

$$(*) \qquad \rho^2 \overline{L}_1 = \delta_2^2 \frac{\partial^2}{\partial y_2^2} + \tilde{L}_2(y_2, y'', D_{y''});$$

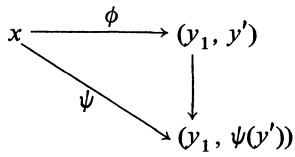
$$\text{define } \overline{L}_2(y'', D_{y''}) = \mathrm{Av}_{|y_2|<\delta_2} \tilde{L}_2(y_2, y'', D_{y''}).$$

$$\vdots$$

$$\rho^2 \overline{L}_{n-1} = \delta_n^2 \frac{\partial^2}{\partial y_n^2}.$$

Thus, in some average sense, $\rho^2 L$ looks like $\sum_k \delta_k^2(\partial/\partial y_k)^2$ on $B_L(x,\rho)$.

SKETCH OF PROOF. The result follows by induction on the dimension using the Geometric Lemma of Chapter III. In fact, we make an initial coordinate change $y=\phi(x)$ to bring about $\rho^2 L = \delta_1^2(\partial/\partial y_1)^2 + L_1(y_1, y', D_{y'})$, and we know by the Geometric Lemma that $B_L(x,\rho)$ looks in $y$-coordinates like $\{(y_1,y') \mid |y_1| \le \delta_1,\ y' \in B_{\overline{L}_1}(x',\rho)\}$. Inductive hypothesis applied to $\overline{L}_1(y', D_{y'})$ produces a coordinate transformation $\overline{\psi}\colon y' \to \overline{y}'$ in $(n-1)$-variables so that $B_{\overline{L}_1}(x',\rho) \sim \{|\overline{y}_k| \le \delta_k, k \ge 2\}$, and all but the first line of $(*)$ holds in the $\overline{y}$ coordinates. Now just define $\psi$ by the diagram




and the lemma follows. Q.E.D.

We now return to the problem of constructing $\mathcal{B}_\alpha, \Phi_\alpha, U_\alpha$, given $L \in S^2$. The geometry is analogous to the simple lemma just proved, only now our constructions work in phase space.

First we define the $\mathcal{B}_\alpha$. By analogy to the geometry of $\sum a_{jk}(x)\partial^2/\partial x_j \partial x_k$, we shall introduce a family of non-Euclidean balls $\mathcal{B} = B_L((x^0,\xi^0),\rho)$ in phase space associated to the symbol $L$. Each $\mathcal{B}$ will be the image of a block $I_{\mathcal{B}} = \{|x| \le 1, |\xi| \le M_{\mathcal{B}}\}$ under a canonical transformation $\Phi_{\mathcal{B}}$. The $\{\mathcal{B}_\alpha\}$ will then be simply those non-Euclidean balls $\mathcal{B}$ with $M_{\mathcal{B}} = B$.

Our construction proceeds by induction on the dimension. First we construct the balls of radius 1. Given $(x^0,\xi^0)$ and $L \in S^2(1 \times M)$, we apply the decomposition of Chapter III, Lemma 7, and let $Q$ be the block containing $(x^0,\xi^0)$. Say $Q$ has sides $\delta \times M\delta$. There are three possibilities:

(A) $L$ is elliptic on $Q$, i.e., $L \ge c(M\delta^2)^2$ on $Q$;

(B) $\mathrm{Vol}(Q) \sim 1$;

(C) $L$ is nondegenerate on $Q$.

In cases (A) and (B) we define $B_L((x^0,\xi^0),1)$ to be simply $Q$. Evidently $Q$ is the image of a suitable block $\{|x| \le 1, |\xi| \le M_{\mathcal{B}}\}$ under a symplectic affine change of scale $\Phi_{\mathcal{B}}$.

In case (C) we proceed by analogy with the Geometric Lemma of Chapter III. Here we start with a canonical transformation $\Psi\colon Q \to Q^*$ with natural bounds, so that

$$L \circ \Psi = \delta^2\tau^2 + \tilde{L}(t,y,\eta),$$

$$\Psi^{-1}(x^0,\xi^0) \text{ is the point } (y,\eta) = 0, t = 0, \tau = \tau_0.$$

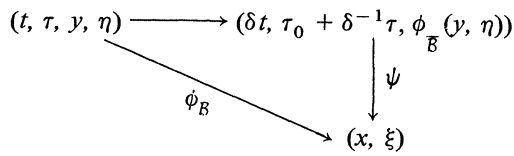
Now set

$$\overline{L}(y,\eta) = C\delta^2\tau_0^2 + \mathrm{Av}_{|t| \le \delta} L(t,y,\eta),$$

$C$ a large constant. Since $L(y,\eta)$ is a symbol in fewer dimensions, the inductive hypothesis says we have already constructed $\overline{\mathcal{B}} = B_{\overline{L}}((0,0),1) = \Phi_{\overline{\mathcal{B}}}(II)$, with $II = \{|y| \le 1, |\eta| \le \overline{M}\}$. Now set $I = \{|t|, |y| \le 1;\ |\tau|, |\eta| \le \overline{M}\}$, and define

$\mathcal{B} = B_L((x^0, \xi^0), 1) = \Phi_{\mathcal{B}}(I)$, where $\Phi_{\mathcal{B}}$ is given by the diagram

(10)




This completes our construction of the ball of radius one. For $\rho \ll 1$, the ball $B_L((x, \xi), \rho)$ is defined simply as $B_{\rho^2 L}((x, \xi), 1)$.

It would certainly be simpler to have a natural intrinsic definition of the balls, rather than an inductive construction. Before continuing, we check that our definition agrees with the familiar non-Euclidean balls associated to $L = \sum a_{jk}(x)\xi_j\xi_k$. In fact, the Geometric Lemma from Chapter III shows by induction on the dimension that $\{(y, \eta) \mid \eta = 0\} \cap B_L((x, 0), \rho)$ is essentially $B_L(x, \rho)$.

So far, we have defined the boxes $\mathcal{B}_\alpha$ and the canonical transformations $\Phi_\alpha$ which straighten them out. Now we must associate Fourier integral operators $U_\alpha$ to the $\Phi_\alpha$. Again, we carry out the construction of a F.I.O. $U_{\mathcal{B}}$ for each of the canonical transformations $\Phi_{\mathcal{B}}$ associated to $\mathcal{B} = B_L((x^0, \xi^0), \rho)$. As before, it is enough to look at the case $\rho = 1$. We proceed by induction on the dimension. The construction of $U_{\mathcal{B}}$ is evident in cases (A) and (B) above—it is nothing but an affine change of scale. The delicate point is to construct $U_{\mathcal{B}}$ in case (C). Here $\Phi_{\mathcal{B}}$ is given as the composition of three canonical transformations, namely

$$(t, \tau, y, \eta) \to (t, \tau, \Phi_{\bar{\mathcal{B}}}(y, \eta)), \tag{11}$$

$$(t, \tau, y, \eta) \to (\delta t, \tau_0 + \delta^{-1}\tau, y, \eta), \tag{12}$$

$$(t, \tau, y, \eta) \xrightarrow{\Psi} (x, \xi), \tag{13}$$

so it is enough to associate Fourier integral operators to these three. Now $\Psi$ satisfies good estimates on $Q$, and so is already covered by Egorov's theorem. Transformation (11) corresponds to $Uf(t, y) = \delta^{+n/2} f(\delta t, y) e^{i\tau_0 t}$, so again there is no problem. The tricky part is (10). However, by the inductive hypothesis we have already succeeded in associating a Fourier integral operator $U_{\bar{\mathcal{B}}}$ to the canonical transformation $\Phi_{\bar{\mathcal{B}}}$. So we can associate an operator to (10) by letting $f(t, y) \to U_{\bar{\mathcal{B}}} f(t, y)$ for each fixed $t$.

This completes the construction of the Fourier integral operators. At last we have the boxes $\mathcal{B}_\alpha$, the straightening transformation $\Phi_\alpha$, and the associated Fourier integral operators $U_\alpha$. These are the ingredients of our $\psi d0$ calculus.

Of course we still have not explained how to prove our theorem on $\psi d0$ calculus. That story is too complicated to present here. At least property (G) of the symbol $L$ in its own calculus is rather clear—it is the analogue of Lemma 1, Chapter IV. This property is the reason our main terms dominate the errors. Our long struggle from the elementary localization at the start of Chapter I to the present $\psi d0$ calculus has gained us exactly this one advantage.

## REFERENCES


1. R. Beals and C. Fefferman, *On local solvability of partial differential equations*, Ann. of Math. (2) **97** (1973), 552–571.
2. ——, *Spatially inhomogenous pseudodifferential operators*, Comm. Pure Appl. Math. **27** (1974), 1–24.
3. M. Beals, C. Fefferman and R. Grossman, *Strictly pseudoconvex domains in* $\mathbf{C}^n$, Bull. Amer. Math. Soc. (N.S.) **8** (1983), 125–322.
4. A. Calderón and R. Vaillancourt, *A class of bounded pseudodifferential operators*, Proc. Nat. Acad. Sci. U.S.A. **69** (1972), 1185–1187.
5. T. Carleman, *Complete works* (A. Pleijel et al, eds.), Malmo Litös, 1960.
6. D. Catlin, *Necessary conditions for subelliptitiicy and hypoellipticity for the* $\overline{\partial}$*-Neumann problem on pseudoconvex domains*, Recent Developments in Several Complex Variables, Ann. of Math. Stud., no. 100, 1981.
7. S. Cheng and S.-T. Yau, *On the existence of complete Kähler metric on noncompact manifolds and the regularity of Fefferman's equation*, Comm. Pure Appl. Math. **33** (1980), 507–544.
8. A. Cordoba and C. Fefferman, *Wave packets and Fourier integral operators*, Comm. Partial Differential Equations **3**(11) (1978), 979-1005.
9. J. D'Angelo, *Real hypersurfaces with degenerate Levi form*, Thesis, Princeton Univ., Princeton, N.J., 1976.

9a. F. Dyson and A. Lenard, *Stability of matter*. I, J. Math. Phys. **8** (1967), 423–434.

10. Yu. V. Egorov, *Subelliptic operators*, Uspekhi Mat. Nauk. **30** (1975), 57–104; English transl, Russian Math. Surveys **30** (1975).
11. C. Fefferman, *Recent progress in classical Fourier analysis*, Proc. Internat. Congress Math. (Vancouver, 1974), Vol. 1, Canad. Math. Congr., Montreal, Que., 1975, pp. 95–118.
12. ——, *The Bergman kernel and biholomorphic mappings of pseudoconvex domains*, Invent. Math. **26** (1974), 1–66.
13. C. Fefferman and D. H. Phong, *On positivity of pseudodifferential operators*, Proc. Nat. Acad. Sci. U.S.A. **75** (1978), 4673–4674.
14. ——, *The uncertainty principle and sharp Gårding inequalities*, Comm. Pure Appl. Math. **34** (1981), 285–331.
15. ——, *On the asymptotic eigenvalue distribution of a pseudodifferential operator*, Proc. Nat. Acad. Sci. U.S.A. **77** (1980), 5622–5625.
16. ——, *Subelliptic eigenvalue problems*, Conf. in Honor of A. Zygmund. (W. Beckner et al, eds.), Wadsworth Math. Series, 1981.
17. ——, *Symplectic geometry and positivity of pseudodifferential operators*, Proc. Nat. Acad. Sci. U.S.A. **79** (1982), 710–713.
18. R. Fefferman, personal communication.
19. G. Folland and E. M. Stein, *Estimates for the* $\overline{\partial}_b$ *complex and analysis on the Heisenberg group*, Comm. Pure Appl. Math. **27** (1974), 429–522.
20. R. Greiner and E. M. Stein, *Estimates for the* $\overline{\partial}$*-Neumann problem*, Math. Notes **19** (1977).
21. R. Hunt, *An extension of the Marcinkiewicz interpolation theorem*, Bull. Amer. Math. Soc. **70** (1964), 803–807.
22. L. Hörmander, *Hypoelliptic second-order differential equations*, Acta Math. **119** (1967), 147–171.
23. ——, *The spectral function of an elliptic operator*, Acta Math. **121** (1968), 192–218.
24. ——, *Subelliptic operators*, Seminar on Singularities of Solutions of Linear Partial Differential Equations, Princeton Univ. Press, Princeton, N.J., 1979.
25. D. Iagolnitzer, Appendix: *Microlocal essential support of a distribution and decomposition theorems—an introduction*, Lecture Notes in Math., Vol. 449, Springer-Verlag, 1975, pp. 121–132.
26. J. J. Kohn, *Harmonic integrals on strongly pseudoconvex manifolds*. I, II, Ann. of Math. (2) **78** (1963), 206–213; ibid. **79** (1964), 450–472.

27. ——, *Subellipticity of the $\bar{\partial}$-Neumann problem on pseudoconvex domains: sufficient conditions*, Acta Math. **142** (1979), 79–122.
28. A. Kolmogorov, *Zufällige Bewegungen*, Ann. of Math. (2) **35** (1934), 116–117.
29. R. Lee and R. Melrose, to appear.
30. E. Lieb, *Stability of matter*, Rev. Modern Phys. **48** (1976), 553–569.
31. R. Moyer, *On the Nirenberg-Treves condition for local solvability*, J. Differential Equations **26** (1977), 223–239.
32. A. Nagel, E. M. Stein and S. Wainger, to appear.
33. L. Nirenberg and F. Trèves, *Solvability of a first-order linear partial differential equation*, Comm. Pure Appl. Math. **16** (1963), 331–351.
34. ——, *On local solvability of linear partial differential equations.* I: *Necessary conditions*; II: *Sufficient conditions*, Comm. Pure Appl. Math. **23** (1970), 1–38; 459–509.
35. O. Oleinik and E. Radkevitch, *Second-order equations with non-negative characteristic form.*
36. D. Phong, *On integral representations of the Neumann operator*, Proc. Nat. Acad. Sci. U.S.A. **76** (1979), 1554–1558.
37. L. Rothschild and E. M. Stein, *Hypoelliptic differential operators and nilpotent groups*, Acta Math. **137** (1976), 247–320.
38. A. Sanchez, *Estimates for kernels associated to some subelliptic operators*, Thesis, Princeton Univ., 1983.
39. B. Simon, to appear.
40. ——, *Functional integration and quantum physics*, Academic Press, 1979.
41. H. Weyl, *Ramifications, old and new, of the eigenvalue problem*, Bull. Amer. Math. Soc. **56** (1950), 115–139.

Department of Mathematics, Princeton University, Princeton, New Jersey 08544