Monographs in Harmonic Analysis

I. *Introduction to Fourier Analysis on Euclidean Spaces*, by E. M. Stein and G. Weiss

II. *Singular Integrals and Differentiability Properties of Functions*, by E. M. Stein

III. *Harmonic Analysis: Real-Variable Methods, Orthogonality, and Oscillatory Integrals*, by E. M. Stein

Harmonic Analysis: Real-Variable Methods, Orthogonality, and Oscillatory Integrals

ELIAS M. STEIN

with the assistance of Timothy S. Murphy

Princeton University Press
Princeton, New Jersey
1993

Prologue

Given the complexity of the matters treated here, it may be helpful to begin by giving an overview of our subject. In sketching its broad outlines, we point first to the principal analytic constructs whose study will be our chief concern. These concepts can be loosely grouped into three categories: maximal averages, singular integrals, and oscillatory integrals.

Like all deep ideas in mathematics, these have each taken several forms, displaying their versatility by adapting to the changing contexts in which they occurred. Let us briefly recall how each appeared in an early version.

Maximal averages. The simplest instance arises when we consider the family of averages of a function f on $\mathbf{R}^1$ given by $\frac{1}{2t}\int_{-t}^{t} f(x-y)\,dy$, $t > 0$, as well as the more sophisticated variant $\frac{t}{\pi}\int_{-\infty}^{\infty} \frac{f(x-y)}{y^2+t^2}\,dy$, $t > 0$, which is the Poisson integral of f. For these, the limiting behavior as $t \to 0$ is the main interest, and its deeper study is subsumed in the properties of the corresponding maximal functions.

Singular integrals. A basic object in the classical theory is the Hilbert transform $f \mapsto \text{p.v.}\frac{1}{\pi}\int_{-\infty}^{\infty} f(x-y)\frac{dy}{y}$. Its indispensable role there is partly explained by the fact that it stands squarely at the crossroads linking real variables and complex function theory.

Oscillatory integrals. Here the primordial example is the Fourier transform $f \mapsto \int_{-\infty}^{\infty} e^{-2\pi i x\cdot\xi} f(x)\,dx$. Of course, when thinking of it, we should also have in mind its n-dimensional form, as well as the oscillatory integrals arising from this by symmetry considerations, such as Bessel functions.

Now it was already understood early that these three concepts were, to a substantial degree, intertwined. Thus the fundamental L^2 estimate for the Hilbert transform was seen as a simple consequence of the use of the Fourier transform, and the weak-type (1,1) estimate was originally proved by using properties of the Poisson integral mentioned above.

What could not be guessed then, and could only be revealed with the passage of time, were the wider and deeper interconnections inherent in these examples and their successive generalizations and refinements. The insights that this yielded provide the foundations of a theory of vast scope and utility that has developed over the last thirty years, spurred by its application to such parts of analysis as partial differential equations, several complex variables, and harmonic analysis related to semisimple Lie groups and symmetric spaces.

While the theory encompassing these ideas does not admit a brief summary, we do wish to touch on some of its main themes.

(i) *The underlying real-variable structure.* A central role in the analysis of maximal functions and singular integrals is played by the covering lemmas of Vitali and Whitney types. While this was first understood in the context of $\mathbf{R}^n$ (with its usual translation and dilation structure), significant parts of these results can be extended to much more general settings, where the analogues of these lemmas continue to hold. Moreover, as it turned out, more refined versions of the older results could be proved by examining further the techniques based on these covering arguments.

(ii) *Hardy space theory.* We comment first on the ubiquitous nature of the L^p spaces, $1 < p < \infty$. First, the pervasiveness of L^2 estimates is a basic fact of analysis, given the essential part played by the Fourier transform and other devices involving orthogonality. Second, while it might have been simpler to limit considerations to L^1 and L^∞ estimates, long experience has shown that deep and interesting assertions of this kind rarely hold. Thus the function of L^p is twofold: as a compromise of the possible; but more importantly, that the analysis it requires often reveals fundamental properties of the operators in question.

Now it is exactly with the failure of L^1 and L^∞ that Hardy space theory may be thought to begin. Originally developed in the context of one complex variable with a different emphasis in mind, in its modern incarnation this topic represents a happy culmination of the study of maximal functions and singular integrals by real variable methods. Not only does it yield a rich H^1 theory, making up for many of the shortcomings of L^1, but it also gives us a fruitful H^p theory in the case $p < 1$, where L^p was entirely barren. That H^p would seem destined to be of further interest in the future can be guessed from the fact that the most common "singularities" in analysis, such as those given by rational functions, or carried on analytic subvarieties, or representable by Fourier integral ("Lagrangian") distributions, are all locally in H^p, for some $p \leq 1$.

(iii) *More extended singular integrals.* The singular integrals alluded to so far have all been of the form

$$(Tf)(x) = \int K(x,y)\, f(y)\, dy,$$

where the singularity of the kernel $K(x,y)$ is concentrated in y near x. A significant departure of the current theory is that it can begin to come to grips with the situation that arises when the singularity is now "spread out", say for y in some variety Σ_x. When an analysis in this context is possible, orthogonality again plays a key role, sometimes via the Fourier transform, but more often using other oscillatory integrals. An important observation is that, at bottom, what makes this possible is some sort of "curvature" property of the family $\{\Sigma_x\}$. In this setting, analogues of maximal functions arise by taking averages over (proper) submanifolds of $\mathbf{R}^n$. Again, curvature properties play a decisive role in their study.

(iv) *Oscillatory integrals.* As indicated above, oscillatory integrals provide a necessary tool in exploiting the geometric properties related to curvature and orthogonality in the more extended maximal operators and singular integrals that have arisen. However, these oscillatory integrals, and others of interest, are not easily classified and come in a multiplicity of forms: variants of the Fourier transform, convolution operators (such as Bochner-Riesz means), and Fourier integral operators are among these forms. What is clear is that this part of the theory is in its infancy, and much more remains to be understood.

(v) *Heisenberg group.* The study of the Heisenberg group illustrates a number of essential ideas treated in this book. In particular, it gives an excellent example of the real-variable structure mentioned above; connected with this is the Cauchy-Szegö projection operator, which is a naturally occurring instance of a singular integral in this general context. In addition, we might point out that inherent in its structure is the notion of "twisted convolution"; it accounts for the composition formula for pseudo-differential operators (in their symmetric form) and also yields important examples of oscillatory singular integrals. But beyond these didactic uses the significance of the Heisenberg group resides in what it has allowed us to do, namely, to explore the way into the broader applications of our subject to such interesting areas as several complex variables and (subelliptic) partial differential equations.

CHAPTER I

Real-Variable Theory

We begin by setting down some of the fundamental real-variable ideas behind the theory of the maximal operator and the boundedness of singular integrals. To proceed here requires that our underlying space be endowed with a certain kind of metric structure. The model for this is $\mathbf{R}^n$, equipped with its usual family of Euclidean balls, which is the setting appropriate for the standard translation-invariant theory.† In fact, by abstracting some simple and basic features of this case (connected with the covering lemmas of Vitali and Whitney), a number of key points of the earlier development can be carried out in a much broader context. The following additional comments may be helpful in placing the subject of this chapter in its proper perspective.

(i) We prove here the weak-type (1,1) and L^p inequalities for the maximal operator in the generality alluded to above. We also deal with the corresponding facts for singular integrals. However, for the latter our results are of a conditional nature, since they depend on an additional assertion (essentially the L^2 boundedness) that must be treated separately. In the translation-invariant case, this is exactly where the Fourier transform is decisive. In our general context, other notions must also come into play, but consideration of these aspects is postponed until they are systematically taken up in chapters 6 and 7.

(ii) What will be even clearer (in later chapters) is that maximal operators and singular integrals can be thought of as part of a threefold unity, in that these two operators are intimately tied to another construct, namely that of square functions. One way to realize this unity is to consider all three as singular integrals, but now as vector-valued versions taking their values in differing Banach spaces.

(iii) When we continue beyond this chapter we shall not feel constrained by the requirement to present matters in the generality used here. Instead, for simplicity of exposition, we shall usually content ourselves with the standard setting of $\mathbf{R}^n$, and invoke the general theory only when needed in particular circumstances. Instances where the general point of view plays an important role are the weighted inequali-

† As developed in, e.g., *Singular Integrals*, chapters 1 and 2.

ties arising in Chapter 5, the maximal functions and singular integrals associated with lower dimensional varieties treated in Chapter 11, the extension of the theory to the Heisenberg group and other nilpotent groups dealt with in chapters 12 and 13, and several further applications sketched in §8 below.

1. *Basic assumptions*

The basic metric notions we shall be interested in have to do with the possibility of measuring the order of magnitude of "size" (or distance), and the order of magnitude of "volume". The situations we envisage will be general enough so that these two quantities will have to be taken, to a degree, independent of each other. As a reflection of this, we shall quantify these notions in terms of different objects: size in terms of a family of "balls", and volume in terms of a Borel measure.

1.1 Our considerations will always take place in the coordinate space $\mathbf{R}^n$.[‡] We shall assume we are given, for each $x \in \mathbf{R}^n$, a collection $\{B(x,\delta)\}_\delta$ of nonempty open, bounded subsets of $\mathbf{R}^n$, parameterized by δ, $0 < \delta < \infty$; that is, $B = B(x,\delta)$ is the "ball", "centered" at x of "radius" δ. We shall suppose that the balls are monotonic in δ in the sense that $B(x,\delta_1) \subset B(x,\delta_2)$ whenever $\delta_1 < \delta_2$. We shall also assume that we are given a nonnegative Borel measure μ with the property that $\mu(\mathbf{R}^n) > 0$.

The basic properties concerning the family of balls and the measure that we postulate are as follows: We assume there exist constants c_1 and c_2, both greater than 1, so that, for all x, y, and δ,

(i) $B(x,\delta) \cap B(y,\delta) \neq \emptyset$ implies $B(y,\delta) \subset B(x, c_1\delta)$.

(ii) $\mu(B(x,c_1\delta)) \leq c_2\mu(B(x,\delta))$.

Statement (i) guarantees the engulfing property crucial in Vitali-type covering lemmas, while assumption (ii) represents the fact that μ is a "doubling" measure, which allows one to exploit the first statement.[†]

1.2 In some circumstances it is convenient to substitute for the assumptions (i) and (ii) a weaker one, which we describe as follows. For each $B = B(x,\delta)$, let $B^* = \bigcup B_1$ where B_1 ranges over all balls of radius δ meeting B. The assumption that replaces (i) and (ii) is then:

(i,ii)* $\mu(B^*(x,\delta)) \leq c_2\mu(B(x,\delta))$.

It is obvious that (i) and (ii) imply (i,ii)*.

[‡] This is mostly a matter of notational convenience. By making slight changes, we can replace $\mathbf{R}^n$ by manifolds or more general spaces. See §8.1 below.

[†] Note that (ii) is equivalent with the inequality $\mu(B(x,2\delta)) \leq c_2'\mu(B(x,\delta))$, from which the terminology "doubling" originates.

1.3 Further assumptions. In addition to the above basic properties, it will be convenient to postulate two other properties about the balls $B(x,\delta)$ and the measure μ. These further assumptions, while not always essential, allow us to avoid certain technical complications. The first will always be assumed in what follows and is in any case easily verifiable in each particular example.

(iii) $\bigcap_\delta \overline{B}(x,\delta) = \{x\}$ and $\bigcup_\delta B(x,\delta) = \mathbf{R}^n$.

(iv) For each open set U and each $\delta > 0$, the function $x \mapsto \mu(\{B(x,\delta) \cap U\})$ is continuous.

1.4 Let us remark that these additional properties easily lead to the following conclusions, among others. First note that $\mu(B) > 0$, for any ball B, which is a consequence of the doubling property, (iii), and the fact that $\mu(\mathbf{R}^n) > 0$. It then follows from (iv) that for any locally integrable f, and any $\delta > 0$, the mean value

$$(A_\delta f)(x) = \frac{1}{\mu(B(x,\delta))}\int_{B(x,\delta)} f(y)\,d\mu(y)$$

is a continuous function of x.

The analysis of the averages $A_\delta(f)$, and in particular the question whether $A_\delta(f) \to f$ almost everywhere as $\delta \to 0$, is intimately connected with the properties of the *maximal function*, defined by

$$(Mf)(x) = \sup_{\delta>0} A_\delta(|f|)(x).$$

This basic real-variable construct (in the standard setting of $\mathbf{R}^n$) was introduced by Hardy and Littlewood for $n = 1$, and by Wiener for general n.

Before proceeding further with the study of maximal functions and covering lemmas, we will illustrate the nature of our postulates by detailing several examples that will give us a better idea of the scope of our assumptions.

2. *Examples*

2.1 We begin by remarking that the standard Euclidean balls, defined by $B(x,\delta) = \{y : |y - x| < \delta\}$, satisfy all of the above properties. Here we take $d\mu$ to be the Euclidean measure dx. The same is true of the following (equivalent) variant. Let B_1 be a fixed open subset of $\mathbf{R}^n$ that is bounded and is star-shaped with respect to the origin (so that dilates are increasing). Define

$$\tilde{B}(x,\delta) = x + \delta B_1 = \{y : (y - x)/\delta \in B_1\}.$$

Then the family $\{\tilde{B}(x,\delta)\}$ also satisfies all the above properties.

The idea of equivalence we have just used can be defined more generally. We will say that two different families $\{B(x,\delta)\}$ and $\{\tilde{B}(x,\delta)\}$ are *equivalent* if there exists a $c > 1$ so that

$$B(x,c^{-1}\delta) \subset \tilde{B}(x,\delta) \subset B(x,c\delta) \quad \text{for all } x \text{ and } \delta.$$

2.2 The next construction leads to a general class of examples. Suppose for each $\delta > 0$ we are given B_δ, an open, convex, symmetric ($x \in B_\delta \Leftrightarrow -x \in B_\delta$), and bounded subset of $\mathbf{R}^n$ with $B_{\delta_1} \subset B_{\delta_2}$ when $\delta_1 < \delta_2$, $\bigcap_{\delta>0} B_\delta = \{x\}$, and $\bigcup_{\delta>0} B_\delta = \mathbf{R}^n$. Set $B(x,\delta) = x + B_\delta = \{y : y - x \in B_\delta\}$ and take $d\mu$ to be Euclidean measure. Let us observe that while the assumptions (i) and (ii) may not be satisfied in general, the substitute property (i,ii)* does hold in all cases. In fact, if we take into account the convexity and symmetry of the B_δ, then it follows easily that $B^*(x,\delta) \subset 3B(x,\delta)$, giving $\mu(B^*(x,\delta)) \le 3^n\mu(B(x,\delta))$.

2.3 An important class of examples, actually subsumed under §2.2, arises if we consider *nonisotropic dilations* of $\mathbf{R}^n$. That is, we fix a sequence $a_1,\dots,a_n$ of strictly positive exponents, and consider the dilations

$$\delta : (x_1,\dots,x_n) \mapsto (\delta^{a_1}x_1,\dots,\delta^{a_n}x_n).$$

Here we take $B_\delta = \{y : \max_k |y_k|^{1/a_k} < \delta\}$, with $B(x,\delta) = x + B_\delta$, and $d\mu = dx$; equivalent situations would be obtained by replacing the rectangular boxes B_δ by the tilted boxes $B'_\delta = \{y : \sum_k \delta^{-a_k}|y_k| < 1\}$ or the ellipsoids $B''_\delta = \{y : \sum_k \delta^{-2a_k}|y_k|^2 < 1\}$. It is not difficult to verify that (i), (ii) of §1.1 and (iii), (iv) of §1.3 hold for these balls. For us, the main application of this structure will occur in Chapter 11 when we study maximal averages taken over k-dimensional submanifolds.

2.4 A general method of constructing families $\{B(x,\delta)\}$ is in terms of a *quasi-distance* defined on $\mathbf{R}^n$. By this we mean a nonnegative function ρ on $\mathbf{R}^n \times \mathbf{R}^n$ for which there exists a positive constant c so that

$$\begin{aligned} &\rho(x,y) = 0 \quad \Leftrightarrow \quad x = y \\ &\rho(x,y) \le c\rho(y,x) \\ &\rho(x,y) \le c(\rho(x,z) + \rho(y,z)) \end{aligned} \tag{1}$$

where x, y, and z are arbitrary points in $\mathbf{R}^n$; we also assume that ρ is upper semicontinuous in the first variable.

Given such a ρ, we can define the $B(x,\delta)$ by

$$B(x,\delta) = \{y : \rho(y,x) < \delta\}. \tag{2}$$

To show (i), assume that $B(x,\delta) \cap B(y,\delta) \neq \emptyset$. By the quasi-triangle inequality, $\rho(y,x) \le 2c\delta$. If $w \in B(y,\delta)$ (i.e., $\rho(w,y) \le \delta$), applying (1)

again gives $\rho(w,x) \le c(2c\delta + c\delta) = 3c^2\delta$, which proves (i) with $c_1 = 3c^2$. Conversely, we may define

$$\rho(y,x) = \inf\{\delta : y \in B(x,\delta)\};$$

similar arguments show that (i) and (iii) imply (1).

By replacing the quasi-distance $\rho(x,y)$ by the equivalent quasi-distance $\frac{1}{2}(\rho(x,y) + \rho(y,x))$, we can assume $\rho(y,x) = \rho(x,y)$. We shall always do this in the future.

Observe that the $B(x,\delta)$ given in §2.3 are defined by

$$\rho(x,y) = \max_k |x_k - y_k|^{1/a_k},$$

while the $B'(x,\delta)$ are determined by

$$\rho'(x,y) = \sum_k |x_k - y_k|^{1/a_k}.$$

2.5 The examples given in §2.1–§2.3 are invariant with respect to the usual translations of the Euclidean space $\mathbf{R}^n$. An interesting extension of this is to replace the Euclidean translations $x - y$ by the more general (noncommutative) group translations $y^{-1} \cdot x$. Here we assume that $\mathbf{R}^n$ is the underlying manifold of a group; the dot denotes multiplication and y^{-1} is the inverse of y. We then set $B(x,\delta) = x \cdot B_\delta$, with B_δ as in §2.3. If we make the key assumption that the dilations

$$\delta : (x_1,\dots,x_n) \mapsto (\delta^{a_1}x_1,\dots,\delta^{a_n}x_n)$$

are automorphisms of the group, then all of our postulates (i)–(iv) hold when $d\mu$ is Lebesgue measure. This is the notion of a *homogeneous* group of which the key example is the Heisenberg group. These matters are taken up in chapters 12 and 13.

2.6 Finally we describe an example that has a simple geometric interpretation, but for which the underlying space has no features of translation invariance or homogeneity. In $\mathbf{R}^2$ we consider the two vector fields $X_1 = \partial/\partial x_1$ and $X_2 = x_1^k\,\partial/\partial x_2$, where k is a nonnegative integer. A natural family of balls $B(x,\delta)$ associated with these vector fields may be defined as follows:

$$y \in B(x,\delta) \;\text{ if }\; |x_1 - y_1| < \delta \text{ and } \begin{cases} |x_2 - y_2| < \delta^{k+1} & \text{when } |x_1| < \delta \\ |x_2 - y_2| < \delta|x_1|^k & \text{when } |x_1| \ge \delta \end{cases}$$

where $x = (x_1,x_2), y = (y_1,y_2)$. If we take $d\mu = dx$, this family verifies all our assumptions above. We can interpret this example in the following (equivalent) way. One has $y \in B(x,\delta)$ if one can join y to x in elapsed time $\le c\delta$ along a path whose velocity vector at any point is of the form $a_1X_1 + a_2X_2$, with $|a_1| \le 1$ and $|a_2| \le 1$. For a general formulation of this, and for further details, see §8.4.

2.7 In all of our examples we have taken $d\mu = dx$; of course, many other choices could have been made for the measure $d\mu$. In particular, when $\{B(x,\delta)\}$ are the standard Euclidean balls discussed in §2.1, then, in order to satisfy the postulates in §1.1 and §1.3, the only restriction on $d\mu$ is the doubling condition $\mu(B(x,2\delta)) \leq c\mu(B(x,\delta))$. Note that (iv) in §1.3 is automatically satisfied (see §8.6 below). We remark that $d\mu(x) = |x|^\alpha dx$ is a doubling measure when $\alpha > -n$, but $d\mu(x) = e^{c|x|}\,dx$, $c \neq 0$, is not. Other less trivial examples are in §8.7–§8.9 below.

3. *Covering lemmas and the maximal function*

Having familiarized ourselves with several families of balls $\{B(x,\delta)\}$ and their associated measures $d\mu$, we come to one of the main points justifying the above conditions: certain covering lemmas of Vitali and Whitney types, and the consequences that can be deduced from them.

3.1 We consider first the simplest of these covering lemmas, a finite version of the Vitali lemma. For it, we need only the postulates in the weaker form (i,ii)*, instead of (i), (ii).

LEMMA 1. *Let E be a measurable subset of $\mathbf{R}^n$ that is the union of a finite collection of balls $\{B_j\}$. Then one can select a disjoint subcollection $B_1,\ldots,B_m$ of the $\{B_j\}$ so that*

$$\sum_{k=1}^m \mu(B_k) \geq c\mu(E).$$

Here c is a positive constant, which we can take to be $c = c_2^{-1}$ (with c_2 as in §1.1 or §1.2).

Proof. Let B_1 be a ball of the collection $\{B_j\}$ of maximal radius. Next choose B_2 to have maximal radius among the subcollection of balls disjoint with B_1. We continue this process until we can go no further. From this it is clear that the chosen subcollection $B_1, B_2, \ldots, B_m$ consists of disjoint balls. Recalling the definition of the ball B_1^* (see §1.2), we observe that B_1^* contains all balls of the original collection that intersect B_1 and whose radii are at most as large as that of B_1. Similarly B_k^* contains all the remaining balls that intersect B_k and whose radii are at most as large as that of B_k. From this it follows that $\bigcup_{k=1}^m B_k^*$ contains the union of the initial collection of balls. Thus $\sum_k \mu(B_k^*) \geq \mu(\bigcup_k B_k^*) \geq \mu(E)$. Since $\mu(B_k^*) \leq c_2\mu(B_k)$, the lemma follows.

The lemma we have just proved allows us to obtain the fundamental results about the averages $A_\delta(f)$ and the maximal function $M(f) = \sup_{\delta>0} A_\delta(|f|)$, which we defined in §1.4. Here we use the notation L^p to stand for the space $L^p(\mathbf{R}^n, d\mu)$, taken with respect to our measure $d\mu$, $\|\cdot\|_p$ will designate the norm in this space, and "almost everywhere" means except for a set of μ-measure zero.

THEOREM 1. *Let f be a function defined on $\mathbf{R}^n$.*

(a) *If $f \in L^p$, $1 \leq p \leq \infty$, then $M(f)$ is finite almost everywhere.*

(b) *If $f \in L^1$, then for every $\alpha > 0$,*

$$\mu(\{x : (Mf)(x) > \alpha\}) \leq \frac{c_2}{\alpha}\int_{\mathbf{R}^n} |f(y)|\,d\mu(y)$$

(c) *If $f \in L^p$, $1 < p \leq \infty$, then $M(f) \in L^p$ and*

$$\|M(f)\|_p \leq A_p\|f\|_p$$

where the bound A_p depends only on c_2 and p.

Remark. It is easy to see that the maximal operator is not bounded as a mapping from L^1 to itself; i.e., the $p = 1$ analogue of conclusion (c) is false. Connected with this is the estimate $A_p = O(1/(p-1))$, as $p \to 1$, for the bound in (c). The situation is similar for the singular integrals taken up below. Further details are in §8.13 and §8.14.

COROLLARY. *If f is locally integrable with respect to $d\mu$, then*

$$\lim_{\delta\to 0}(A_\delta f)(x) = f(x)$$

for almost every x.

Proof. We shall prove the inequalities stated for M by showing that they hold for the larger "uncentered" maximal function $\tilde{M}f$ defined by

$$(\tilde{M}f)(x) = \sup_{x\in B} \frac{1}{\mu(B)}\int_B |f(y)|\,d\mu(y),$$

where the supremum is taken over all balls B containing x. Note of course that $(Mf)(x) \leq (\tilde{M}f)(x)$. The maximal operator $\tilde{M}$ will also have a useful role below. Incidentally, $(\tilde{M}f)$ is automatically lower semicontinuous and, because of this, axiom (iv) in §1.3 is obviated in most of what follows. Finally we remark that, by axiom (i,ii)*, $(\tilde{M}f)(x) \leq c_2(Mf)(x)$, since $B(y,\delta) \subset B^*(x,\delta)$ if $x \in B(y,\delta)$.

We first prove conclusion (b). Let $E_\alpha = \{x : \tilde{M}f(x) > \alpha\}$ and let $E \subset E_\alpha$ be any compact subset. By definition of E_α, for each $x \in E$, there exists a ball B_x so that $x \in B_x$ and

$$\mu(B_x) < \frac{1}{\alpha}\int_{B_x} |f(y)|\,d\mu(y). \tag{3}$$

Since $x \in B_x$, by the compactness of E we can select a finite collection of these balls that cover E. The lemma allows us to select a disjoint subcollection $B_1, \ldots, B_m$ of this covering with $\mu(E) \le c_2 \sum_{k=1}^{m} \mu(B_k)$. Since each ball B_k satisfies (3), adding these inequalities shows that $\mu(E) \le (c_2/\alpha)\int |f(y)|\,d\mu(y)$. If one takes the supremum over all such $E \subset E_\alpha$, the conclusion (b) is proved.

Conclusions (a) and (c) of the theorem and the corollary follow from (b) in a well known manner. Indeed, note that $\tilde{M}(f) \le \tilde{M}(f_1) + \alpha/2$, where $f_1(x) = f(x)$ if $|f(x)| > \alpha/2$, and $f_1(x) = 0$ if $|f(x)| \le \alpha/2$. Thus $\{\tilde{M}(f) > \alpha\} \subset \{\tilde{M}(f_1) > \alpha/2\}$, and we get

$$\mu(\{x : \tilde{M}f(x) > \alpha\}) \le \frac{c}{\alpha}\int_{\{|f| > \frac{\alpha}{2}\}} |f|\,d\mu.$$

However,

$$\int (\tilde{M}f)^p\,d\mu = p\int_0^\infty \mu(\{\tilde{M}f > \alpha\})\alpha^{p-1}\,d\alpha,$$

which, by the above, is majorized by

$$cp\int\Big(\int_0^{2|f|} \alpha^{p-2}\,d\alpha\Big)|f|\,d\mu = \frac{cp}{p-1}2^{p-1}\int |f|^p\,d\mu,$$

proving inequality (c).

To prove the corollary it is useful to remark that since by our assumptions we have that $\bigcap_{\delta > 0} \overline{B}(x,\delta) = \{x\}$, and the $\overline{B}(x,\delta)$ are compact, then for each x the (Euclidean) diameters of $B(x,\delta)$ tend to zero as $\delta \to 0$; which implies that $(A_\delta f)(x) \to f(x)$, whenever f is continuous at x.

3.2 We now turn from the study of maximal functions to the closely related theory of singular integrals. This requires that we change our point of view slightly, and from now on we shall assume the stronger versions of our postulates (namely those of §1.1 rather than those of §1.2). This is needed for our next main step: an extension of the basic covering idea of Whitney, which has the following formulation.

Suppose F is a nonempty closed set and $O = {}^cF$ is its complement. We can "cover" O by a collection of balls that are essentially disjoint, and whose sizes are comparable to their distances from the set F. In the standard setting of $\mathbf{R}^n$, the actual covering of O that can be achieved is by closed cubes whose interiors are disjoint and whose side lengths are comparable to their distances from the set F.‡

In the general setting we consider here, things are not quite as elegant and we need to begin by fixing a pair of positive constants c^* and c^{**} (with $1 < c^* < c^{**}$), which will depend only on the structural constant involved in our basic postulates (i.e., the quantity c_1 appearing in assumption (i)), and not on the particular set F under consideration. The values of c^* and c^{**} will be made precise below. Using them, for any ball B we define the balls B^* and B^{**} that have the same "centers" as B but whose "radii" are expanded by the factors c^* and c^{**} respectively. That is, if $B = B(x,\delta)$ then $B^* = B(x, c^*\delta)$ and $B^{**} = B(x, c^{**}\delta)$. Clearly, $B \subset B^* \subset B^{**}$.†

LEMMA 2. *Given F, a closed nonempty set, there exists a collection of balls $B_1, \ldots, B_k, \ldots$ so that*

(a) *The B_k are pairwise disjoint.*

(b) $\bigcup_k B_k^* = O = {}^cF$.

(c) $B_k^{**} \cap F \neq \emptyset$ *, for each k.*

Remarks. (i) It is easy to construct from the above collection $\{B_k\}$ a collection of sets $\{Q_k\}$ so that the Q_k are disjoint, $B_k \subset Q_k \subset B_k^*$, and $\bigcup_k Q_k = O$. Take, for example,

$$Q_k = B_k^* \cap \big({}^c\bigcup_{j<k} Q_j\big) \cap \big({}^c\bigcup_{j>k} B_j\big), \quad k = 1, 2, \ldots$$

These Q_k can be taken as substitutes for the standard cubes that appear in the usual Whitney lemma.

(ii) One can also show that, while the B_k^* are not necessarily disjoint, they do have the bounded intersection property. See §7.1 below.

Proof of Lemma 2. We begin by choosing ε to be sufficiently small; later we will see that $\varepsilon = 1/8c_1^2$ will do. With ε fixed, we consider the covering $\{B(x, \varepsilon\delta(x))\}_{x \in O}$ of O, where $\delta(x)$ is the "distance" of x from F, that is, $\delta(x) = \sup\{\delta : B(x,\delta) \subset O\}$. That for each $x \in O$ the function $\delta(x)$ is strictly positive and finite follows from axiom (iii) in §1.3.

‡ See, for instance, the lemma on p. 16 of *Singular Integrals.*

† Note that the B^* defined here are different from those occurring in §1.2.

We now select a maximal disjoint subcollection of $\{B(x,\varepsilon\delta(x))\}_{x\in O}$; for this subcollection $B_1,\ldots,B_k,\ldots$ with $B_k = B(x_k,\varepsilon\delta(x_k))$, we shall prove the assertions (a), (b), and (c) above. We define

$$B_k^* = B(x_k,\delta(x_k)/2), \qquad B_k^{**} = B(x_k, 2\delta(x_k)),$$

(so we have set $c^* = 1/2\varepsilon$, $c^{**} = 2/\varepsilon$). Note that (a) and (c) hold automatically by our choice of B_k. It is also clear that $B_k^* \subset O$; what remains to be shown is that $\bigcup_k B_k^* \supset O$.

Now let $x \in O$; then, by the maximality of the collection $\{B_k\}$,

$$B\big(x_k,\varepsilon\delta(x_k)\big) \cap B\big(x,\varepsilon\delta(x)\big) \neq \emptyset, \quad \text{for some } k.$$

We claim that $\delta(x_k) \geq \delta(x)/4c_1$. If not, taking $\varepsilon < 1/2c_1 (< 1)$, we have

$$B\big(x_k,\delta(x_k)\big) \cap B\Big(x,\frac{\delta(x)}{2c_1}\Big) \neq \emptyset.$$

Since $2\delta(x_k) < \delta(x)/2c_1$, by the engulfing property

$$B\big(x_k,2\delta(x_k)\big) \subset B\Big(x,\frac{\delta(x)}{2}\Big),$$

which gives a contradiction since $B(x_k,2\delta(x_k))$ meets $F = {}^cO$, while $B(x,\delta(x)/2) \subset O$.

Using $4c_1\varepsilon\delta(x_k) \geq \varepsilon\delta(x)$ and the engulfing property again gives

$$x \in B\big(x_k, c_1 \cdot 4c_1\varepsilon\delta(x_k)\big).$$

We take $B(x_k, c_1\cdot 4c_1\varepsilon\delta(x_k)) = B_k^* = B(x_k,\delta(x_k)/2)$; i.e., $c^* = 4c_1^2$, $\varepsilon = 1/2c^* = 1/8c_1^2$, $c^{**} = 4c^* = 16c_1^2$, finishing the proof.

4. Generalization of the Calderón-Zygmund decomposition

The Calderón-Zygmund decomposition is a key step in the real-variable analysis of singular integrals. The idea behind this decomposition is that it is often useful to split an arbitrary integrable function into its "small" and "large" parts, and then use different techniques to analyze each part.

The scheme is roughly as follows. Given a function f and an altitude α, we write $f = g + b$, where $|g|$ is pointwise bounded by a constant multiple of α. While b is large, it does enjoy two redeeming features: it is supported in a set of reasonably small measure, and its mean value is zero on each of the balls that constitute its support. To obtain the decomposition $f = g+b$, one might be tempted to "cut" f at the height α; however, this is not what works. Instead, one bases the decomposition on the set where the *maximal function* of f has height α.

In the general context we are concerned with, the decomposition can be formulated as follows:

THEOREM 2.‡ *Suppose we are given a function $f \in L^1$ and a positive number α, with $\alpha > \dfrac{1}{\mu(\mathbf{R}^n)}\displaystyle\int_{\mathbf{R}^n}|f|\,d\mu$.*† *Then there exists a decomposition of f, $f = g + b$, with $b = \sum_k b_k$, and a sequence of balls $\{B_k^*\}$, so that*

(i) $|g(x)| \leq c\alpha$, *for a.e. x.*

(ii) *Each b_k is supported in B_k^*,*

$$\int |b_k(x)|\,d\mu(x) \leq c\alpha\mu(B_k^*), \quad \textit{and} \quad \int b_k(x)\,d\mu(x) = 0.$$

(iii) $\displaystyle\sum_k \mu(B_k^*) \leq \frac{c}{\alpha}\int |f(x)|\,d\mu(x)$.

4.1 Proof. Let $E_\alpha = \{x : \tilde{M}f(x) > \alpha\}$, where $\tilde{M}$ is the uncentered maximal function defined in §3. E_α is an open set, and we consider first the case when its complement is nonempty.

We can apply the lemma of §3.2 (and the remarks that follow it) to $O = E_\alpha$. Thus we obtain collections of balls $\{B_k\}$, $\{B_k^*\}$, and "cubes" $\{Q_k\}$, so that

$$B_k \subset Q_k \subset B_k^*, \quad \text{with } \bigcup_k Q_k = E_\alpha, \tag{4}$$

and where the Q_k are mutually disjoint. It follows immediately that

$$\sum_k \mu(B_k) \leq \mu(E_\alpha). \tag{5}$$

Now define $g(x) = f(x)$ for $x \notin E_\alpha$, and

$$g(x) = \frac{1}{\mu(Q_k)}\int_{Q_k} f(y)\,d\mu(y), \quad \text{if } x \in Q_k.$$

Hence $f = g + \sum b_k$, where

$$b_k(x) = \chi_{Q_k}\cdot\left[f(x) - \frac{1}{\mu(Q_k)}\int_{Q_k} f(y)\,d\mu(y)\right], \tag{6}$$

with χ_{Q_k} denoting the characteristic function of Q_k.

‡ Compare with the classical version on pp. 17–20 and p. 31 of *Singular Integrals*. See also §2.1 of Chapter 3 and §3.1 of Chapter 4 below.

† Of course, this assumption is vacuous if $\mu(\mathbf{R}^n) = \infty$.

Because of the corollary in §1.5 (the differentiation theorem), we have $|f(x)| \le \alpha$ for a.e. $x \in {}^c\bigcup Q_k = \{x : \tilde{M}f(x) \le \alpha\}$. So $|g(x)| \le \alpha$ for $x \in {}^c\bigcup Q_k$. Next, we observe that

$$\frac{1}{\mu(B_k^{**})} \int_{B_k^{**}} |f(x)|\, d\mu(x) \le \alpha \tag{7}$$

because the ball B_k^{**} intersects ${}^cE_\alpha$.

So from (7), and the fact that $B_k \subset Q_k \subset B_k^{**}$, it follows that $|g(x)| \le \bar{c}\alpha$, whenever $x \in Q_k$. Thus (i) is proved. That b_k is supported in B_k^* is a consequence of the inclusion $Q_k \subset B_k^*$. Also

$$\int |b_k(x)|\, d\mu(x) \le 2 \int_{Q_k} |f(x)|\, d\mu(x) \le c\alpha\mu(B_k^*)$$

by (7) and the doubling property. Moreover the assertion $\int b_k(x)\, d\mu(x) = 0$ is obvious from (6); thus conclusion (ii) is proved. Again by the doubling property, $\sum \mu(B_k^*) \le c\mu(\{\tilde{M}f > \alpha\})$ because of (5), and the quantity on the right is majorized by $(c/\alpha)\int |f|\, d\mu$, as we see if we invoke the maximal theorem of §3.1. With this the proof of Theorem 2 is concluded, under the assumption $\{x : \tilde{M}f(x) \le \alpha\} \ne \emptyset$.

If we now consider the special situation where $\{x : \tilde{M}f(x) > \alpha\} = \mathbf{R}^n$ (which can happen only when $\mu(\mathbf{R}^n) < \infty$), then we see by the maximal theorem that

$$\mu(\mathbf{R}^n) \le \frac{c}{\alpha} \int_{\mathbf{R}^n} |f|\, d\mu.$$

We then obtain the decomposition $f = g + b_1$, with

$$g = \frac{1}{\mu(\mathbf{R}^n)} \int_{\mathbf{R}^n} f\, d\mu,$$

$b_1 = f - g$; here b_1 is supported in the "ball" $B_1^* = \mathbf{R}^n$. Our assumption

$$\alpha > \frac{1}{\mu(\mathbf{R}^n)} \int_{\mathbf{R}^n} |f|\, d\mu$$

guarantees that $|g| \le \alpha$.

5. *Singular integrals*

The main result for singular integrals we shall prove in this chapter is a conditional one, guaranteeing the boundedness in L^p, for p in the

range $1 < p \le q$, on the assumption that the boundedness in L^q is already known. The singular integrals one is interested in are operators T, expressible in the form

$$(Tf)(x) = \int_{\mathbf{R}^n} K(x,y) f(y)\, d\mu(y), \tag{8}$$

where the kernel K is singular near $x = y$, and so the expression (8) is meaningful only if K is treated as a distribution or in some limiting sense. Now the particular regularization of (8) that may be appropriate depends much on the context, and a complete treatment of the issues thereby raised could take us quite far afield.

Here we shall limit ourselves to two closely related ways of dealing with the questions concerning the definability of the operator. One is to prove estimates for the (dense) subspace where the operator is initially defined. The other is to regularize the given operator by replacing it with a suitable family, and to prove uniform estimates for this family. The most common method of regularizing these operators is by truncation, and ideas relevant to this are detailed in §7. Common to both methods is the *a priori* approach: We assume some additional properties of the kernel, but then prove estimates that are independent of these "regularity" properties.

We now carry out the first approach in detail. There will be two kinds of assumptions made about the operator. The first is quantitative: we assume that we are given a bound A, so that the operator T is defined and bounded on L^q with norm A; that is,

$$\|T(f)\|_q \le A\|f\|_q, \quad \text{for all } f \in L^q. \tag{9}$$

Moreover, we assume that there is associated to T a measurable function K (that plays the role of its kernel), so that for the same constant A and some constant $c > 1$,

$$\int_{{}^cB(y,c\delta)} |K(x,y) - K(x,\bar{y})|\, d\mu(x) \le A, \quad \text{whenever } \bar{y} \in B(y,\delta), \tag{10}$$

for all $y \in \mathbf{R}^n$, $\delta > 0$.

The further (regularity) assumption on the kernel K is that for each f in L^q that has compact support, the integral (8) converges absolutely for almost all x in the complement of the support of f, and that equality (8) holds for these x.

THEOREM 3. *Under the assumptions* (9) *and* (10) *made above on* K, *the operator* T *is bounded in* L^p *norm on* $L^p \cap L^q$, *when* $1 < p < q$. *More precisely,*

$$\|T(f)\|_p \le A_p\|f\|_p, \tag{11}$$

for $f \in L^p \cap L^q$ *with* $1 < p < q$, *where the bound* A_p *depends only on the constant* A *appearing in* (9) *and* (10) *and on* p, *but not on the assumed regularity of* K, *or on* f.

5.1 Proof. The key point is to prove that the mapping $f \mapsto T(f)$ is of weak-type (1,1); that is

$$\mu\{x : |Tf(x)| > \alpha\} \leq \frac{A'}{\alpha} \int |f|\, d\mu \tag{12}$$

for $f \in L^1 \cap L^q$ and $\alpha > 0$, where again the constant A' is to depend on A, but not on the other properties of K.

We take the precaution of replacing the left side of (12) by $\mu(\{x : |Tf(x)| > c'\alpha\})$, where c' is a (large) constant to be chosen momentarily. With α fixed we invoke the decomposition of f as $g + b$ given by the theorem of §4,‡ and (12) will be established once we show

$$\mu\{x : |Tg(x)| > (c'/2)\alpha\} + \mu\{x : |Tb(x)| > (c'/2)\alpha\} \leq \frac{A'}{\alpha} \int |f(x)|\, d\mu(x). \tag{13}$$

First we claim that $g \in L^q$; once we have established this we shall be able to use the assumption (9). We begin with the case $q < \infty$. Now

$$\int |g|^q\, d\mu = \int_{^c\cup B_k^*} |g|^q\, d\mu + \int_{\cup B_k^*} |g|^q\, d\mu.$$

On ${}^c\bigcup B_k^*$, $g(x) = f(x)$, so $\int_{^c\cup B_k^*} |g|^q\, d\mu \leq c\alpha^{q-1}\|f\|_1$, by conclusion (i) of Theorem 2. Moreover,

$$\int_{\cup B_k^*} |g|^q\, d\mu \leq c\alpha^q \mu(\cup B_k^*) \leq c\alpha^{q-1}\|f\|_1,$$

if we also use conclusion (iii). Therefore

$$\|g\|_q^q \leq c\alpha^{q-1}\|f\|_1, \quad \text{when } q < \infty. \tag{14}$$

Chebycheff's inequality, (9) (with f replaced by g), and (14) then give

$$\mu\{x : |Tg(x)| > (c'/2)\alpha\} \leq [(c'/2)\alpha]^{-q}\|Tg\|_q^q \leq A'\alpha^{-q}\|g\|_q^q \leq \frac{A'}{\alpha}\|f\|_1.$$

Therefore that part of (13) that involves Tg is proved, when $q < \infty$.

When $q = \infty$, observe that $\|g\|_\infty \leq c\alpha$ (again by conclusion (i) of Theorem 2). Thus, by (9) (here $q = \infty$), the set $\{x : |Tg(x)| > (c'/2)\alpha\}$ is empty if we choose $c' \geq 2Ac$; once this choice is made, the required estimate for Tg is proved in all cases.

‡ We need only consider $\alpha > \frac{1}{\mu(\mathbf{R}^n)} \int |f|\, d\mu$; otherwise (12) follows trivially, because the left side is always bounded by $\mu(\mathbf{R}^n)$.

We come now to $T(b)$. We know that $b = \sum b_k$, where each b_k is supported in the ball B_k^*. Observe that by the definition of the b_k (see (6) above), we have $b_k \in L^q$, if $f \in L^q$. Let B_k^{**} denote the ball with the same center as B_k^*, but whose radius is expanded by the factor c; here c is the same constant as is used in hypothesis (10) regarding our kernel.† This hypothesis then immediately implies that

$$\int_{^cB_k^{**}} |K(x, y) - K(x, \bar{y}^k)|\, d\mu(x) \leq A, \quad \text{if } y \in B_k, \tag{15}$$

where $\bar{y}^k$ denotes the common center of B_k^* and B_k^{**}.

However, when $x \notin B_k^{**}$ we can use the representation (8) for $T(b_k)(x)$, since b_k is supported in B_k^*. So

$$\int_{^c\cup B_k^{**}} |Tb(x)|\, d\mu(x) \leq \sum_k \int_{^cB_k^{**}} |Tb_k(x)|\, d\mu(x),$$

which is dominated by

$$\sum_k \int_{^cB_k^{**}} \left(|K(x, y) - K(x, \bar{y}^k)| \cdot \int_{B_k^*} |b_k(y)|\, d\mu(y) \right) d\mu(x).$$

Here we have used that

$$Tb_k(x) = \int [K(x, y) - K(x, \bar{y}^k)] b_k(y)\, d\mu(y),$$

since $\int b_k(y)\, d\mu(y) = 0$. Therefore, by (15) and conclusions (ii) and (iii) of Theorem 2, the last sum is majorized by $Ac\alpha \sum_k \mu(B_k^*)$, which in turn is bounded by $A' \int |f(x)|\, d\mu(x)$. As a result

$$\int_{^c\cup B_k^{**}} |Tb(x)|\, d\mu(x) \leq A' \int |f(x)|\, d\mu(x),$$

which shows that

$$\mu(\{x : |Tb(x)| > c'\alpha/2\} \cap {}^c\cup B_k^{**}) \leq \frac{A'}{\alpha} \int |f|\, d\mu.$$

However,

$$\mu(\cup B_k^{**}) \leq \sum \mu(B_k^{**}) \leq c \sum \mu(B_k^*) \leq \frac{c}{\alpha} \int |f|\, d\mu.$$

† Note that the B^{**} are not the same as those that occur in the Whitney lemma of §3.2, since in the present context they play a different role. Here the passage from B^* to B^{**} is given by the factor c, suited to the hypothesis (10). For the Whitney lemma the corresponding factor was 4, large enough to guarantee the intersection of the B^{**} with ${}^c\mathcal{O}$.

Together then, the last two estimates complete the proof of (13), and with it (12), showing that the mapping $f \mapsto T(f)$ is of weak-type (1,1). The Marcinkiewicz interpolation theorem then gives (11), as a combination of (9) and (12).[‡] If we examine the argument above, we see that the bound A' in (12), and therefore the bounds A_p in (11), do not depend on the regularity of K, but are a function of only the bound A in (9) and (10), the exponent p, and the constant c appearing in the decomposition theorem of §4. This concludes the proof of Theorem 3.

5.2 Since $L^p \cap L^q$ is a dense linear subspace of L^p (when $p < \infty$), we can use Theorem 3 to extend T to all of $L^p(\mathbf{R}^n)$, $1 < p < q$; this extension also satisfies the inequality (11), but now for all of L^p. Similarly we can extend T to L^1, and there it satisfies the weak-type inequality (12). In fact for $p > 1$, whenever $\{f_n\}$ is a sequence in $L^p \cap L^q$ that converges in L^p norm, then by (11) the sequence $\{T(f_n)\}$ is Cauchy in the L^p norm. Likewise, when $p = 1$, $T(f_n)$ converges in measure by (12). Observe that the extension of T so obtained is unique. We summarize our discussion as a corollary:

COROLLARY. *The operator T in Theorem 3 has a unique extension to all of L^p, $1 \le p < q$, that satisfies the inequalities* (11) *and* (12).

In what follows we shall also use the symbol T to denote the extension given by the corollary.

5.3 It is worthwhile to point out where the key assumption (10) is used in the proof of Theorem 3. It enters only via the inequality

$$\int_{^cB(\bar{y},c\delta)} |Tf|\, d\mu \le A \int_{B(\bar{y},\delta)} |f|\, d\mu,$$

which holds whenever f is a function supported in the ball $B(\bar{y},\delta)$ and satisfies the cancellation property

$$\int_{B(\bar{y},\delta)} f(x)\, d\mu(x) = 0.$$

This inequality will be a crucial fact in the extension of results of this kind to Hardy spaces, as we will see Chapter 3, §3.1.

5.4 We return to the interpolation theorem used in §5.1 above. The case for $q = \infty$ is already subsumed in the proof of Theorem 1

[‡] We prove this particular case of the interpolation theorem in §5.4 below. See *Fourier Analysis*, Chapter 5, §2, and *Singular Integrals*, Chapter 1, §4 (and Appendix B of *Singular Integrals*) for more complete versions.

(in §3.1), so we may assume that $q < \infty$. The main point is to observe that, whenever $f \in L^1 \cap L^q$ and $\alpha > 0$, we have

$$\mu\{x : |(Tf)(x)| > \alpha\} \le A''\Big(\alpha^{-1} \int_{|f|>\alpha} |f|\, dx + \alpha^{-q} \int_{|f|\le\alpha} |f|^q\, dx\Big),$$

where A'' depends only on the constants A and A' appearing in (9) and (12). This results by splitting $f = f^\alpha + f_\alpha$, where $f^\alpha(x) = f(x)$ when $|f(x)| > \alpha$, and $f_\alpha(x) = f(x)$ when $|f(x)| \le \alpha$. If we then apply (9) to Tf_α, and apply (12) to Tf^α, we get the desired distribution inequality for $Tf = Tf^\alpha + Tf_\alpha$. Now we can complete the argument by noting that

$$\int |(Tf)(x)|^p\, dx = p \int_0^\infty \mu\{x : |(Tf)(x)| > \alpha\}\, \alpha^{p-1}\, d\alpha.$$

6. *Examples of the general theory*

Having presented the basic ideas concerning maximal functions and singular integrals in their general abstract setting, we intend now to illustrate the theory by briefly indicating several examples that will be the subject of later study. Our aim in this section will be only to state in simplest form their salient facts; we postpone the actual proofs of most of the asserted properties to later chapters where these matters are taken up in detail. In keeping with our wish to quickly get to the point, we have chosen most of our examples from the context of the "classical" setting of the theory: here $\mathbf{R}^n$ is given its usual Euclidean structure; i.e., the balls $B(x,\delta)$ are $\{y : |y-x| < \delta\}$, where $|\cdot|$ is the standard Euclidean norm; also $d\mu(x)$ is the usual Lebesgue measure dx.

6.1 Approximations of the identity. Suppose Φ is a fixed function on $\mathbf{R}^n$ that is appropriately small at infinity; for example, take $|\Phi(x)| \le A(1+|x|)^{-n-\varepsilon}$. We also assume here that Φ is normalized by the condition $\int \Phi\, dx = 1$. From such a Φ, we fashion an "approximation to the identity" via convolutions as follows. For each $t > 0$, we set $\Phi_t(x) = t^{-n}\Phi(x/t)$. The relevant point here is that

$$\lim_{t\to 0} (f * \Phi_t)(x) = f(x), \qquad \text{for a.e. } x \tag{16}$$

whenever $f \in L^p(\mathbf{R}^n)$ for some p, $1 \le p \le \infty$.

Indeed, this result (like the corollary in §3.1 above) is a consequence of the following pointwise estimate.

PROPOSITION.

$$\sup_{t>0} |f * \Phi_t(x)| \le c_\Phi Mf(x).$$

The proof of this majorization (and its variants) can be found in Chapter 2, §2.1. For a more precise form of (16), see §8.16 below.

We give two of the original and most important examples. First, if

$$\Phi(x) = c_n(1+|x|^2)^{-(n+1)/2},$$

where

$$c_n = \frac{\Gamma(\frac{n+1}{2})}{\pi^{(n+1)/2}},$$

then $\Phi_t(x)$ is the *Poisson kernel*, and

$$u(x,t) = (f * \Phi_t)(x)$$

gives the solution of the Dirichlet problem for the upper half space

$$\mathbf{R}^{n+1}_+ = \{(x,t) : x \in \mathbf{R}^n, t > 0\},$$

namely

$$\Delta u = \Big(\frac{\partial^2}{\partial t^2} + \sum_{j=1}^n \frac{\partial^2}{\partial x_j^2}\Big)u(x,t) = 0, \qquad u(x,0) \equiv f(x).$$

The second example is the *Gaussian kernel*

$$\Phi(x) = (4\pi)^{-n/2}e^{-|x|^2/4}.$$

This time, if $u(x,t) = (f * \Phi_{t^{1/2}})(x)$, then u is the solution of the heat equation

$$\Big(\frac{\partial}{\partial t} - \sum_{j=1}^n \frac{\partial^2}{\partial x_j^2}\Big)u(x,t) = 0, \qquad u(x,0) \equiv f(x).$$

6.2 Singular Integrals. The main result proved in Theorem 3 for singular integrals is a conditional one, guaranteeing the boundedness on L^p for a range $1 < p \le q$, on the assumption that the boundedness on L^q is already known; the most important instance of this occurs when $q = 2$. In keeping with this, we consider bounded linear transformations T from $L^2(\mathbf{R}^n)$ to itself that commute with translations. As is well known, such operators are characterized by the existence of a bounded function m on $\mathbf{R}^n$ (the "multiplier"), so that T can be realized as

$$\widehat{Tf}(\xi) = m(\xi)\hat{f}(\xi),$$

where $\hat{}$ denotes the Fourier transform. Alternatively, at least on test functions $f \in \mathcal{S}$, T can be realized in terms of convolution with a kernel K,

$$Tf = f * K, \tag{17}$$

where K is the distribution given by $\widehat{K} = m$.†

We shall now examine how Theorem 3 (for $q = 2$) applies to this class of operators. To proceed further, we assume that the distribution K agrees away from the origin with a function that is locally integrable away from the origin; we denote this function by $K(x)$. Then (17) implies that

$$Tf(x) = \int K(x-y)\, f(y)\, dy, \qquad \text{for a.e. } x \notin \operatorname{supp} f$$

whenever f is in L^2 and f has compact support. This is the representation (8) in the present context. Next, the basic condition (10) is then equivalent with

$$\int_{|x|\ge c|y|} |K(x-y) - K(x)|\, dx \le A, \tag{18}$$

for all $y \ne 0$, where $c > 1$.

6.2.1 Let us consider the condition (18) further. First, as is easily seen, it is a consequence of the differential inequalities

$$\Big|\Big(\frac{\partial}{\partial x}\Big)^\alpha K(x)\Big| \le A_\alpha |x|^{-n-|\alpha|} \qquad \text{for all } \alpha, \tag{18_α}$$

or its weaker form, (here $\gamma > 0$ is fixed)

$$|K(x-y) - K(x)| \le A\frac{|y|^\gamma}{|x|^{n+\gamma}}, \qquad \text{whenever } |x| \ge c|y|. \tag{18_γ}$$

How do K, satisfying such conditions, come about? It turns out that, roughly speaking, such conditions on K have equivalent versions when stated in terms of the Fourier transform of K, namely the multiplier m. In Chapter 6, we shall prove the following proposition:

† For these two realizations of translation-invariant operators on $L^2(\mathbf{R}^n)$, see *Fourier Analysis*, Chapter 1, §3; here $\mathcal{S}$ denotes the Schwartz class of testing functions (see, e.g., Chapter 3, §1.1 for a definition).

PROPOSITION. (a) *If we assume that*

$$\left|\left(\frac{\partial}{\partial\xi}\right)^{\alpha} m(\xi)\right| \le A'_{\alpha}|\xi|^{-\alpha}$$

holds for all α, then K satisfies (18_α) *for all α.*

(b) *If we assume that m satisfies the above inequality for all $0 \le |\alpha| \le \ell$, where ℓ is the smallest integer $> n/2$, then K satisfies* (18).

We make two remarks about the above singular integrals.

1. If m is homogeneous of degree 0 and C^∞ away from the origin, then it satisfies (a) of the proposition. It can then be shown that, in addition, the function $K(x)$ is homogeneous of degree $-n$ and (besides being smooth away from the origin) satisfies the cancellation condition

$$\int_{|x|=1} K(x)\, d\sigma(x) = 0.$$

The distribution K can be realized in terms of a principal value singular integral involving the function $K(x)$ (see §8.18 and §8.19 below). The simplest and most basic instances of such operators are the *Riesz transforms* R_j, $j = 1, \dots, n$. These are given by

$$R_j(f) = f * K_j, \quad \text{where} \quad \widehat{K}_j(\xi) = \frac{\xi_j}{i|\xi|}, \quad K_j(x) = \frac{c_n x_j}{|x|^{n+1}};$$

here c_n is the constant appearing in the definition of the Poisson kernel (see §6.1 above).

Of course, the classical example is the one that arises when the dimension $n = 1$. Then $R_1 = H$ is the *Hilbert transform* and is given by convolution with $K(x) = 1/\pi x$; here $\widehat{K}(\xi) = -i\,\mathrm{sign}(\xi)$.

2. The multipliers m satisfying condition (b) of the proposition are essentially those that arise as *Marcinkiewicz multipliers*; see Chapter 6, §4.4 and §7.6.

6.3 Maximal functions, singular integrals, and square functions. There are three interrelated concepts arising in a fundamental way in harmonic analysis that may be thought of as three different manifestations of the same essential idea; they are maximal functions, singular integrals, and square functions. The first two we have already considered, and the third (defined below) will play an important role later. Here we want to illustrate this threefold unity in terms of some of its common roots, again in the standard setting for $\mathbf{R}^n$.

We begin with a fixed function Φ that is sufficiently small at infinity (e.g., $\Phi \in \mathcal{S}$, but weaker assumptions will do) with (say) $\int \Phi = 1$. First, recall that the operator

$$f \mapsto \sup_{t>0} |f * \Phi_t|$$

is closely related to the maximal operator M, as was mentioned in §6.1. Second, if we make the alternate assumption that $\int \Phi\, dx = 0$, then the operator

$$f \mapsto \int_0^\infty (f * \Phi_t)\frac{dt}{t}, \tag{19}$$

when appropriately defined, will be a singular integral of the kind discussed in §6.2. Indeed,

$$\lim_{\substack{N\to\infty \\ \varepsilon\to 0}} \int_\varepsilon^N \Phi_t \frac{dt}{t}$$

converges in the sense of distributions to K, where $\widehat{K}(\xi)$ is bounded, and K satisfies (a) of the proposition in §6.2; in fact, $\widehat{K} = m$ is homogeneous of degree 0.[‡] For sketches of the proofs of these statements concerning (19), see §8.19 below.

Third, coming to square functions, each Φ with $\int \Phi = 0$ leads to the square function s_Φ, which is the operator $f \mapsto s_\Phi(f)$ given by

$$(s_\Phi f)(x) = \left(\int_0^\infty |(f * \Phi_t)(x)|^2 \frac{dt}{t}\right)^{1/2}. \tag{20}$$

Also occurring often in practice is the "nontangential" version of s_Φ. It is given by

$$(S_\Phi f)(x) = \left(\int\!\!\int_\Gamma |f * \Phi_t(x-y)|^2 \frac{dt\, dy}{t^{n+1}}\right)^{1/2}, \tag{20$'$}$$

where Γ is the cone $\Gamma = \{(y,t) : |y| < t\}$.

6.3.1 Square functions, as their name suggests, have a very direct connection with L^2 estimates. It is easily seen that

$$\|S_\Phi(f)\|_{L^2} = c\|s_\Phi(f)\|_{L^2} \le A\|f\|_{L^2}.$$

Indeed, observe that, by Plancherel's theorem,

$$\begin{aligned}\|s_\Phi(f)\|_{L^2}^2 &= \int_{\mathbf{R}^n}\int_0^\infty |\widehat{f}(\xi)|^2|\widehat{\Phi}(t\xi)|^2\frac{dt}{t}\,d\xi \\ &\le \left(\sup_\xi \int_0^\infty |\widehat{\Phi}(t\xi)|^2\frac{dt}{t}\right)\int_{\mathbf{R}^n}|\widehat{f}(\xi)|^2\,d\xi \le A^2\|f\|_{L^2},\end{aligned}$$

[‡] A converse is also true and, for certain Φ, leads to the representation $f = \int_0^\infty (f * \Phi_t)\, dt/t$.

since

$$\sup_\xi \int_0^\infty |\widehat{\Phi}(t\xi)|^2 \frac{dt}{t} \le A^2.$$

The last fact follows easily because

$$|\widehat{\Phi}(u)| \le A|u| \quad \text{and} \quad |\widehat{\Phi}(u)| \le A|u|^{-1}.$$

In addition, Fubini's theorem shows that

$$\|S_\Phi(f)\|_{L^2}^2 = v_n \|s_\Phi(f)\|_{L^2}^2,$$

where v_n is the volume of the unit ball in $\mathbf{R}^n$; for this point see also Chapter 3, §4.4.3.

One of the keys to a better understanding of square functions is to view them as ordinary singular integrals, but now as taking their values in a Hilbert space. For this reason we now turn to the extension of the theory of this chapter for functions whose values lie in a Banach space.

6.4 Banach space valued functions. The results for singular integrals in Theorem 3 go through for functions that take their values in Banach spaces. Thus if B_1, B_2 are a pair of Banach spaces, and $\mathcal{B}(B_1, B_2)$ is the Banach space of bounded operators from B_1 to B_2, we may assume that in the definition of T given by (8), f takes its values in B_1, K takes its values in $\mathcal{B}(B_1, B_2)$, and Tf takes its values in B_2. Throughout, the absolute value $|\cdot|$ must be replaced by the norm in B_1, $\mathcal{B}(B_1, B_2)$, or B_2, respectively.

Instead of trying to formulate here a precise theorem to this effect, we mention two particular results. The first deals with the square function s_Φ discussed in §6.3 above. Here we take $B_1 = \mathbf{R}^1$, and B_2 to be the Hilbert space $L^2[(0,\infty); dt/t]$. With Φ fixed (say $\Phi \in \mathcal{S}$ and $\int \Phi\, dx = 0$), we define $K(x)$ to be the B_2-valued function given by

$$K(x) = t^{-n}\Phi(x/t) = \Phi_t(x).$$

Now

$$|K(x)| = \left(\int_0^\infty |\Phi_t(x)|^2 \frac{dt}{t}\right)^{1/2},$$

and hence $|K(x)| \le A|x|^{-n}$, since

$$|\Phi_t(x)| \le At^{-n} \quad \text{for } |x| \le t, \quad \text{and} \quad |\Phi_t(x)| \le At|x|^{-n-1} \quad \text{for } |x| \ge t.$$

In the same way,

$$\left|\left(\frac{\partial}{\partial x}\right)^\alpha K(x)\right| \le A_\alpha |x|^{-n-\alpha},$$

so that K satisfies the vector-valued version of condition (18_α). Since we have already verified the L^2 boundedness of $Tf = f * K$, we see that T is bounded on L^p for $1 < p < 2$ (by duality, T is also bounded on L^p for $2 < p < \infty$; see §7.4). Similar considerations hold if we consider the nontangential version of s_Φ, namely S_Φ. In this case we take $B_2 = L^2(\Gamma; dy\, dt/t^{n+1})$, with $K(x) = \Phi_t(x - y)$, $(y,t) \in \Gamma$.

A second example of the Banach space version of singular integrals is taken up in Chapter 2, §4.2, where it is seen that the maximal theorem can be viewed as a result on singular integrals that take their values in an appropriate Banach space. These two examples are further evidence why the threesome—singular integrals, maximal operators, and square functions—are manifestations of the same unity. In terms of singular integrals, the usual theory arises when the kernel is scalar valued, maximal operators arise when the kernel is L^∞-valued, and square functions arise when the kernel takes its values in a Hilbert space.

6.5 We return to the general structure described in §1 and §2, and ask what are the conditions analogous to (18_α) or (18_γ), which we had in the translation-invariant case, that would play a similar role in guaranteeing the crucial condition (10). To deal with this requires the following construct. We set

$$V(x,y) = \inf\{\mu(B(y,\delta)) : x \in B(y,\delta)\}.$$

That is, $V(x,y)$ is the volume of the "smallest" ball $B(y,\delta)$ that contains x. In view of the remark in §2.4, the volume function V is essentially symmetric; i.e., $V(x,y)$ and $V(y,x)$ are comparable. The extension of (18_α) (at least when $\alpha = 0$) is then the condition

$$|K(x,y)| \le \frac{A}{V(x,y)}.$$

To formulate its differential analogue, we recall the distance function $\rho(x,y)$ (see §2.4), for which $B(y,\delta)$ is given by $\{x : \rho(x,y) < \delta\}$. Let $\eta(s)$, $0 \le s \le 1$, be a Dini modulus of continuity, i.e., a nondecreasing function with $\eta(0) = 0$ and $\int_0^1 \eta(s)\frac{ds}{s} < \infty$. The variant of (18_γ) we shall need is that, for some Dini modulus η and some constant c,

$$|K(x,y) - K(x,\bar{y})| \le \eta\Big(\frac{\rho(y,\bar{y})}{\rho(x,\bar{y})}\Big)(V(x,\bar{y}))^{-1} \tag{18'}$$

whenever $\rho(x,\bar{y}) \ge c\rho(y,\bar{y})$.

PROPOSITION. *If K satisfies (18′), then it also satisfies (10).*

To prove this, write the integral in (10) as

$$\int_{^cB(\bar{y},c\delta)} |K(x,y) - K(x,\bar{y})|\,d\mu(x)$$
$$= \sum_{k=0}^{\infty} \int_{B(\bar{y},2^{k+1}c\delta)\setminus B(\bar{y},2^k c\delta)} |K(x,y) - K(x,\bar{y})|\,d\mu(x).$$

If $x \notin B(\bar{y}, 2^k c\delta)$, then $\rho(x,\bar{y}) \geq 2^k c\delta$ and $V(x,\bar{y}) \geq \mu(B(\bar{y},2^k c\delta))$, while $\rho(y,\bar{y}) < \delta$, if $y \in B(\bar{y},\delta)$. Taking into account the monotonicity of η and (18′) we see that the above sum is bounded by

$$\sum_{k=0}^{\infty} \mu(B(\bar{y},2^{k+1}c\delta)) \cdot \eta(2^{-k}c^{-1}) \cdot \mu(B(\bar{y},2^k c\delta))^{-1}$$

which, by the doubling condition in §1.1, is bounded by

$$c\sum_{k=0}^{\infty} \eta(2^{-k}c^{-1}) \leq c' \int_0^1 \eta(s)\frac{ds}{s} = A < \infty.$$

The proposition is therefore proved.

The reader might note how the two independent elements of our metric structure—distance and volume—combine neatly in (18′) and its consequences; as in other arguments, each plays its role, but it is in their interplay that the interest lies.

6.6 We must emphasize that the theory developed so far for singular integrals tells only part of the story: the conditional nature of Theorem 3 makes everything depend on an additional L^q result. Sometimes (e.g., when $q = \infty$, as in the case of maximal functions), this result is immediate; however, the more general situation (and in particular the case $q = 2$) requires further analysis. Here the translation-invariant case has served as our guide and brought us to the province of the Fourier transform; from there we will be led to oscillatory integrals, pseudo-differential operators, and almost orthogonality. These are matters that we will take up in chapters 6 and 7 below.

7. *Appendix: Truncation of singular integrals*

We shall now discuss the procedure of regularizing singular integrals by "truncation" — the method applicable to the classical Hilbert transform and its generalizations.[†]

[†] See also *Singular Integrals*, Chapter 2, §4.

7.1 We assume that we are given an operator T bounded from $L^q(d\mu)$ to itself for some q, $1 < q < \infty$; i.e.,

$$\|T(f)\|_q \leq A\|f\|_q. \tag{21}$$

In addition we suppose that T has associated with it a kernel $K(x,y)$ that satisfies the estimate

$$|K(x,y)| \leq \frac{A}{V(x,y)} \tag{22}$$

so that

$$Tf(x) = \int K(x,y)f(y)\,d\mu(y) \tag{23}$$

for a.e. x outside the support of f.

We define the truncated kernels K_ε by

$$K_\varepsilon(x,y) = K(x,y), \qquad \text{if} \quad \rho(x,y) \geq \varepsilon,$$

and $K_\varepsilon(x,y) = 0$ otherwise; T_ε is then given by

$$(T_\varepsilon f)(x) = \int K_\varepsilon(x,y)f(y)\,d\mu(y).$$

PROPOSITION 1. *Under the assumptions* (21), (22), *and* (23), *the* T_ε *satisfy the estimate*

$$\|T_\varepsilon(f)\|_q \leq A'\|f\|_q,$$

with A' *independent of* ε.

Proof. Let $\varepsilon > 0$ be fixed throughout.

To begin with, let us observe that if $\lambda(\alpha)$ is the distribution function of $V(y,x)^{-1}$ as a function of y (with x fixed), then

$$\lambda(\alpha) = \mu\{y : y \in B(x,\delta) \text{ for some } \delta \text{ with } \mu(B(x,\delta)) < \alpha^{-1}\} \leq \frac{1}{\alpha}. \tag{24}$$

It follows therefore by (22) that $K_\varepsilon(x,y)$, as a function of y, is in L^r for $1 < r \leq \infty$; and therefore by Hölder's inequality, T_ε is well defined on L^q.

Set $\tilde{T}_\varepsilon = T - T_\varepsilon$; $\tilde{T}_\varepsilon$ is the "near" part of T. Fix $\bar{x} \in \mathbf{R}^n$ and $f \in L^q$; all of our estimates will be independent of ε, $\bar{x}$, and f. Let χ_δ denote the characteristic function of $B(\bar{x},\delta)$. We first show that

$$\|\chi_{a\varepsilon}\tilde{T}_\varepsilon f\|_q \leq C\|\chi_{b\varepsilon}f\|_q \tag{25}$$

for some small a and a somewhat larger b; this is the crux of the proof.

Notice that $\tilde{T}_\varepsilon f(x) = 0$ if $\operatorname{supp}(f) \subset {}^cB(x,\varepsilon)$ and that $\tilde{T}_\varepsilon f(x) = Tf(x)$ if $\operatorname{supp}(f) \subset B(x,\varepsilon)$. Taking c as in the triangle inequality (1), we have

$$\chi_{a\varepsilon}\tilde{T}_\varepsilon f = \chi_{a\varepsilon}\tilde{T}_\varepsilon \chi_{b\varepsilon} f, \qquad \text{if } b \geq c(1+a).$$

Next, we split $\chi_{a\varepsilon}\tilde{T}_\varepsilon\chi_{b\varepsilon}f$ into two parts:

$$\chi_{a\varepsilon}\tilde{T}_\varepsilon\chi_{b\varepsilon}f = \chi_{a\varepsilon}\tilde{T}_\varepsilon\chi_{d\varepsilon}f + \chi_{a\varepsilon}\tilde{T}_\varepsilon(\chi_{b\varepsilon} - \chi_{d\varepsilon})f.$$

We assume $c(a+d) < 1$, so that

$$B(\bar{x}, d\varepsilon) \subset B(x, \varepsilon) \quad \text{whenever } x \in B(\bar{x}, a\varepsilon).$$

This gives $\chi_{a\varepsilon}\tilde{T}_\varepsilon\chi_{d\varepsilon} = \chi_{a\varepsilon}T\chi_{d\varepsilon}$ and

$$\|\chi_{a\varepsilon}\tilde{T}_\varepsilon\chi_{d\varepsilon}f\|_q = \|\chi_{a\varepsilon}T\chi_{d\varepsilon}f\|_q \le A\|\chi_{d\varepsilon}f\|_q \le A\|\chi_{b\varepsilon}f\|_q,$$

provided $d < b$.

Now we handle $\chi_{a\varepsilon}\tilde{T}_\varepsilon(\chi_{b\varepsilon} - \chi_{d\varepsilon})f(x)$, which equals

$$\int_{B(x,\varepsilon)\cap B(\bar{x},b\varepsilon)\setminus B(\bar{x},d\varepsilon)} K(x,y)f(y)\,d\mu(y) \quad \text{for a.e. } x \in B(\bar{x}, a\varepsilon), \tag{26}$$

if we assume that $a < d$ and apply (23). For y in the above range of integration we have that

$$d\varepsilon \le \rho(\bar{x}, y) \le c[\rho(x,\bar{x}) + \rho(x,y)] \le c[a\varepsilon + \rho(x,y)].$$

Choosing a so that $ca < d/2$, we have

$$\rho(x,y) \ge c'\varepsilon \ge c''b\varepsilon \ge c''\rho(\bar{x},y);$$

the doubling condition then shows that $V(x,y) \ge CV(\bar{x},y)$. Thus

$$|K(x,y)| \le \frac{C}{V(\bar{x},y)} \le \frac{C}{\mu(B(\bar{x},d\varepsilon))} \le \frac{C'}{\mu(B(\bar{x},b\varepsilon))}.$$

By Hölder's inequality, $\|\chi_{b\varepsilon}f\|_1 \le \|\chi_{b\varepsilon}f\|_q \cdot \mu(B(\bar{x},b\varepsilon))^{1-1/q}$. Combining this with (26) gives

$$\|\chi_{a\varepsilon}\tilde{T}_\varepsilon(\chi_{b\varepsilon} - \chi_{d\varepsilon})f\|_q \le \frac{C\mu(B(\bar{x},a\varepsilon))^{1/q}\|\chi_{b\varepsilon}f\|_1}{\mu(B(\bar{x},b\varepsilon))} \le C\|\chi_{b\varepsilon}f\|_q$$

and, with it, (25).

It is now only a question of adding these inequalities for a suitable collection of balls covering $\mathbf{R}^n$. To do this, let c_1 be an engulfing constant (as in §1.1) and choose a maximal disjoint collection of balls $\{B(\bar{x}^k, a\varepsilon/c_1)\}$. So $\bigcup B(\bar{x}^k, a\varepsilon) = \mathbf{R}^n$ by construction. Also, $\{B(\bar{x}^k, b\varepsilon)\}$ have the bounded overlapping property: for some N, no point in $\mathbf{R}^n$ belongs to more than N of the $B(\bar{x}^k, b\varepsilon)$; N is of course independent of ε. To see this, assume $z \in \bigcap_{k=1}^{N} B(\bar{x}^k, b\varepsilon)$. Then $B(\bar{x}^k, a\varepsilon/c_1) \subset B(z, k\varepsilon)$, for k sufficiently large. By the doubling property, $\mu(B(z,k\varepsilon)) \le C\mu(B(\bar{x}^k, a\varepsilon/c_1))$, and by the disjointness of the $B(\bar{x}^k, a\varepsilon/c_1)$, we get $N \le C$, verifying the bounded overlapping property. Finally then, by (25),

$$\begin{aligned}\int_{\mathbf{R}^n} |\tilde{T}_\varepsilon f|^q\,d\mu &\le \sum_k \int_{B(\bar{x}^k,a\varepsilon)} |\tilde{T}_\varepsilon f|^q\,d\mu \\ &\le A'\sum_k \int_{B(\bar{x}^k,b\varepsilon)} |f|^q\,d\mu \le A'N\int_{\mathbf{R}^n} |f|^q\,d\mu,\end{aligned}$$

and the proposition is proved.

COROLLARY 1. *Suppose T has a kernel K that satisfies*

$$\int_{{}^cB(\bar{y},c\delta)} |K(x,y) - K(x,\bar{y})|\,d\mu(x) \le A \quad \textit{whenever } y \in B(\bar{y},\delta), \tag{10}$$

in addition to (21), (22), and (23). Then $\|T_\varepsilon f\|_p \le A_p\|f\|_p$ for $1 < p \le q$ and

$$\mu\{x : |T_\varepsilon f(x)| > \alpha\} \le \frac{A}{\alpha}\|f\|_1, \quad \textit{for all } \alpha > 0,$$

with A_p and A independent of ε.

Proof. We use Theorem 3 of §5. We have just shown that $\|T_\varepsilon f\|_q \le C\|f\|_q$ uniformly in ε. It suffices to show

$$\int_{{}^cB(\bar{y},c\delta)} |K_\varepsilon(x,y) - K_\varepsilon(x,\bar{y})|\,d\mu(x) \le A \quad \text{whenever } y \in B(\bar{y},\delta) \tag{10_ε}$$

uniformly in δ, $\bar{y}$, y, and ε, for some possibly larger choice of constant c. Since we have such an estimate for K, we need only bound

$$\int_{\substack{\varepsilon\le\rho(x,y)\\ c\delta\le\rho(x,\bar{y})<\varepsilon}} |K(x,y)|\,d\mu(x) \tag{$*$}$$

and

$$\int_{\substack{\rho(x,y)<\varepsilon\\ \rho(x,\bar{y})\ge\max\{c\delta,\varepsilon\}}} |K(x,\bar{y})|\,d\mu(x). \tag{$**$}$$

For $(*)$ to be nonzero, we must have $c\delta \le \varepsilon$. We have $(*) \le \mu(B(\bar{y},\varepsilon)) \cdot A/\mu(B(y,\varepsilon))$. Using the doubling property and $\rho(y,\bar{y}) \le \delta < \varepsilon/c$ shows that $\mu(B(\bar{y},\varepsilon))$ and $\mu(B(y,\varepsilon))$ are comparable, giving a uniform bound for $(*)$.

Similarly, for $(**)$ to be nonzero, we must have $c\delta \le c'(\varepsilon+\delta)$, where c' is the "quasimetric constant" of ρ. Taking $c > c'$ gives $c''\delta \le \varepsilon$ and shows that $\mu(B(y,\varepsilon))$ and $\mu(B(\bar{y},\varepsilon))$ are comparable. Therefore $(**) \le \mu(B(y,\varepsilon)) \cdot A/\mu(B(\bar{y},\varepsilon))$ is uniformly bounded and the proof is complete.

Note. It should be observed that Proposition 1, its corollary, and the further results derived from them also hold if we replace the assumption (22) by the weaker assumption

$$\int_{\varepsilon<\rho(x,y)<2\varepsilon} (|K(x,y)| + |K(y,x)|)\,d\mu(x) \le A, \quad \text{for all } 0 < \varepsilon < \infty. \tag{22$'$}$$

7.2 We keep to our assumptions above. Then, in view of the uniform boundedness of the T_ε, there exists a sequence $\varepsilon_j \to 0$ so that the operators T_{ε_j} tend weakly (in L^q) to a limit T_0. We claim that there exists a bounded measurable function $a(x)$,‡ so that

$$(Tf)(x) = (T_0f)(x) + a(x)\cdot f(x), \quad \text{for every } f \in L^q. \tag{27}$$

‡ Taking T to be the identity operator shows that $a(x)$ need not be zero. Moreover, the full family of truncations $\{T_\varepsilon\}$ does not necessarily tend to a limit, as can be seen by considering fractional integration of imaginary order and other examples. See §8.22 below.

We set $\Delta = T - T_0$. The key observation to be made is that Δ is "local", in the sense that $\Delta(g)(x) = 0$ if g vanishes in a neighborhood of x. In fact, with g given, $(T - T_\varepsilon)(g)(x) = 0$ as soon as $\varepsilon \le \rho(x, \operatorname{supp} g)$, which makes clear the local character of Δ. This implies that, for every cube Q,

$$\Delta(\chi_Q g) = \chi_Q(\Delta g) \quad \text{a.e.,} \tag{28}$$

because $\Delta(\chi_Q g)$ vanishes outside $\bar{Q}$, and $\Delta(\chi_Q g) = \Delta(g) - \Delta(\chi_{^cQ} g)$, while $\Delta(\chi_{^cQ} g)$ vanishes in Q^o. (Observe also that $\mu(bQ) = 0$ for "most" Q.)

Using linear combinations of characteristic functions of cubes, the identity (28), the regularity of μ, and the L^q-boundedness of $\Delta = T - T_0$, we get by a passage to the limit that

$$\Delta(fg) = f\Delta(g)$$

whenever $f \in L^q$ and g is bounded with compact support. Let $\{O_m\}$ be an increasing sequence of bounded open sets that exhaust $\mathbf{R}^n$ and let χ_m be the characteristic function of O_m. Thus if f_m is supported in O_m then $\Delta(f_m) = f_m \Delta(\chi_{m'})$, whenever $m' \ge m$. Therefore, $\Delta(\chi_m)$ forms a coherent set of functions; that is,

$$\Delta(\chi_{m_1}) = \Delta(\chi_{m_2}) \text{ in } O_{m_1}, \text{ if } m_1 \le m_2.$$

Hence there exists a function $a(x)$, so that $a = \Delta(\chi_m)$ in O_m and, as a result, $\Delta(f_m) = a \cdot f_m$, whenever f_m is supported in O_m. Finally, $\Delta(f) = a \cdot f$ for all $f \in L^q$, and $a(x)$ is bounded since Δ is, proving (27).

7.3 So far we have seen how L^p and weak-type L^1 results for the operator T are related to corresponding estimates for the truncated operators T_ε. For some purposes—e.g., when one tries to make more precise the sense in which $Tf = \lim_{\varepsilon \to 0} T_\varepsilon f + a \cdot f$ for certain T—it is essential to consider the associated maximal operator. We therefore define $T_* f = \sup_{\varepsilon > 0} |T_\varepsilon f|$, and our intention is to prove for T_* the same kind of estimates that hold for T. Experience shows that it is possible to do this only when the kernel K satisfies an additional hypothesis: estimates of the type (10) or (18), but with the roles of x and y reversed. More precisely, we assume there exists a Dini modulus η with

$$|K(x,y) - K(\bar{x},y)| \le \eta\Big(\frac{\rho(x,\bar{x})}{\rho(\bar{x},y)}\Big)[V(\bar{x},y)]^{-1} \tag{29}$$

whenever $\rho(\bar{x},y) \ge c\rho(x,\bar{x})$.

Proposition 2. *Suppose T and its associated kernel K satisfy* (21), (22), (23), (10), *and* (29). *Then*

$$T_* f \le A\{M(Tf) + M(f)\}, \tag{30}$$

and, more generally, for any $r > 0$,

$$T_* f \le A_r\{M(|Tf|^r)^{1/r} + M(f)\}. \tag{30'}$$

Here M is the maximal operator of §3.

Proof. As a preliminary matter we remark that for each $\varepsilon > 0$, $T_\varepsilon f(x)$ is actually continuous in x, and hence $T_* f(x)$ is semicontinuous and measurable.†

Let us now fix an $\bar{x} \in \mathbf{R}^n$ and an $\varepsilon > 0$. Write $f = f_1 + f_2$, where $f_1 = f$ on $B(\bar{x},\varepsilon)$, and $f_2 = f$ on ${}^cB(\bar{x},\varepsilon)$. Thus $T_\varepsilon f(\bar{x}) = Tf_2(\bar{x})$, by the definition of K_ε. The first point to keep in mind is that

$$|Tf_2(\bar{x}) - Tf_2(x)| \le A' Mf(\bar{x}) \quad \text{whenever } \rho(x,\bar{x}) < \varepsilon/c. \tag{31}$$

In fact, the difference in (31) is bounded by

$$\int_{\rho(\bar{x},y) \ge \varepsilon} |K(x,y) - K(\bar{x},y)| \cdot |f(y)|\, d\mu(y)$$
$$= \sum_{k=0}^{\infty} \int_{2^{k+1}\varepsilon > \rho(\bar{x},y) \ge 2^k\varepsilon} |K(x,y) - K(\bar{x},y)| \cdot |f(y)|\, d\mu(y)$$

If we invoke (29), then the right side is majorized by

$$\sum_{k=0}^{\infty} \eta\Big(\frac{1}{2^k c}\Big)\mu(B(\bar{x},\varepsilon 2^k))^{-1} \int_{B(\bar{x},\varepsilon 2^{k+1})} |f(y)|\, d\mu(y) \le c' \sum_{k=0}^{\infty} \eta\Big(\frac{1}{2^k c}\Big) Mf(\bar{x}),$$

verifying (31). Therefore

$$|T_\varepsilon f(\bar{x})| \le |Tf(x)| + |Tf_1(x)| + A'(Mf)(\bar{x}) \quad \text{whenever } x \in B(\bar{x},\varepsilon/c). \tag{32}$$

Inequality (32) provides us with a substantial set of x to exploit; it is just a matter of choosing $x \in B(\bar{x},\varepsilon/c)$ so that neither $Tf(x)$ nor $Tf_1(x)$ is too large. Now

$$\mu\{x \in B(\bar{x},\varepsilon/c) : |Tf(x)| > \alpha\} \le \alpha^{-r} \int_{B(\bar{x},\varepsilon/c)} |Tf(x)|^r\, d\mu(x)$$
$$\le \alpha^{-r}\mu(B(\bar{x},\varepsilon/c))M(|Tf|^r)(\bar{x})$$

for any $r > 0$. Thus if $\alpha \ge 4^{1/r}[M(|Tf|^r)(\bar{x})]^{1/r}$, then

$$\mu\{x \in B(\bar{x},\varepsilon/c) : |Tf(x)| > \alpha\} \le \frac{1}{4}\mu(B(\bar{x},\varepsilon/c)).$$

Also, by Theorem 3,

$$\mu\{x \in B(\bar{x},\varepsilon/c) : |Tf_1(x)| > \alpha\} \le \frac{A}{\alpha}\int |f_1|\, d\mu$$
$$= \frac{A}{\alpha}\int_{B(\bar{x},\varepsilon)} |f|\, d\mu \le \frac{A}{\alpha}\mu(B(\bar{x},\varepsilon))Mf(\bar{x});$$

so if $\alpha \ge 4A'Mf(\bar{x})$, then

$$\mu\{x \in B(\bar{x},\varepsilon/c) : |Tf_1(x)| > \alpha\} \le \frac{1}{4}\mu(B(\bar{x},\varepsilon/c)).$$

Therefore if $\alpha \ge \max\{4^{1/r}[M(|Tf|^r)(\bar{x})]^{1/r}, 4AMf(\bar{x})\}$, then there exists an $x \in B(\bar{x},\varepsilon/c)$ so that $|Tf(x)| \le \alpha$ and $|Tf_1(x)| \le \alpha$. Substituting this in (32) yields

$$T_\varepsilon f(\bar{x}) \le A\{[M(|Tf|^r)(\bar{x})]^{1/r} + Mf(\bar{x})\},$$

which gives (30′) and, in particular, (30).

† This follows from (29) and the L^p control of V^{-1} given by §8.12 below.

COROLLARY 2. *Under the assumptions of Proposition 2,*

$$\|T_* f\|_p \le A_p \|f\|_p, \quad 1 < p \le q$$

and

$$\mu\{x : T_* f(x) > \alpha\} \le \frac{A}{\alpha}\|f\|_1, \quad \text{for all } \alpha > 0.$$

The first conclusion follows directly from (30), the corollary in §5.2, and the L^p boundedness of the maximal function (Theorem 1 in §3). The proof of the second conclusion is in the same spirit but is a little more complicated. We need two observations. First, F satisfies a weak-type L^1 inequality,

$$\mu\{x \in \mathbf{R}^n : |F(x)| > \alpha\} \le \frac{A}{\alpha} \quad \text{for all } \alpha > 0, \tag{33}$$

exactly when $|F|^r$ belongs to the Lorentz space $L^{1/r,\infty}$, if $0 < r < 1$. Moreover, if we choose the smallest A occurring in (33), then A^r is equivalent to the $L^{1/r,\infty}$ norm of $|F|^r$. We then apply this observation successively to $F = T_*(f)$ and $F = T(f)$, once we note that, by the general form of the Marcinkiewicz interpolation theorem, the mapping $f \mapsto M(f)$ is bounded from $L^{1/r,\infty}$ to itself, if $0 < r < 1$.[‡]

7.4 Three concluding remarks.

(i) One has $Tf(x) = \lim_{\varepsilon \to 0} T_\varepsilon f(x) + a(x)f(x)$ for almost every x, whenever $f \in L^p$, $1 \le p \le q$, if one can prove the convergence for f lying in a dense subspace of L^p. This follows the usual pattern of proving the existence of limits almost everywhere as a consequence of the corresponding maximal inequality.[†]

(ii) An immediate consequence of (27) and the definition of T_* is the inequality

$$|Tf(x)| \le |T_* f(x)| + c|f(x)|. \tag{34}$$

(iii) Under all the assumptions we have made (namely, (21), (22), (23), (10), and (29)), we can also conclude that T and T_* are bounded on L^p for every p, $1 < p < \infty$, and not just for $1 < p \le q$.

To see this, let T^* be the dual of T, which is the bounded operator from $L^{q'}$ to itself ($1/q' + 1/q = 1$) determined by the identity

$$\int (T^* f) g \, d\mu = \int (Tg) f \, d\mu, \tag{35}$$

holding whenever $f \in L^{q'}$ and $g \in L^q$. T^* is represented (in the sense of (23)) by a kernel $K^*(x,y)$, where $K^*(x,y) = K(y,x)$. Now the assumption (29) and the proposition in §6.1 show that the corollary in §5.2 is applicable to T^*, proving its boundedness for $1 < p \le q'$, and so (35) gives the desired conclusion for T in the range $q \le p < \infty$. The result for T_* then follows by appealing to Corollary 2 in §7.3.

[‡] For the properties of the Lorentz spaces and the general form of the Marcinkiewicz theorem, consult *Fourier Analysis*, Chapter 5.

[†] See, for instance, *Singular Integrals*, p. 45.

8. Further results

A. Real-variable structures

8.1 In §1, we can replace the underlying space $\mathbf{R}^n$ by a locally compact space; it is of course assumed that a family of balls (or, alternatively, a quasi-distance ρ) and a measure $d\mu$ are given that satisfy the assumptions required there. Three illustrations of this are as follows.

(i) The first occurs when the underlying space is discrete. A particular instance is a finitely generated discrete group G of *polynomial growth*. By this, we mean that there is a finite set $U \subset G$ so that, if $U^0 = \{\mathrm{id}\}$, $U^{k+1} = U \cdot U^k$, then $G = \bigcup U^k$, while the cardinality of U^k is bounded by a constant multiple of a fixed power of k. In this case, we can take

$$B(x,\delta) = \{y \in G : y^{-1} \cdot x \in U^k, \text{ for some } k < \delta\}$$

and $d\mu$ to be the counting measure; all the requirements in §1 are then satisfied. For the facts about groups of polynomial growth, see Gromov [1981], Tits [1981].

(ii) Here we take our underlying space to be a smooth compact Riemannian manifold M, with the quasi-distance ρ given by the Riemannian metric; $d\mu$ is the induced measure. Again, all the requirements in §1 are satisfied.

(iii) If M is a noncompact Riemannian manifold, our basic assumptions may not hold because of the behavior of the metric at infinity. A simple example is furnished by the hyperbolic unit disc $\{z \in \mathbf{C} : |z| < 1\}$, with the holomorphically invariant metric $ds^2 = |dz|^2/(1-|z|^2)^2$. The volume of a ball of radius δ grows exponentially as $\delta \to \infty$; hence the doubling condition in §1.1 fails for large δ.

8.2 Suppose M is a smooth compact manifold of dimension n. We shall describe a construction of a family of "nonisotropic" balls on M. To do this, assume we have a smooth mapping $\Theta : M \times M \to \mathbf{R}^n$ so that $\Theta(x,x) = 0$ for all $x \in M$ and, for each fixed $x \in M$, the mapping $y \mapsto \Theta(x,y)$ is a diffeomorphism of a neighborhood of $x \in M$ to a neighborhood of $0 \in \mathbf{R}^n$. Suppose also that we are given an n-tuple $a_1, \dots, a_n$ of strictly positive numbers. We define a norm function ρ on $\mathbf{R}^n$ in terms of these numbers by

$$\rho(x) = \rho(x_1, \dots, x_n) = \max_k |x_k|^{1/a_k}.$$

With Θ, ρ as above, we define

$$B(x,\delta) = \{y : \rho(\Theta(x,y)) < \delta\}$$

and take $d\mu$ to be any measure on M with a smooth strictly positive density. One can then assert the following.

(a) If the exponents a_k are all equal (more strictly, if $a_k \equiv 1$) then the balls defined above are equivalent to those described in §8.1, part (ii).

(b) If $\max a_k \le 2\min a_k$, all the requirements in §1 are satisfied. For closely related results that are relevant to complex analysis, see §8.3 below. Other instances arise in Folland and Stein [1974], Nagel and Stein [1979].

(c) When $\max a_k > 2\min a_k$, it is not necessarily true that the resulting balls satisfy the crucial engulfing property (1) in §1.1. An example is given in Nagel and Stein [1979].

8.3 Suppose M arises as the boundary of a smooth bounded domain Ω in $\mathbf{C}^N$; i.e., $M = b\Omega$ and $n = 2N-1$. For each boundary point x, let ν_x denote the unit (outward) normal to M at x. The directions orthogonal to $\mathbf{C}\cdot\nu_x$ are the "complex tangential" directions at x. Define the "polydisc" $P(x,\delta) \subset \mathbf{C}^N$ to be the product of: (1) A one-dimensional complex disc in the direction of ν_x, with radius δ^2, and (2) An $(N-1)$-dimensional complex ball lying in the orthogonal complement of $\mathbf{C}\cdot\nu_x$, having radius δ. Set

$$B(x,\delta) = M \cap P(x,\delta), \qquad \text{for } x \in M.$$

Then this family of balls, together with a fixed measure $d\mu$ given by a strictly positive smooth density, satisfies all our assumptions in §1.

This construction is equivalent to a special case of §8.2(b), in which $a_k = 1$ for $1 \le k < n$, and $a_n = 2$. See Stein [1972]; for the case of the unit ball in $\mathbf{C}^N$, see Korányi [1969].

8.4 Let $X_1,\ldots,X_k$ be a collection of real smooth vector fields on a compact manifold M of dimension n; we suppose that these vector fields and their commutators of order at most m span the tangent space at each point of M.

For each $x,y \in M$, define $\rho(x,y)$ to be the least time taken to move from x to y along a path pointing in the directions of the X_j's. More precisely, $\rho(x,y)$ is the infimum of the T for which there exists a piecewise smooth path $\gamma : [0,T] \to M$, with $\gamma(0) = x$, $\gamma(T) = y$, and

$$\dot\gamma(t) = \sum_{j=1}^{k} a_j(t)X_j(\gamma(t)), \qquad \text{with} \qquad \sum_{j=1}^{k}[a_j(t)]^2 \le 1.$$

Set $B(x,\delta) = \{y : \rho(x,y) < \delta\}$. Then these balls, together with a measure $d\mu$ given by a strictly positive smooth density, satisfy all the assumptions in §1.

In fact, let $\{Y_j\}_{j=1}^N$ be an enumeration of the vector fields $X_1,\ldots,X_k$, and all their commutators of order $\le m$, and let degree(Y_j) be the order of the corresponding commutator. Let I denote any n-tuple of integers $I = (j_1,\ldots,j_n)$, with $1 \le j_k \le N$, and let degree$(I) = \sum_{\ell=1}^{n}$ degree(Y_{j_ℓ}). Write also

$$\Lambda(x,\delta) = \sum_I |\det(Y_{j_1},\ldots,Y_{j_n})| \cdot \delta^{\text{degree}(I)}.$$

Then for each $x \in M$, the function $\delta \mapsto \Lambda(x,\delta)$ is a polynomial in δ of degree at most nm. The significant fact is that

$$\mu(B(x,\delta)) \approx \Lambda(x,\delta) \qquad \text{for small } \delta,$$

and hence $d\mu$ satisfies the crucial doubling property. For further details, see Nagel, Stein, and Wainger [1981], [1985].

Several further comments are in order.

(i) The balls described here are equivalent with those arising in §8.3 in the special case in which $X_1,\ldots,X_k$ are the real and imaginary parts of a basis of the tangential Cauchy-Riemann operators; in addition we require that Ω be strongly pseudoconvex or, more generally, that its Levi form has, at each point, at least one nonzero eigenvalue.

(ii) These balls are also of interest when the domain Ω is in $\mathbf{C}^2$ and is of "finite type"; see Christ [1988], C. Fefferman and Kohn [1988], McNeal [1989], Nagel, Rosay, Stein, and Wainger [1989].

(iii) These balls play a crucial role in the study of hypoelliptic operators such as $\sum_{j=1}^{k} X_j^2$. See Rothschild and Stein [1976], also §8.5 below.

8.5 Here are two extensions of the structure described in §8.4.

(a) For each $x \in M$, let T_x^* denote the cotangent space of M at x. Suppose we are given, for each x, a nonnegative quadratic form Q_x defined on T_x^* that depends smoothly on x. Define $\rho(x,y)$ to be the least τ for which there exists a piecewise smooth curve $\gamma : [0,\tau] \to M$ with $\gamma(0) = x$, $\gamma(\tau) = y$, and for which $|\langle\dot\gamma(t),\xi\rangle| \le 1$ for all $t \in [0,\tau]$ and $\xi \in T^*_{\gamma(t)}$ with $Q_{\gamma(t)}(\xi) \le 1$. We make the key assumption that for some $\varepsilon > 0$, we have $\rho(x,y) \le c\,d(x,y)^\varepsilon$, where d is a Riemannian distance. With $d\mu$ as above, all the properties in §1 are then satisfied. See C. Fefferman and Phong [1983], Sanchez-Calle [1984], C. Fefferman and Sanchez-Calle [1986]. This extends the results alluded to in §8.4(iii) to more general second-order operators; here Q_x is not necessarily the sum of squares of linear forms.

(b) For the treatment of other hypoelliptic operators that are polynomials in vector fields, the following extension is needed. We now assume that we are given a double-indexed family of vector fields $\{X_j^i\}$, with $1 \le i \le r$, where X_j^i will be thought of as having degree i. We suppose that these vector fields and their commutators span the tangent space at each point. Let us define $B(x,\delta)$ to consist of all y that can be joined to x by a piecewise smooth path $\gamma : [0,1] \to M$, $\gamma(0) = x$, $\gamma(1) = y$, with

$$\dot\gamma(t) = \sum_{i,j} a_j^i(t)X_j^i(t), \quad \text{and} \quad \sum_{i,j}\delta^{-2i}[a_j^i(t)]^2 \le 1.$$

Then these balls, together with $d\mu$ as above, satisfy all the properties in §1. There is also a formula for the volume $\mu(B(x,\delta))$ analogous to that in §8.4; see Nagel, Stein, and Wainger [1985].

8.6 In the next four sections we shall consider $\mathbf{R}^n$ with its usual balls $B(x,\delta) = \{y \in \mathbf{R}^n : |x-y| < \delta\}$. In this section, $d\mu$ is any measure that is doubling with respect to these balls, and we note two elementary facts.

(a) $\mu(\mathbf{R}^n) = \infty$, unless $\mu(\mathbf{R}^n) = 0$. To see this, observe that there is a $c > 1$ so that $\mu(B(0,2\delta)) \ge c\mu(B(0,\delta))$ for all $\delta > 0$; iterating this inequality shows that $\mu(\mathbf{R}^n) = \infty$.

(b) If $S \subset \mathbf{R}^n$ is a smooth submanifold of dimension $k < n$, then $\mu(S) = 0$. In particular, $\mu(\{x\}) = 0$ for all $x \in \mathbf{R}^n$. Indeed, let $x \in S$, and let ν be a unit normal to S at x. Then for small δ, $B(x + (\delta\nu/2), \delta/4) \cap S = \emptyset$, and therefore

$$\mu(S \cap B(x, \delta)) \leq c\mu(B(x, \delta))$$

for a fixed $c < 1$ that does not depend on δ (provided δ is small). Now apply the corollary in §3 to $f = \chi_S$.

8.7 (a) Let P be a polynomial on $\mathbf{R}^n$ of degree d. If dx is Lebesgue measure, then $d\mu(x) = |P(x)|^a \, dx$ is a doubling measure (with respect to the standard balls) whenever $a > -1/d$. This is essentially in Ricci and Stein [1987]; related results are in Chapter 5, §6.5.

(b) A variant of this result is as follows. Suppose M is a smooth compact manifold, and f is a smooth function on M that does not vanish to infinite order at any point. Let $d\mu = |f|^a \, d\sigma$, where $d\sigma$ is the measure on M induced by some Riemannian metric. Then there is a positive ε so that $d\mu$ is a doubling measure when $a > -\varepsilon$. More precisely, let k be the smallest integer so that, for each $x \in M$, there is an α with $|\alpha| \leq k$ so that, in some coordinate system, $\partial_x^\alpha f(x) \neq 0$; then we can take $\varepsilon = 1/k$.

8.8 (a) There exist doubling measures (with respect to the usual balls in $\mathbf{R}^n$) that are totally singular. Indeed, on $\mathbf{R}^1$, the Riesz product

$$d\mu = \prod_{k=1}^{\infty} [1 + a\cos(3^k \cdot 2\pi x)] \, dx, \qquad \text{where } -1 < a < 1,$$

is such a measure.

For the proof that $d\mu$ is totally singular, see Zygmund [1959], Chapter 5. To verify the doubling property, it suffices to check that $d\mu(I) \approx d\mu(J)$, where $I = [(\ell - 1)/3^j, \ell/3^j]$, $J = [\ell/3^j, (\ell + 1)/3^j]$ are two adjacent intervals of length 3^{-j}. Now

$$d\mu = P_j(x) \prod_{k=j}^{\infty} [1 + a\cos(3^k \cdot 2\pi x)] \, dx,$$

and it is easily seen that $P_j(x) \approx P_j(\bar{x})$, if $|x - \bar{x}| \leq c3^{-j}$.

(b) There exist doubling measures $d\mu = f\,dx$ that are absolutely continuous, but where f vanishes on a set of positive measure.

To see this, partition $\mathbf{R}^1$ using the measure $d\mu$ above, so that $\mathbf{R}^1 = A \cup B$ with $\mu(A) = 0$, $|A| > 0$, and $\mu(B) > 0$, $|B| = 0$. Now let

$$F : \mathbf{R}^1 \to \mathbf{R}^1, \qquad F(x) = \int_0^x (d\mu + dx).$$

Since μ has no atoms, F is an increasing homeomorphism of the line, mapping intervals to intervals, and converts $d\mu + dx$ to dx. Then $E_1 = F(A)$ and

$E_2 = F(B)$ are disjoint sets of positive Lebesgue measure whose union is $\mathbf{R}^1$, while both $\chi_{E_1}\,dx$ and $\chi_{E_2}\,dx$ are doubling measures. Indeed, let I and J be adjacent intervals of the same length. Since $d\mu$ is doubling,

$$|I \cap E_1| = |F^{-1}(I)| \approx |F^{-1}(J)| = |J \cap E_1|,$$

from which it follows that

$$|I \cap E_2| = \mu(F^{-1}(I)) \approx \mu(F^{-1}(J)) = |J \cap E_2|.$$

Journé [1989].

8.9 Doubling measures arise in a natural way in various problems in analysis.

(a) Suppose that $d\mu$ is a doubling measure on $\mathbf{R}^1$, and let $F(x) = \int_{x_0}^x d\mu$. Then $F : \mathbf{R}^1 \to \mathbf{R}^1$ extends to a quasi-conformal homeomorphism of the closed upper half-plane $\mathbf{R}^2_+$ to itself. Conversely, every such mapping, provided that it preserves orientation, gives rise to a doubling measure as above when restricted to the boundary. See Beurling and Ahlfors [1956]; this paper also contains further examples of singular doubling measures.

(b) Suppose Ω is a bounded domain in $\mathbf{R}^n$ with smooth boundary $M = b\Omega$ and that $\{a_{ij}(x)\}$ is bounded, measurable, symmetric, and strictly positive definite on Ω. We consider the equation

$$\sum_{i,j} \frac{\partial}{\partial x_j}\left(a_{ij}\frac{\partial u}{\partial x_i}\right) = 0 \quad \text{on } \Omega, \qquad \text{with } u|_{b\Omega} = f.$$

Fix a point $\bar{x} \in \Omega$ and consider the "harmonic measure" $d\mu$ associated to $\bar{x}$, defined by

$$u(\bar{x}) = \int_M f(x)\, d\mu(x).$$

In general, $d\mu$ is a doubling measure (with respect to the usual balls defined on M), but it may be singular. See Caffarelli, Fabes, Mortola, and Salsa [1981], Caffarelli, Fabes, and Kenig [1981], and §6.20 of Chapter 5.

8.10 The quasi-distance ρ in §2.4 can always be replaced by an equivalent $\tilde{\rho}$ that satisfies a "Hölder condition". More precisely, we can find a $\tilde{\rho}$ with $\tilde{\rho} \approx \rho$, so that $\tilde{\rho}(x, y) = \tilde{\rho}(y, x)$ and

$$|\tilde{\rho}(x, z) - \tilde{\rho}(y, z)| \leq C\tilde{\rho}(x, y)^\gamma [\tilde{\rho}(x, z) + \tilde{\rho}(y, z)]^{1-\gamma}$$

for some fixed constants C and γ, $0 < \gamma \leq 1$. The exponent γ can be chosen to depend only on the structural constant c_1 appearing in §1.1. Macias and Segovia [1979a].

8.11 Suppose that for each δ, B_δ is an open, bounded, convex, and symmetric subset of $\mathbf{R}^n$ so that B_δ increases with δ. As remarked in §2.2, if $d\mu$ is the Lebesgue measure dx, then the family of balls $B(x, \delta) = x + B_\delta$ satisfies §1.2. As a result, the maximal theorem (Theorem 1) holds in this setting also. Examples where this situation occurs are in Marcinkiewicz and Zygmund [1939b], Carbery, Christ, Vance, Wainger, and D. Watson [1989].

B. Maximal functions and singular integrals

8.12 Suppose $V(x,y)$ is the volume function appearing in §6.5. Set $f(x) = 1/V(x,y)$, with y fixed, and let $\lambda(\alpha)$ be the distribution function of f; that is, $\lambda(\alpha) = |\{x : f(x) > \alpha\}|$.

(i) We have $\lambda(\alpha) \le \alpha^{-1}$, for all $\alpha > 0$.

(ii) If we make the additional assumption that $\mu(B(x,\delta))$ is continuous in δ, $0 < \delta < \infty$, then we have the more precise assertion:

$$\lambda(\alpha) = \min(\alpha^{-1}, \mu(\mathbf{R}^n)), \qquad \text{all } \alpha > 0.$$

(iii) When the assumption in (ii) is satisfied, we also have that

$$\int_{{}^cB} V(x,y)^{-p}\, d\mu(x) = (p-1)^{-1}[\mu(B)^{1-p} - \mu(\mathbf{R}^n)^{1-p}],$$

whenever B is a ball centered at y, and $p > 1$. When $p = 1$ the integral diverges, unless $\mu(\mathbf{R}^n) < \infty$, in which case it equals $\log(\mu(\mathbf{R}^n)/\mu(B))$.

To prove (i) and (ii), observe that $\{x : V(x,y) < \alpha^{-1}\} = \bigcup B(y,\delta)$, where the union is taken over all balls $B(y,\delta)$ with $\mu(B(y,\delta)) < \alpha^{-1}$.

8.13 (a) The proof of $\|Mf\|_{L^p} \le A_p\|f\|_{L^p}$ given in §2.1 shows that $A_p = O([p-1]^{-1})$, as $p \to 1$. To see that this bound is best possible, let $B_1 = B(y,\delta)$, $B = B(y, c\delta)$, where c is a large constant. Set $f = \mu(B_1)^{-1/p}\chi_{B_1}$, then $\|f\|_{L^p} = 1$. However,

$$(Mf)(x) \ge c'\mu(B_1)^{1-1/p}[V(x,y)]^{-1}, \quad \text{if } x \in {}^cB.$$

Thus if we apply §8.12, and let $\delta \to 0$, we get that $A_p \ge \bar{c}(p-1)^{-1}$ as $p \to 1$. One can prove similarly that M is not bounded on L^1. Note that this argument uses the assumption that $\mu(B(x,\delta))$ is continuous in δ; this premise can be dropped (see §8.14 below).

(b) The proof that $\|Tf\|_{L^p} \le A_p\|f\|_{L^p}$ for the singular integral operators T given in §5.1 shows that $A_p = O([p-1]^{-1})$, as $p \to 1$. In general, this bound is best possible. Indeed, assume that the kernel K of T satisfies $|K(x,y)| \ge c[V(x,y)]^{-1}$, as well as the regularity property (18′). Then for $x \notin B$, we have that

$$(Tf)(x) = \mu(B_1)^{-1+1/p}K(x,y) + \int_{B_1} [K(x,y) - K(x,\bar{y})]\, f(\bar{y})\, d\mu(\bar{y}),$$

where f and B_1 are as above.

The estimate (from below) of the first term is again a consequence of §8.12, while the second term provides an inessential contribution, as the argument in §6.5 shows. A similar proof shows that T is not bounded on L^1. If we make the further regularity assumption (29) (so that duality applies), then we can see that the estimate $A_p = O(p)$ as $p \to \infty$ holds, is best possible, and moreover that T does not extend to a bounded operator from L^∞ to itself.

8.14 The crucial weak-type inequality for the maximal function may be reversed. In fact, for appropriate constants c, $\bar{c}$, we have

$$\mu\{x : (Mf)(x) > \bar{c}\alpha\} \ge \frac{c}{\alpha}\int_{|f|>\alpha} |f|\, dx.$$

To prove this, let $E = \{x : (\tilde{M}f)(x) > \alpha\}$, and decompose E as $\bigcup Q_k$, according to §4.1. Since B_k^{**} intersects the complement of E, we have

$$\int_{B_k^{**}} |f|\, dx \le \alpha\mu(B_k^{**}),$$

and thus $\int_{Q_k} |f|\, dx \le c\alpha\mu(Q_k)$. Adding these inequalities and using the fact that $|f| \le \tilde{M}(f) \le c_1 M(f)$ gives the asserted inequality with $\bar{c} = 1/c_1$. As a consequence, it is not difficult to show the following.

(a) Suppose f is supported in a set of finite measure; then Mf is integrable on sets of finite measure if and only if $|f|\log(1+|f|)$ is integrable.

(b) There is a constant c_p, $p > 1$, with $c_p \ge c/(p-1)$, so that if $f \in L^p(\mathbf{R}^n)$ then $\|Mf\|_{L^p} \ge c_p\|f\|_{L^p}$. Moreover, M is not bounded on L^1.

See also Chapter 1, §5.2 of *Singular Integrals*.

8.15 (a) If $f \in L^1(\mathbf{R}^n, d\mu)$ then $Mf \in L^q(E)$, whenever $0 < q < 1$ and $E \subset \mathbf{R}^n$ has finite μ-measure. In fact, for $f \in L^1$ we have

$$\int_E [(Mf)(x)]^q\, d\mu(x) \le c_q\mu(E)^{1-q}\|f\|_{L^1}^q.$$

(b) The weak-type inequality for the maximal function goes through when the L^1 function f, more precisely the measure $f(x)\, d\mu(x)$, is replaced by a measure dm (possibly singular with respect to $d\mu$), if dm is supposed to have finite total mass. Indeed, with such a measure dm, let

$$M(dm)(x) = \sup_{\delta>0} \frac{1}{\mu(B(x,\delta))}\int_{B(x,\delta)} |dm|.$$

Then $\mu\{x : M(dm)(x) > \alpha\} \le c\alpha^{-1}\int_{\mathbf{R}^n} |dm|$. Also, the L^q inequality in (a) holds, with $\|f\|_{L^1}$ replaced by $\int_{\mathbf{R}^n} |dm|$.

To prove (a), let $\lambda(\alpha) = \mu(\{x : Mf(x) > \alpha\} \cap E)$. Then

$$\int_E (Mf)^q\, d\mu = q\int_0^\infty \alpha^{q-1}\lambda(\alpha)\, d\alpha = \int_0^A + \int_A^\infty.$$

Now choose $A = \|f\|_{L^1}/\mu(E)$ and use the fact that $\lambda(\alpha) \le \mu(E)$ in the first integral, while $\lambda(\alpha) \le c\alpha^{-1}\|f\|_{L^1}$ in the second.

That the maximal inequalities for L^1 functions extend to finite measures as asserted is evident from the proofs given in §3.

8.16 The corollary in §3.1 has the following extension. Whenever f is locally integrable, then

$$\frac{1}{\mu(B(x,\delta))}\int_{B(x,\delta)}|f(y)-f(x)|\,d\mu(x)\to 0, \quad \text{as } \delta\to 0,$$

for almost every x. The set of x for which this holds is called the "Lebesgue set" of f.

In the case of $\mathbf{R}^n$ with its usual Euclidean structure, if $\{\Phi_t\}$ is an approximation of the identity (as in §6.1), then $\lim_{t\to 0}(f*\Phi_t)(x)=f(x)$ for every x in the Lebesgue set of f, whenever $f\in L^p(\mathbf{R}^n)$, for some p, $1\le p\le\infty$.

8.17 The Besicovitch covering lemma gives a more refined version of the lemma in §3.1, in the setting of $\mathbf{R}^n$ with the standard Euclidean balls. It has the advantage of being applicable when the underlying measure is not assumed to be doubling.

The lemma states that there is a constant c_n so that the following holds. Suppose $\{B_{x^\alpha}\}$ is a collection of balls, with B_{x^α} centered at x^α; assume also that the set $E=\{x^\alpha\}$ is a bounded subset of $\mathbf{R}^n$. Then there is a subcollection $\{B'\}$ that covers E, and so that no point belongs to more than c_n of the B'.

This has the following consequences. Let μ and ν be positive measures on $\mathbf{R}^n$ so that, for all balls B, $\nu(B)=0$ implies $\mu(B)=0$. Define

$$(M_\nu f)(x)=\sup_{\delta>0}\frac{1}{\nu(B(x,\delta))}\int_{B(x,\delta)}|f(y)|\,d\mu(y),$$

where $B(x,\delta)=\{y:|x-y|<\delta\}$. Then

$$\nu\{x:(M_\nu f)(x)>\alpha\}\le\frac{c_n}{\alpha}\int|f(y)|\,d\mu(y).$$

In particular, if we take $\nu=\mu$ we see that the maximal theorem in §3.1 holds, for the standard Euclidean balls, without requiring that $d\mu$ be a doubling measure. Note however that the assertion is made for the centered maximal operator M, and not for the uncentered version $\tilde{M}$.

Similar conclusions also hold for certain families besides the centered Euclidean balls. See Besicovitch [1945], Morse [1947], de Guzmán [1981]; the case of rectangles is described in *Fourier Analysis*, Chapter 2.

8.18 We summarize briefly the theory of singular integrals as presented in Calderón and Zygmund [1952]. Suppose we are given a function $K_0(x)$ that is homogeneous of degree $-n$ on $\mathbf{R}^n$, with $|K_0(x)|\le A|x|^{-n}$, so that $|K_0(x-y)-K_0(x)|\le A\eta(|y|)$ when $|x|=1$, as $y\to 0$; here (as in §6.5) η is a Dini modulus of continuity. We assume also the cancellation condition $\int_{|x|=1}K_0(x)\,d\sigma(x)=0$. Let us define the principal-value distribution $K=$ p.v. K_0 by

$$K(f)=\lim_{\varepsilon\to 0}\int_{|x|\ge\varepsilon}f(x)\,K_0(x)\,dx,$$

for $f\in\mathcal{S}$. The following assertions then hold.

(a) $\widehat{K}$ is a bounded function on $\mathbf{R}^n$, and hence the convolution operator $Tf=f*K$ extends to a bounded operator on $L^2(\mathbf{R}^n)$.

(b) The operator T satisfies the assumptions (10), (21)–(23), and (29) (in the setting of $\mathbf{R}^n$ with the usual balls, with $d\mu=dx$ and $q=2$). Thus all the conclusions stated in Theorem 3 and the Appendix §7 apply to T.

The proof of the boundedness of $\widehat{K}$ can be given by adapting the argument in §4.5 of Chapter 6, where somewhat more regularity of K is required.

8.19 The distributions K that arise in §8.18 are *homogeneous distributions* (having degree $-n$). If we assume further regularity, such distributions can be characterized by the following four equivalent properties.

(1) The distribution K is of the form $c\delta+\text{p.v.}K_0$, where δ is the Dirac delta function, and K_0 is a Calderón-Zygmund kernel (of the type specified in §8.18) with the additional property that K_0 is C^∞ away from the origin.

(2) K is homogeneous of degree $-n$ and, away from the origin, agrees with a C^∞ function. The first statement means that $K(\phi_t)=t^{-n}K(\phi)$ for all $t>0$ and $\phi\in\mathcal{S}$; here we write $\phi_t(x)=t^{-n}\phi(x/t)$.

(3) $\widehat{K}$ is a function that is homogeneous of degree 0 and is C^∞ away from the origin.

(4) K can be written as $\lim_{\varepsilon\to 0,N\to\infty}\int_\varepsilon^N\Phi_t\,dt/t$ for some $\Phi\in\mathcal{S}$ with $\int\Phi\,dx=0$.

The equivalence of (1), (2), and (3) may be proved by using arguments of the kind that appear in Chapter 6, §4.4; see also §7.5 of that chapter, as well as *Singular Integrals*, Chapter 3. The assertion (3) follows directly from (4), via the formula $\widehat{K}(\xi)=\int_0^\infty\widehat{\Phi}(t\xi)\,dt/t$. Conversely, if $\widehat{K}$ is given, one may obtain a representation (4) by taking $\widehat{\Phi}(\xi)=\eta(|\xi|)\widehat{K}(\xi)$, where $\eta\in C_0^\infty([1,2])$ and $\int_1^2\eta(t)\,dt/t=1$.

Some related results for homogeneous distributions of degree $d\ne -n$ can be found in Chapter 6, §7.5.

8.20 The results in §8.18 and §8.19 refer to homogeneity in the setting of isotropic dilations. There are closely parallel analogues that hold in the context of the nonisotropic dilations described in §2.3. Some further details may be found in B. Jones [1964], Fabes and Rivière [1966], Kree [1965].

8.21 The Hardy-Littlewood-Sobolev inequality for fractional integration extends to the general context treated in §1. Indeed, if

$$(I_\alpha f)(x)=\int_{\mathbf{R}^n}[V(x,y)]^{-1+\alpha}f(y)\,d\mu(y),$$

then $\|I_\alpha f\|_q\le A_{p,q}\|f\|_p$ whenever $1<p<q<\infty$ and $q^{-1}=p^{-1}-\alpha$.

The proof can be given by adapting the argument in §4.2 of Chapter 8 or, alternatively, by using the reasoning of *Singular Integrals*, Chapter 5. In applications, the relevant operator is more often

$$(J_\beta f)(x) = \int [\rho(x,y)]^\beta \, [V(x,y)]^{-1} f(y)\, d\mu(y).$$

Similar conclusions then hold with $\alpha = \beta\gamma$, if $\rho(x,y) \le c[V(x,y)]^\gamma$. In particular, the operator J_2 gives a majorant for the fundamental solution of the sub-Laplacian $\sum_{j=1}^n X_j^2$ arising in §8.4.

8.22 Let T be the fractional integration operator, having imaginary order, that is defined (via the Fourier transform) by

$$\widehat{Tf}(\xi) = |\xi|^{-it} \cdot \widehat{f}(\xi);$$

here $t \neq 0$ is a fixed real number. Then $Tf = f * K$, and the distribution K, away from the origin, agrees with the function

$$K(x) = \gamma_t^{-1} |x|^{-n+it},$$

where γ_t is an appropriate constant.

The operator T is of a kind described in §6.2: it is bounded on $L^2(\mathbf{R}^n)$ and the differential inequalities (18_γ) hold. As a result, it also enjoys the properties stated in (10), (21)-(23), and (29), and all the conclusions of the appendix §7 apply to it. However, if T_ε is the corresponding truncated operator, it is not true that $T_\varepsilon(f)$ converges as $\varepsilon \to 0$. Indeed, if f is (say) smooth with compact support, then

$$(T_{\varepsilon_1} - T_{\varepsilon_2})f(x) - c_t(\varepsilon_1^{it} - \varepsilon_2^{it})f(x) \to 0, \qquad \text{as } \varepsilon_1, \varepsilon_2 \to 0.$$

See Muckenhoupt [1960]; also *Singular Integrals*, Chapter 2, §6.12.

C. Vector-valued singular integrals

8.23 Whenever $\Phi \in \mathcal{S}$, and $\int_{\mathbf{R}^n} \Phi \, dx = 0$, we have defined in §6.3 the square functions s_Φ and S_Φ. One has:

(a) $\|s_\Phi(f)\|_{L^p} \le A_p \|f\|_{L^p}$, $1 < p < \infty$, with a similar inequality for S_Φ.

(b) Suppose Φ is nondegenerate, in the sense that there exists a $\Psi \in \mathcal{S}$ with $\int_{\mathbf{R}^n} \Psi \, dx = 0$ so that

$$\int_0^\infty \widehat{\Phi}(t\xi)\, \widehat{\Psi}(t\xi) \frac{dt}{t} = 1.$$

Then we have the converse inequality $\|f\|_{L^p} \le A'_p \|s_\Phi(f)\|_{L^p}$, for $f \in L^p(\mathbf{R}^n)$, $1 < p < \infty$; again the analogous result holds for S_Φ.

(c) Similar results are valid under weaker assumptions on Φ and Ψ. For instance, it suffices to have $|\Phi(x)| \le A(1+|x|)^{-n-1}$, $|\nabla\Phi(x)| \le A(1+|x|)^{-n-1}$, with the same conditions on Ψ; we must assume, of course, that the integrals of Φ and Ψ vanish. An important example occurs when these functions arise as the first derivatives of the Poisson kernel; then the corresponding s_Φ gives the n-dimensional version of the Littlewood-Paley g-function, and S_Φ is the area integral of Lusin.

Indeed, the inequalities (a) for $1 < p \le 2$ are treated in §6.3. The case $p > 2$ follows by duality: As in §6.4, think of $s_\Phi(x)$ as the norm of the vector $(\mathbf{s}_\Phi f)(x)$ in the Hilbert space $H = L^2[(0,\infty); dt/t]$; $\mathbf{s}_\Phi$ is then a linear operator that is given by convolution with an H-valued kernel. Now observe that $[L^p(\mathbf{R}^n, H)]^* = L^{p'}(\mathbf{R}^n, H^*) = L^{p'}(\mathbf{R}^n, H)$ and use the inequality for $1 < p < 2$. The same argument works for S_Φ.

To prove (b), one notes that if $\tilde{\Psi}(x) = \overline{\Psi}(-x)$, then

$$\langle f * \Phi_t, g * \tilde{\Psi}_t \rangle = \langle f * \Phi_t * \Psi_t, g \rangle,$$

and so $\langle f, g \rangle = \int_0^\infty \langle f * \Phi_t, g * \tilde{\Psi}_t \rangle \, dt/t$. Thus

$$|\langle f, g \rangle| = \langle \mathbf{s}_\Phi(f), \mathbf{s}_{\tilde{\Psi}}(g) \rangle \le \langle s_\Phi(f), s_{\tilde{\Psi}}(g) \rangle;$$

and our assertion follows from the direct inequality (a). The corresponding result for S_Φ is proved in the same way, if one uses the integral identities found in Chapter 3, §4.4.3.

For the condition of nondegeneracy imposed on Φ, see also §6.19 of Chapter 4. Further information about the g-function and area integral can be found in *Singular Integrals*, Chapter 4.

8.24 Let $T : L^p(\mathbf{R}^n) \to L^p(\mathbf{R}^n)$ be a bounded linear mapping of scalar-valued functions, for some p, $1 \le p \le \infty$. Let B be a Banach space and consider (as in §6.4) the space L^p_B of strongly measurable‡ B-valued functions f, for which $|f|_B \in L^p(\mathbf{R}^n)$. We define the extension T_B of T to B-valued functions by

$$T_B(f \otimes v) = (Tf) \otimes v,$$

when $v \in B$, and f is scalar valued. The question we address is: Given T, for what B is T_B bounded from L^p_B to itself?

If B is a Hilbert space and $T : L^p \to L^p$ is an arbitrary bounded linear transformation, then $T_B : L^p_B \to L^p_B$ is also bounded, with the same norm. This is essentially proved in Chapter 10, §2.5.1, and goes back to Marcinkiewicz and Zygmund [1939a].

8.25 (a) Continuing the discussion in §8.24, we suppose that T is a singular integral operator that satisfies conditions (10), (21)-(23), and (29), $\mathcal{M}$ is a general measure space, and $B = L^r(\mathcal{M})$, where $1 < r < \infty$. Then $T_B : L^p_B \to L^p_B$ is bounded for $1 < p < \infty$.

In fact, when $p = r$, the result for T_B is immediate from that for T; one then uses Theorem 3 (in the vector-valued form indicated in §6.4) with $q = r$. See Benedek, Calderón, and Panzone [1962].

‡ A definition of strong measurability may be found in, e.g., Journé [1983].

(b) Let B be a "noncommutative" analogue of an L^r space. An example is the trace-class space C_r: it consists of all bounded operators A on a fixed Hilbert space for which

$$\|A\|_{C_r} = (\operatorname{tr}(A^*A)^{r/2})^{1/r} < \infty.$$

In this setting, results like those in (a) are valid when $B = C_r$, $1 < r < \infty$. See J. A. Guitérrez [1982], Bourgain [1986b]; also Gohberg and Krein [1970], E. Davies [1988].

8.26 In connection with §8.24, the following condition ("ζ-convexity") on a Banach space B is decisive: There exists a function $\zeta : B \times B \to \mathbf{R}$ that is convex in each variable separately, so that $\zeta(x, y) \le |x + y|$ whenever $x, y \in B$ with $|x| = |y| = 1$, and with $\zeta(0, 0) > 0$.

(i) If T is one of the singular integral operators considered in §8.25 and $T_B : L^p_B \to L^p_B$ is bounded for some p, then B is ζ-convex.

(ii) Conversely, when B is ζ-convex, then $T_B : L^p_B \to L^p_B$ is bounded for all p, $1 < p < \infty$, for a large class of such operators T.

The proofs require consideration of probabilistic analogues of the operators T, given as multipliers involving martingale differences. Burkholder [1981] and [1983], McConnell [1984], Bourgain [1983].

Notes

The model for the real variable methods described in §1–§5 is the theory of maximal functions and singular integrals in the standard translation-invariant setting of $\mathbf{R}^n$, as may be found in the first two chapters of *Singular Integrals.* The key source of that theory is the paper of Calderón and Zygmund [1952], together with some earlier work in $\mathbf{R}^1$ by Besicovitch, Titchmarsh, and Marcinkiewicz.

The general point of view set out here has many roots in past work. Among these are: an ergodic theorem in Calderón [1953]; a paper of Smith [1956] on maximal functions for Poisson integrals in domains in $\mathbf{R}^n$; the development of the nonisotropic (translation-invariant) theory of singular integrals of B. Jones [1964], Fabes and Rivière [1966], Sadosky [1966]; the maximal functions on homogeneous groups in Stein [1968]; its application to Fatou's theorem for the complex ball in Korányi [1969]; generalizations to boundary behavior of holomorphic functions on domains in $\mathbf{C}^n$ by Stein [1972]; and the singular integrals that appear as intertwining operators in Knapp and Stein [1971]. Several more systematic approaches were then developed in Korányi and Vági [1971], Coifman and G. Weiss [1971]. It is the latter we have followed more closely in this chapter.

Further details concerning the examples and topics mentioned in §6—such as the Riesz transforms, translation-invariant singular integrals, their vector valued versions and square functions, and multiplier theorems—may be found in *Singular Integrals*, chapters 2–4.

CHAPTER II

More about Maximal Functions

The basic properties of the maximal function, which were the subject of Theorem 1 in the previous chapter, were obtained as a direct consequence of the real-variable structure described there. It turns out that significant extensions and refinements of these properties can be derived from the same circle of ideas. These deeper results are interesting in their own right, but they also foreshadow some later developments of importance. Our presentation of this material will be organized along three main lines.

(1) *Vector-valued inequalities.* The passage to a Hilbert space valued version of a (scalar valued) operator is a useful device that arises in many situations. When the operator in question is linear, this technical step is subsumed under a general theorem of Marcinkiewicz and Zygmund (see Chapter 1, §8.24) However, the maximal operator M cannot be treated by this method; this is because it is not linear or, put another way, although it can be reformulated as a linear operator, it then takes its values in L^∞ (and not in a Hilbert space). Thus, the maximal operator requires its own further analysis. This analysis is based in part on a weighted inequality that anticipates some of the ideas treated in Chapter 5, and particularly the role of the class A_1.

(2) *The tent space $\mathcal{N}$.* The importance of nontangential behavior is highlighted by the definition of a certain function space $\mathcal{N}$. A key point here is that the dual of this space consists of the Carleson measures or, equivalently, that functions in $\mathcal{N}$ have an atomic decomposition of a simple nature. These facts anticipate fundamental theorems taken up later, such as the duality between H^1 and BMO and the atomic decomposition in H^p. In addition, the space $\mathcal{N}$ is useful in a variety of applications. One that we describe allows us to characterize those collections $\mathcal{B}$ of (standard) balls in $\mathbf{R}^n$ for which maximal operators fashioned from $\mathcal{B}$ satisfy analogues of the usual L^p and weak-type (1,1) inequalities.

(3) *Singular approximations to the identity.* These do not admit a pointwise majorization by the standard maximal function but nevertheless do arise in a variety of situations, most interestingly for Poisson integrals on symmetric spaces. The relevant weak-type and L^p inequalities still hold, but the proof requires that we use the Calderón-Zygmund

decomposition (Chapter 1, §4) and, in effect, that we think of the corresponding maximal operator as being made up of vector valued singular integrals.

Two remarks about our exposition here are in order. First, on a minor note, and as was mentioned above, our presentation is limited to the classical real variable setting of $\mathbf{R}^n$ with the usual Euclidean balls. However, given the ideas presented in Chapter 1, the extension of this material to the general situation treated there is a routine exercise for many of the results in question. Second, and more importantly, the results of the present chapter (together with the weighted inequalities of Chapter 5) may well represent the limit of what can be understood by using only the real-variable theory centered around covering lemmas. Important further developments in the theory of maximal functions, which involve the use of orthogonality (via the Fourier transform and oscillatory integrals), are treated in chapters 10 and 11.

1. Vector-valued maximal functions

1.1 As we have said above, we shall present the theory in this chapter in the usual setting of $\mathbf{R}^n$. In the present context, the maximal function discussed in the previous chapter becomes

$$Mf(x) = \sup_{r>0} \frac{c_n}{r^n} \int_{|y|\le r} |f(x-y)|\, dy.$$

(The set $\{|y| < r\}$ is the standard Euclidean ball of radius r, and r^n/c_n is its volume.)

Our first aim is to extend the basic L^p and weak-type inequalities for the maximal function (Theorem 1 of Chapter 1) to vector-valued functions. The most natural generalization of this type occurs when our functions take their values in a Hilbert space.

It will be convenient to fix the Hilbert space as the usual sequence space l^2. Thus we envisage the following situation. We write

$$f(x) = \{f_j(x)\}_{j=1}^{\infty},$$

where each f_j is a complex-valued function, and we set

$$|f(x)| = \Big(\sum_{j=1}^{\infty} |f_j(x)|^2\Big)^{1/2} = \|f(x)\|_{l^2}.$$

With this definition in mind, we say that $f = \{f_j\}$ belongs to L^p if each f_j is measurable and $|f(x)| \in L^p$. We write $\|f\|_p = \|\,|f|\,\|_{L^p}$.

We next define the vector-valued maximal operator $\overline{M}$ by

$$\overline{M}f(x) = \Big(\sum_{j=1}^{\infty} (Mf_j(x))^2\Big)^{1/2}. \tag{1}$$

The generalization of the maximal theorem of the previous chapter can then be stated as follows.

THEOREM 1. (a) *If $f \in L^p$, $1 \le p < \infty$, then $\overline{M}f$ is finite almost everywhere.*

(b) *If $f \in L^1$ then, for every $\alpha > 0$,*†

$$|\{x : \overline{M}f(x) > \alpha\}| \le \frac{A}{\alpha} \int_{\mathbf{R}^n} |f(y)|\, dy. \tag{2}$$

(c) *If $f \in L^p$, $1 < p < \infty$, then $\overline{M}f \in L^p$ and*

$$\|\overline{M}f\|_p \le A_p \|f\|_p. \tag{3}$$

Before we come to the proof, three clarifying comments may be in order.

(i) As opposed to the case when one deals with a linear operator, the vector-valued inequalities are not necessarily consequences of the scalar-valued case. (For the case of a linear operator, see §8.24 in Chapter 1.)

(ii) The untoward effect of the nonlinearity of $\overline{M}$ is highlighted when we consider the L^∞ case. For this purpose take $\mathbf{R}^1$, and set $f_j = \chi_{(2^{j-1},2^j)}$, $j = 1, 2, \ldots$. Then $|f| = \chi_{(1,\infty)} \in L^\infty$, but since $Mf_j(x) \ge 1/8$ if $|x| \le 2^j$, we get that $(\overline{M}f)^2(x) \ge \sum_{2^j \ge |x|} 1/64$; hence $\overline{M}f(x) = \infty$ everywhere.

The unboundedness of $\overline{M}$ on L^∞ is also reflected in the fact that the bound A_p appearing in (3) is of the order $p^{1/2}$ as $p \to \infty$. For further discussion and a substitute result for bounded functions with compact support, see §5.2 below.

(iii) While the "trivial" case $p = \infty$ fails, and so cannot be used in proving the inequalities for $\overline{M}$, the starting point in the present situation is the case $p = 2$ of the inequality (3), which follows from Theorem 1 of Chapter 1 because

$$\|\overline{M}f\|_2^2 = \sum_j \|Mf_j\|_2^2 \le A \sum_j \|f_j\|_2^2 = A\|f\|_2^2. \tag{4}$$

1.2 Weak-type inequality. In this section, we prove the weak-type inequality (2). We will use a variant of the Calderón-Zygmund decomposition (§4 of Chapter 1). Here it is important to observe that the

† Here and in the sequel, $|E|$ denotes the Lebesgue measure of the set E.

decomposition is valid for functions that take their values in a Banach space.

It suffices to prove (2) when f is nonnegative, in the sense that $f_j(x) \geq 0$, for all j and x. For any fixed $\alpha > 0$, the aforementioned construction gives us a collection $\{Q_k\}$ of disjoint "cubes"‡ so that $|f(x)| \leq \alpha$ on ${}^c\bigcup Q_k$ and

$$\frac{1}{|Q_k|}\int_{Q_k} |f(x)|\,dx \leq A\alpha, \quad \text{for all } k.$$

Now take $f = g+b$, where $g = f$ on ${}^c\bigcup Q_k$ and $b = f$ on $\bigcup Q_k$. Thus $|g(x)| \leq \min\{\alpha, |f(x)|\}$ and $\int |g|^2 \leq \alpha \int |f|$. Combining this with (4) gives

$$|\{\overline{M}g > \alpha/2\}| \leq \frac{4}{\alpha^2}\|\overline{M}g\|_2^2 \leq \frac{A}{\alpha^2}\|g\|_2^2 \leq \frac{A}{\alpha}\|f\|_1. \tag{5}$$

Since $\overline{M}f \leq \overline{M}g + \overline{M}b$, it suffices to prove that

$$|\{\overline{M}b > \alpha/2\}| \leq \frac{A}{\alpha}\|f\|_1. \tag{6}$$

We prove (6) by deducing it from a simpler variant, to wit, the one obtained by replacing the function b (supported on the cubes Q_k) by its average value on each cube. Thus we set $b^o(x) = |Q_k|^{-1}\int_{Q_k} f(y)\,dy$ if $x \in Q_k$, and $b^o(x) = 0$ if $x \notin \bigcup Q_k$.

Now $|b^o(x)| \leq |Q_k|^{-1}\int_{Q_k} |f(y)|\,dy \leq A\alpha$ on each Q_k. Recalling that $\sum |Q_k| \leq A\|f\|_1/\alpha$, we get $\|b^o\|_2^2 \leq A^2\alpha^2 \sum |Q_k| \leq A\alpha\|f\|_1$. Using again the case $p = 2$ gives

$$|\{\overline{M}b^o > \alpha/2\}| \leq \frac{4}{\alpha^2}\|\overline{M}b^o\|_2^2 \leq \frac{A}{\alpha^2}\|b^o\|_2^2 \leq \frac{A}{\alpha}\|f\|_1.$$

So we have proved (6) with b replaced by b^o, and finally it suffices to see that

$$\overline{M}b(x) \leq c\overline{M}b^o(x), \text{ whenever } x \notin \bigcup Q_k^*. \tag{7}$$

Here Q_k^* denotes the cube with the same center as Q_k and with twice the diameter. Notice that

$$\sum |Q_k^*| = c\sum |Q_k| \leq \frac{A}{\alpha}\|f\|_1,$$

so the set where $\overline{M}b$ is not controlled by (7) has acceptable size.

‡ In the present case of $\mathbf{R}^n$ with the standard structure, the Q_k can be taken to be genuine cubes.

Write $b_j = b\chi_{Q_j}$, $b_j^o = B^o\chi_{Q_j}$, so $b = \sum b_j$ and $b^o = \sum b_j^o$. To show (7), it is enough to observe that

$$Mb_j(x) \leq cMb_j^o(x), \quad \text{whenever } x \notin \bigcup Q_k^*.$$

Let $B = B(x,r)$, then $|B|^{-1}\int_B b_j = |B|^{-1}\sum_k \int_{B\cap Q_k} b_j$, where the sum ranges over those k with $B \cap Q_k \neq \emptyset$.

The key observation is that if $x \notin \bigcup Q_k^*$ and $B(x,r) \cap Q_k \neq \emptyset$, then $B(x,3r) \supset Q_k$. Therefore the above sum is bounded by

$$|B|^{-1}\sum_k \int_{Q_k} b_j = |B|^{-1}\sum_k \int_{Q_k} b_j^o \leq |B(x,r)|^{-1}\int_{B(x,3r)} b_j^o \leq 3^n Mb_j^o(x).$$

This proves (7) and, with it, the weak-type inequality (2). Also, the L^p inequality (3) for $1 < p \leq 2$ follows from this and the case $p = 2$ discussed first, by the Marcinkiewicz interpolation theorem (§5.4 in Chapter 1).

1.3 The case $p \geq 2$; weighted maximal inequality. To deal with the case $p \geq 2$, we shall need to examine the behavior of maximal inequalities when weights are introduced. This will be our first result of its kind, and here the class A_1 already makes its appearance. The general theory of weighted inequalities will be taken up in Chapter 5.

We take $\omega(x)$ to be a nonnegative, locally integrable function (a "weight"), and are interested in inequalities for the maximal function when Lebesgue measure dx is replaced by $\omega(x)\,dx$. We write $\omega(E) = \int_E \omega(x)\,dx$. The proposition below is the weighted version of the scalar-valued maximal inequality appropriate to this context.

PROPOSITION. *For ω as above and M the usual maximal operator, we have*

$$\omega\{Mf(x) > \alpha\} \leq \frac{A}{\alpha}\int_{\mathbf{R}^n} |f(x)|M\omega(x)\,dx \tag{8}$$

and

$$\int_{\mathbf{R}^n} (Mf(x))^q\omega(x)\,dx \leq A_q\int_{\mathbf{R}^n} |f(x)|^q M\omega(x)\,dx, \quad 1 < q < \infty, \tag{9}$$

with A and A_q independent of f and ω.†

† Of course, for (8) and (9) to be non-vacuous, $M\omega$ needs to be finite somewhere. Since ω is locally integrable, this happens exactly when $\int_{|x|\leq r}\omega(x)\,dx \leq Ar^n$ for all large r.

Our proposition is a direct generalization of the boundedness of the classical maximal operator (which corresponds here to the case $\omega \equiv 1$). Although its statement does not follow from the classical counterpart, the proof is nearly identical.

Proof. The inequality (9) says that $f \mapsto Mf$ is bounded from $L^q((M\omega)\,dx)$ to $L^q(\omega\,dx)$. By the Marcinkiewicz interpolation theorem, (9) is a consequence of the trivial fact that M is bounded on L^∞ together with the weak-type inequality (8), to which we now turn.

The proof of (8) is a reprise of the argument given for conclusion (b) of Theorem 1 in the first chapter. Let $E_\alpha = \{x : Mf(x) > \alpha\}$, and let E be any compact subset of E_α. For each $x \in E_\alpha$, there is a ball B_x centered at x with

$$|B_x| < \frac{1}{\alpha}\int_{B_x} |f(y)|\,dy.$$

We can then select a disjoint collection $B_1, \ldots, B_m$ of such balls so that

$$\bigcup_k B_k^* \supset E;$$

here B_k^* is the ball with the same center as B_k and three times the radius.

The key observation is that

$$\int_{B_k^*} \omega\,dx \le \frac{A}{\alpha}\int_{B_k} |f(y)|(M\omega)(y)\,dy. \tag{10}$$

Indeed,

$$\int_{B_k^*} \omega\,dx \le A|B_k|M\omega(y) \qquad \text{for any } y \in B_k,$$

because $|\tilde{B}_k|^{-1}\int_{\tilde{B}_k} \omega\,dx \le M\omega(y)$, where $\tilde{B}_k$ is the smallest ball centered at y that contains B_k^*. Note that the radius of $\tilde{B}_k$ is at most four times that of B_k; thus $|\tilde{B}_k| \le A|B_k|$. The positions of y, B_k, B_k^*, and $\tilde{B}_k$ are schematized in Figure 1.

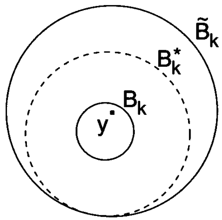

Figure 1. Situation in the proof of the proposition.

If we integrate the above over $y \in B_k$, and use the fact that

$$|B_k| \le \alpha^{-1}\int_{B_k} |f(y)|\,dy,$$

we get (10). Having established this, we can then sum both sides of (10) over k, using the disjointness of the B_k, to obtain

$$\omega(E) \le \frac{A}{\alpha}\int |f(x)|\,(M\omega)(x)\,dx,$$

which implies (8).

With the aid of the proposition we can now easily finish the proof of Theorem 1.

Returning to the case when $f = \{f_j\}$ is vector valued, we recall that $|f(x)| = (\sum |f_j(x)|^2)^{1/2}$ and $\overline{M}f(x) = (\sum |Mf_j(x)|^2)^{1/2}$. Then by (9), when $q = 2$, we have

$$\int (Mf_j)^2\,\omega\,dx \le A\int |f_j|^2(M\omega)\,dx,$$

and summing over j gives us

$$\int (\overline{M}f)^2\,\omega\,dx \le A\int |f|^2(M\omega)\,dx. \tag{11}$$

Suppose now that $2 \le p < \infty$, and let r denote the exponent dual to $p/2$ (thus $1 < r \le \infty$). If we take the supremum of (11) over $\omega \in L^r$ with $\|\omega\|_r \le 1$, the left side is $\|\overline{M}f\|_p^2$ (since $L^r = (L^{p/2})^*$), while the right side is majorized by

$$\|f\|_p^2 \cdot \|M\omega\|_r \le \|f\|_p^2 \cdot c_r\|\omega\|_r \le c_r\|f\|_p^2,$$

in view of the maximal theorem in §3 of Chapter 1. Therefore

$$\|\overline{M}f\|_p \le c_r^{1/2}\|f\|_p, \tag{12}$$

and (3) is proved for $2 \le p < \infty$, concluding the proof of Theorem 1.

1.3.1 A few additional remarks may be in order.

(i) The theorem just proved also holds when the sequence space l^2 is replaced by l^q, for some q, $1 < q \le \infty$; here $|f(x)| = |f(x)|_q = (\sum_j |f_j(x)|^q)^{1/q}$ when $1 < q < \infty$, $|f(x)| = |f(x)|_\infty = \sup_j |f_j(x)|$ when $q = \infty$. One then sets $\overline{M}_q f(x) = (\sum_j |Mf_j|^q(x))^{1/q}$ when $1 < q < \infty$, and $\overline{M}_\infty f(x) = \sup_j |Mf_j(x)|$. Then, in analogy with the case $q = 2$ handled by Theorem 1, one can show that

$$|\{\overline{M}_q f(x) > \alpha\}| \leq \frac{A_q}{\alpha} \int |f|_q \, dx \quad \text{and} \qquad \qquad (13)$$
$$\|\overline{M}_q f\|_p \leq A_{p,q} \| \, |f|_q \|_p, \quad 1 < p < \infty.$$

The proof for $1 < q < \infty$ is nearly identical to that for $q = 2$, except one begins with the easy case $p = q$, instead of $p = 2$. Observe also that $q = \infty$ reduces to the scalar-valued case, because then $\overline{M}_\infty f \leq M(|f|_\infty)$. Finally, it should be observed that analogues of (13) fail when $q = 1$ (see §5.1 below).

(ii) The weighted inequalities (8) and (9) naturally raise the question as to what happens when the weights on both sides of the inequality are the same. In view of the proposition, an obvious sufficient condition is that $M\omega \leq c\omega$. This is the A_1 condition; its variants, which give necessary and sufficient conditions for $q > 1$, are the subject of Chapter 5. One can also observe that (8) is sharp in the sense that, if $\omega\{Mf > \alpha\} \leq A/\alpha \int |f|\omega \, dx$, for all f and α, then $\omega \geq cM\omega$; this follows by choosing for f a sequence converging weakly to a point mass at an arbitrary $x \in \mathbf{R}^n$.

2. *Nontangential behavior and Carleson measures*

In many situations, the study of a function f defined on $\mathbf{R}^n$ can be closely connected with related properties of a corresponding function F defined on the (open) upper half-space $\mathbf{R}^{n+1}_+$, with F constructed from f by some averaging process. The simplest example arises when the average is carried out by a suitable "approximation of the identity" as follows.

Fix an integrable function Φ on $\mathbf{R}^n$ with $\int_{\mathbf{R}^n} \Phi \, dx = 1$. For $t > 0$, set $\Phi_t(x) = t^{-n}\Phi(x/t)$, so that $\int \Phi_t = 1$ for all t. Now let

$$F(x,t) = (f * \Phi_t)(x). \qquad (14)$$

Then, as is well known, for appropriate f and Φ, $F(x,t) \to f(x)$ as $t \to 0$ in a variety of senses. Some are quite easy to see, such as convergence in the L^p norm for $f \in L^p$, $1 \leq p < \infty$. Others, such as the one we shall consider here, are deeper and have to do with the possibility of pointwise convergence in the nontangential sense; this involves a corresponding maximal function

$$F^*(x) = \sup_{|x-y|<t} |F(y,t)|. \qquad (15)$$

For the systematic study of such functions F, it will be important to free ourselves from the restriction that F be given as an average of the form (14), and we will therefore turn our attention to a general class of functions in $\mathbf{R}^{n+1}_+$ for which we have nontangential control. However, before proceeding in this more general setting, it will be helpful to review briefly some of the properties of functions F that do arise in the form (14).

2.1 Nontangential maximal functions and averages. To begin with, we recall the relation of the averages (14) with the maximal function Mf given by

$$(Mf)(x) = \sup_{r>0} c_n r^{-n} \int_{|y|<r} |f(x-y)| \, dy.$$

Whenever Φ is a nonnegative function on $\mathbf{R}^n$ that is radial and (radially) decreasing, then

$$\sup_{t>0} |f * \Phi_t(x)| \leq Mf(x) \cdot \int_{\mathbf{R}^n} \Phi \, dy. \qquad (16)$$

To see this, it suffices to verify that the inequality

$$|f * \Phi(x)| \leq Mf(x) \qquad (17)$$

holds for functions Φ of the above type that are normalized by the condition that $\int \Phi = 1$.

First take Φ to be of the form $\sum_{j=1}^{N} a_j \chi_{B_j}$, where each a_j is a positive constant and each χ_{B_j} is the characteristic function of a ball B_j that is centered at the origin. Then, since $\sum a_j |B_j| = 1$ and $(f * \chi_{B_j})(x) \leq |B_j| Mf(x)$, the inequality (17) follows immediately. In general, any nonnegative, integrable, radial, and radially decreasing Φ can be approximated by such finite sums; so the inequalities (17) and (16) hold as claimed.

The implication of the above for nontangential control is then a simple consequence.

PROPOSITION. *Assume that Φ has a radial majorant that is nonincreasing, bounded, and integrable. Then, with $F(x,t) = f * \Phi_t(x)$, we have*

$$F^*(x) = \sup_{|y-x|<t} |F(y,t)| \leq cMf(x).$$

Proof. If the radial nonincreasing majorant of $\Phi(x)$ is $\tilde{\Phi}(|x|)$, then $\sup_{|u|\leq 1} |\Phi(x-u)|$ has a radial majorant $\tilde{\Psi}(|x|)$, where $\tilde{\Psi}(r) = \tilde{\Phi}(0)$ for $0 \leq r \leq 1$, and $\tilde{\Psi}(r) = \tilde{\Phi}(r-1)$ for $r > 1$. The integrability of $\tilde{\Psi}$ then follows from that of $\tilde{\Phi}$ and the finiteness of $\tilde{\Phi}(0)$. We can therefore apply (16) to obtain our conclusion.

2.2 The main theorem. As mentioned at the beginning of §2, we now free ourselves from the restriction that our functions F, defined on $\mathbf{R}^{n+1}_+$, arise in a particular way from a given f on $\mathbf{R}^n$. For such a general F, we define as above its nontangential maximal function F^* by using cones of aperture 1; that is,

$$F^*(x) = \sup_{(y,t)\in\Gamma(x)} |F(y,t)|, \quad \text{where } \Gamma(x) = \{(y,t) : |y-x| < t\}. \tag{18}$$

We let $\mathcal{N}$ be the linear space of all (everywhere-defined) Borel measurable functions F on $\mathbf{R}^{n+1}_+$ having the property that $F^* \in L^1(\mathbf{R}^n)$. With the norm $\|F\|_{\mathcal{N}} = \|F^*\|_{L^1(\mathbf{R}^n)}$, $\mathcal{N}$ becomes a Banach space.

Several remarks may be helpful at this point:

(i) No matter what F is, $\{x \in \mathbf{R}^n : F^*(x) > \alpha\}$ is always open. In particular, F^* is always measurable.

(ii) If $F^*(x) = 0$ for almost every x (or, more generally, for a dense set of x), then F vanishes identically on $\mathbf{R}^{n+1}_+$.

(iii) Instead of the cones with unit aperture used in the definition of F^*, we could have used any fixed aperture a, and defined

$$F^*_a(x) = \sup\{|F(y,t)| : |y-x| < at\}.$$

The new norm thus obtained, namely $\|F^*_a\|_{L^1(\mathbf{R}^n)}$, turns out to be equivalent to the one above (the case $a = 1$); see §2.5 below.

The study of the space $\mathcal{N}$ is intimately connected with the study of another space $\mathcal{C}$, whose elements are measures on $\mathbf{R}^{n+1}_+$. Roughly speaking, these two spaces are in duality; the "tents" that occur in the definition of $\mathcal{C}$ may be thought of as dual to the cones that occur in the definition of $\mathcal{N}$. To make these ideas precise we need the following notation: If $B = B(x_0, r)$ is any open ball in $\mathbf{R}^n$, then its "tent" $T(B)$ is the closed set in $\mathbf{R}^{n+1}_+$ given by

$$T(B) = \{(x,t) : |x - x_0| \le r - t\}.$$

More generally, if O is any open set in $\mathbf{R}^n$, let F be its complement and define $T(O) = {}^c\mathcal{R}(F)$, where $\mathcal{R}(F)$ is the union of the cones based in F, i.e., $\mathcal{R}(F) = \bigcup_{x\in F} \Gamma(x)$. One should observe that the tent $T(O)$ is the set in $\mathbf{R}^{n+1}_+$ lying on or below the graph $\{(x,t) : t = \operatorname{dist}(x, F)\}$; put another way,

$$T(O) = \bigcup_{x\in O} T\big(B(x, \operatorname{dist}(x, {}^cO))\big).$$

We remark that tents are quite easily visualized when $n = 1$; see Figure 1.

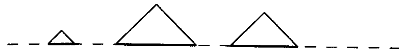

Figure 2. A tent over a simple open set in $\mathbf{R}^1$.

Next, given a Borel measure $d\mu$ on $\mathbf{R}^{n+1}_+$, we define the function $C(d\mu)$ by

$$C(d\mu)(x) = \sup_{x\in B} \frac{1}{|B|}\int_{T(B)} |d\mu|, \tag{19}$$

where the supremum is over all balls containing x. We then define $\mathcal{C}$ to be the space of measures $d\mu$ for which $C(d\mu)$ is a bounded function and set $\|d\mu\|_{\mathcal{C}} = \sup_{x\in\mathbf{R}^n} |C(d\mu)(x)|$. Each such $d\mu$ is called a *Carleson measure* and $\|d\mu\|_{\mathcal{C}}$ is the *Carleson norm* of $d\mu$.

The duality alluded to above is essentially contained in the following theorem.

THEOREM 2. *If $F \in \mathcal{N}$ and $d\mu \in \mathcal{C}$, then*

$$\left|\int_{\mathbf{R}^{n+1}_+} F(x,t)\, d\mu(x,t)\right| \le c\|F\|_{\mathcal{N}} \cdot \|d\mu\|_{\mathcal{C}}. \tag{a}$$

A more precise version of the above is the inequality

$$\left|\int_{\mathbf{R}^{n+1}_+} F(x,t)\, d\mu(x,t)\right| \le c\int_{\mathbf{R}^n} F^*(x)\, C(d\mu)(x)\, dx. \tag{b}$$

Before we come to the proof, let us remark that while, heuristically speaking, part (a) is the asserted duality, the spaces $\mathcal{N}$ and $\mathcal{C}$ are not, strictly speaking, duals of each other. An exact version of the duality is given in §2.5 below.

2.3 The proof of the theorem is based on two simple but key observations. The first is that

$$\{(x,t) : |F(x,t)| > \alpha\} \subset T(O), \quad \text{where } O = \{x : F^*(x) > \alpha\}. \tag{20}$$

The second observation is as follows: If we assume that $d\mu$ is positive and $\|d\mu\|_{\mathcal{C}} \le 1$—i.e., that $\mu(T(B)) \le |B|$ for all balls B—then this implies that

$$\mu(T(O)) \le c|O|, \quad \text{for any open set } O \subset \mathbf{R}^n. \tag{21}$$

Assuming (20) and (21) for the moment, we can immediately prove part (a) of the theorem. We may assume that $F \geq 0$, $d\mu \geq 0$. Let $O = \{x \in \mathbf{R}^n : F^*(x) > \alpha\}$; then

$$\mu\{(x,t) \in \mathbf{R}^{n+1}_+ : F(x,t) > \alpha\} \leq \mu(T(O)) \leq c\|d\mu\|_C \cdot |O|.$$

Integrating both sides with respect to α yields

$$\int_{\mathbf{R}^{n+1}_+} F(x,t)\, d\mu(x,t) \leq c\|d\mu\|_C \int_{\mathbf{R}^n} F^*(x)\, dx,$$

as required.

2.3.1 We next verify (20). If (x,t) is such that $|F(x,t)| > \alpha$, then for any $y \in B(x,t)$ one has $F^*(y) > \alpha$. Thus $B = B(x,t) \subset O$, and clearly $(x,t) \in T(B) \subset T(O)$:

2.3.2 The second observation (21) is obvious in the one-dimensional case. Indeed, writing O as the disjoint union $\bigcup I_k$ of its maximal open subintervals I_k, we have that $T(O) = \bigcup T(I_k)$. This simple decomposition does not work in the higher-dimensional case, and to prove (21) here we need to invoke the Whitney decomposition.‡ This allows us to express O as a disjoint union of cubes Q_k with $\operatorname{diam}(Q_k) \geq c_1 \cdot \operatorname{dist}(Q_k, {}^cO)$. It follows that if $x \in Q_k$, $\operatorname{diam}(Q_k) \geq c_2 \cdot \operatorname{dist}(x, {}^cO)$.

Now for each cube Q_k, we let B_k denote the ball with the same center as Q_k, with $\operatorname{diam}(B_k) = c_3 \cdot \operatorname{diam}(Q_k)$; c_3 will be chosen to be much larger than $1/c_2$. We claim that

$$T(O) \subset \bigcup_k T(B_k). \tag{22}$$

To see this pick $(x,t) \in T(O)$; that is, $x \in O$ and $t \leq \operatorname{dist}(x, {}^cO)$. Let Q_k be the cube that contains x. It suffices to show that $(x,t) \in T(B_k)$; this occurs exactly when $B(x,t) \subset B_k$. Now

$$\operatorname{dist}(x, {}^cB_k) \geq \bar{c} \cdot \operatorname{diam}(Q_k),$$

since $x \in Q_k$, and $\bar{c}$ can be made large with c_3. Finally,

$$\operatorname{diam}(Q_k) \geq c_2 \cdot \operatorname{dist}(x, {}^cO) \geq c_2,$$

which (upon taking $\bar{c}$ large) gives $\operatorname{dist}(x, {}^cB_k) \geq t$; that is $(x,t) \in T(B_k)$, proving (22).

Now by (22),

$$\mu(T(O)) \leq \sum_k \mu(T(B_k)) \leq \tilde{c} \sum_k |B_k| = \tilde{c}\, c_3^n \sum_k |Q_k| = c|O|,$$

which establishes the observation (21).

‡ One can use either the standard cubes in $\mathbf{R}^n$ (see the lemma on p. 9 of *Singular Integrals*) or the "cubes" that arise in remark (i) following Lemma 2 in §3.2 of Chapter 1.

2.3.3 To prove part (b) of the theorem, we merely retrace our steps in the proof of part (a), and note that, by the maximal definition (19) of the functional C, one has (if $d\mu \geq 0$)

$$\mu(T(B_k)) = \int_{T(B_k)} d\mu \leq C(d\mu)(x)|B_k|, \quad \text{if } x \in Q_k.$$

Thus, $\mu(T(B_k)) \leq c \int_{Q_k} C(d\mu)(x)\, dx$. By (21) and (22) we have

$$\begin{aligned} \mu\{|F(x,t)| > \alpha\} &\leq \sum_k \mu(T(B_k)) \\ &\leq \sum_k c \int_{Q_k} C(d\mu)(x)\, dx = c \int_{\{F^*(x) > \alpha\}} C(d\mu)(x)\, dx. \end{aligned}$$

Integrating this inequality in α produces the second conclusion of the theorem.

2.4 The theorem implies the following corollary.

Corollary. *Assume that $d\mu$ is a fixed positive measure in $\mathcal{C}$. Let F on $\mathbf{R}^{n+1}_+$ be such that $F^* \in L^p(\mathbf{R}^n)$ for some p, $0 < p < \infty$. Then*

$$\int_{\mathbf{R}^{n+1}_+} |F(x,t)|^p\, d\mu(x,t) \leq c \int_{\mathbf{R}^n} [F^*(x)]^p\, dx. \tag{23}$$

In particular, if $F(x,t) = f * \Phi_t$, where Φ has a radial, nonincreasing, bounded, and integrable majorant, then

$$\int_{\mathbf{R}^{n+1}_+} |F(x,t)|^p\, d\mu(x,t) \leq c_p \int_{\mathbf{R}^n} |f(x)|^p\, dx, \tag{24}$$

for $1 < p < \infty$.

In fact, (23) follows from part (a) of the theorem by replacing $|F|$ by $|F|^p$. Next, (24) is a consequence if we use the proposition in §2.1 above, together with the L^p inequality for the maximal operator M given in §3 of Chapter 1.

Notice, incidentally, that (when $d\mu$ is positive) the condition $d\mu \in \mathcal{C}$ is necessary for (24) to hold for some p. Indeed, if Φ is positive and $f = \chi_B$, then $F(x,t) \geq c > 0$ on $T(B)$.

2.5 Remarks about the space $\mathcal{N}$. We now clarify two points concerning the space $\mathcal{N}$, elaborating some earlier comments.

2.5.1 First, we make precise the sense in which matters do not really depend on the aperture of the cones used to define $\mathcal{N}$. For any $a > 0$, we have the maximal function associated with cones of aperture a, given by $F_a^*(x) = \sup_{|y-x|<at} |F(y,t)|$. We claim that, if $a \geq b$,

$$|\{x : F_a^*(x) > \alpha\}| \leq c_{a,b}|\{x : F_b^*(x) > \alpha\}|, \quad \text{for all } \alpha > 0. \tag{25}$$

Here $c_{a,b} = c_n\left(\frac{a+b}{b}\right)^n$ where c_n depends only on the dimension n. Integrating (25) with respect to α gives

$$\int_{\mathbf{R}^n} F_a^*(x)\,dx \leq c_{a,b} \int_{\mathbf{R}^n} F_b^*(x)\,dx.$$

One can establish an assertion of the type (25) by a point-of-density argument, and here we need a quantitative version that can be formulated as follows. Suppose A is a closed set and γ is a fixed parameter, $0 < \gamma < 1$. We say that a point $x \in \mathbf{R}^n$ has *global γ-density* with respect to A, if $|A \cap B|/|B| \geq \gamma$, for *all* balls B centered at x. Let A^* be the points of global γ-density of A; then A^* is closed, $A^* \subset A$, and, most importantly,

$$|{}^cA^*| \leq c_\gamma |{}^cA|.$$

In fact, if $O = {}^cA$ and $O^* = {}^cA^*$, then

$$O^* = \{x : M(\chi_O)(x) > 1 - \gamma\},$$

and the assertion follows from the maximal theorem (Theorem 1(b) in Chapter 1), with $c_\gamma = c_2/(1-\gamma)$.

Returning to (25), we let $O = \{x : F_b^*(x) > \alpha\}$, and claim that $\{x : F_a^*(x) > \alpha\}$ is contained in the complement of the set of points of global γ-density of $A = {}^cO$, provided we take γ sufficiently close to 1. Indeed, suppose $F_a^*(x) > \alpha$ for a given x. Then there exists $(\bar{x}, t)$ with $F(\bar{x}, t) > \alpha$, $|x - \bar{x}| < at$. Now $B(\bar{x}, bt) \subset O$, therefore $O \cap B \supset B(\bar{x}, bt)$, and

$$\frac{|O \cap B|}{|B|} \geq \left(\frac{b}{a+b}\right)^n,$$

where B is the ball of radius $a + b$ centered at x. Thus

$$\frac{|A \cap B|}{|B|} \leq 1 - \left(\frac{b}{a+b}\right)^n,$$

and $x \notin A^*$ if $\gamma > 1 - [b/(a+b)]^n$. Therefore $\{x : F_a^*(x) > \alpha\} \subset {}^cA^*$ and (25) is proved.[†]

[†] The fact that the space $\mathcal{N}$ does not depend on the size of the aperture can also be established using Theorem 3 below.

2.5.2 Next, we shall give a more precise formulation of the duality between nontangential control and Carleson measures (whose essence is contained in in Theorem 2). To this end, we define first the space $\mathcal{N}_0$ to be the closure in $\mathcal{N}$ of the continuous functions with compact support in $\bar{\mathbf{R}}_+^{n+1}$. We note that $\mathcal{N}_0$ consists exactly of the $F \in \mathcal{N}$ that are continuous in $\mathbf{R}_+^{n+1}$ and have nontangential limits a.e. in $\mathbf{R}^n$ (for this equivalence, see §5.6 below).

Together with $\mathcal{N}_0$, we consider a corresponding variant of $\mathcal{C}$, namely the space $\overline{\mathcal{C}}$ of Carleson measures in the extended sense, i.e., the space of all Borel measures $d\mu$ on the closure $\bar{\mathbf{R}}_+^{n+1}$ of $\mathbf{R}_+^{n+1}$, satisfying

$$\sup_B |B|^{-1} \int_{\bar{T}(B)} |d\mu| = \|d\mu\| < \infty.$$

The duality can be expressed as follows. Fix a measure $d\mu \in \overline{\mathcal{C}}$. For every function F that is continuous and has compact support in $\bar{\mathbf{R}}_+^{n+1}$, consider the functional

$$F \mapsto \int_{\bar{\mathbf{R}}_+^{n+1}} F(x,t)\,d\mu(x,t). \tag{26}$$

Then (26) extends to a bounded linear functional on $\mathcal{N}_0$ and, conversely, every bounded linear functional on $\mathcal{N}_0$ arises in this way; in other words, $\overline{\mathcal{C}}$ is identified with the dual of $\mathcal{N}_0$ via (26).

Indeed, (26) extends to all of $\mathcal{N}_0$ by the inequality (a) of Theorem 2 applied to the definition of $\mathcal{N}_0$, which also shows its boundedness. Conversely, given a bounded linear functional ℓ on $\mathcal{N}_0$, restrict it to the continuous functions with compact support in $\mathcal{N}_0$. By the Riesz representation theorem, there exists a Borel measure $d\mu$ so that ℓ is given by (26); also we have

$$\left|\int_{\bar{\mathbf{R}}_+^{n+1}} F(x,t)\,d\mu(x,t)\right| \leq c \int_{\mathbf{R}^n} F^*(x)\,dx$$

for all such F. Taking the supremum over F with $|F| \leq 1$ on $\bar{T}(B)$, we get $\int_{\bar{T}(B)} |d\mu(x,t)| \leq c|B|$ for all balls B, showing that $d\mu \in \overline{\mathcal{C}}$.

2.6 Atomic decomposition. The last general fact we want to explore about the space $\mathcal{N}$ is the possibility of expressing elements of $\mathcal{N}$ using an "atomic decomposition". This will be the simplest example of analogous decompositions, variants of which will be important in other settings below (see Chapter 3, §2).

Suppose $B \subset \mathbf{R}^n$ is a ball. An *atom* associated to B is a measurable function a supported in the tent $T(B) \subset \mathbf{R}_+^{n+1}$ with $\|a\|_\infty \leq |B|^{-1}$. Note that $a^*(x) \leq |B|^{-1}$ when $x \in B$ and that $a^*(x) = 0$ when $x \notin B$. Thus

$a \in \mathcal{N}$ and $\|a\|_{\mathcal{N}} = \|a^*\|_1 \le 1$. Atoms will be the basic building blocks for functions in $\mathcal{N}$.

Observe first that if a_k are atoms, $k = 1, 2, \ldots$, and λ_k are positive constants with $\sum \lambda_k < \infty$, then $\sum \lambda_k a_k \in \mathcal{N}$ and $\|\sum \lambda_k a_k\|_{\mathcal{N}} \le \sum \lambda_k$. The converse also holds.

THEOREM 3. *Any $F \in \mathcal{N}$ can be written*

$$F = \sum \lambda_k a_k, \tag{27}$$

where the a_k are atoms, $\lambda_k \ge 0$, and

$$\sum \lambda_k \le c\|F\|_{\mathcal{N}}, \tag{28}$$

so that the sum (27) *converges in $\mathcal{N}$.*

Before proceeding to the proof we make two remarks about the atomic decomposition. We first point out its utility: One can often prove statements about $\mathcal{N}$ by verifying them in the special case of atoms. Examples of this appear in the next section. Second, the decomposition is equivalent (heuristically, at least) with the duality given by part (a) of Theorem 2 (see §5.7 below).

Turning to the proof, we define $O^j = \{x : F^*(x) > 2^j\}$; thus O^j is an open set. If $T(O^j)$ is the tent over O^j, then clearly $\cdots \supset O^j \supset O^{j+1} \supset \cdots$, and $\cdots \supset T(O^j) \supset T(O^{j+1}) \supset \cdots$; also $\bigcup_{j \in \mathbf{Z}} T(O^j)$ contains the support of F in $\mathbf{R}^{n+1}_+$ (see Figure 3).

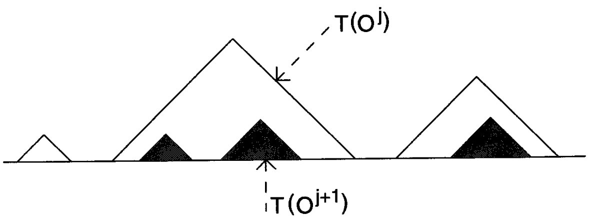

Figure 3. Nested tents.

As we have seen in the proof of Theorem 2 (see §2.3.1 and (22)), O^j can be decomposed as a disjoint union of cubes Q^j_k, with corresponding balls $B^j_k \supset Q^j_k$, so that‡

$$T(O^j) \subset \bigcup_k \big(T(B^j_k) \cap [Q^j_k \times (0, \infty)]\big). \tag{29}$$

‡ Recall that, if $(x,t) \in T(O^j)$, then for some k, $x \in Q^j_k$ and $(x,t) \in T(B^j_k)$.

Next, set

$$\Delta^j_k = T(B^j_k) \cap [Q^j_k \times (0,\infty)] \cap [T(O^j) \backslash T(O^{j+1})];$$

it is obvious from (29) that supp$F \subset \bigcup_{j,k} \Delta^j_k$ and that the Δ^j_k are mutually disjoint. Let $F^j_k = F \cdot \chi_{\Delta^j_k}$; clearly $F = \sum_{j,k} F^j_k$.

Now set $F^j_k = \lambda^j_k a^j_k$, where $a^j_k = 2^{-j-1}|B^j_k|^{-1} \cdot F^j_k$, and $\lambda^j_k = 2^{j+1}|B^j_k|$. Each a^j_k is an atom supported on B^j_k because $|F^j_k| \le 2^{j+1}$ on $\Delta^j_k \subset {}^cT(O^{j+1})$.

Finally, it is easy to see that $F = \sum_{j,k} \lambda^j_k a^j_k$ is an atomic decomposition (that is, (27) and (28) are satisfied), because

$$\begin{aligned}\sum_{j,k} \lambda^j_k = \sum_{j,k} 2^{j+1}|B^j_k| &\le c \sum_{j,k} 2^{j+1}|Q^j_k| \\ &= c\sum_j 2^{j+1}|O^j| \le c\|F^*\|_{L^1} = c\|F\|_{\mathcal{N}},\end{aligned}$$

and the proof of Theorem 3 is complete.

3. *Two applications*

In the previous section we explored the properties of functions for which we had suitable nontangential control. We shall be rewarded here by the fact, paradoxical as it seems, that this study leads us to surprising conclusions about properties of approaches that are either wider or narrower than the nontangential one.

In the first subsection, we return to one of the basic problems of real-variable theory, namely the behavior of averages of functions. The particular problem we shall deal with here concerns the averages

$$\frac{1}{|B|} \int_B f(x-y)\, dy,$$

as B ranges over a suitable collection of balls.

3.1 Differentiation with respect to a general family of balls. Let $\mathcal{B} = \{B\}$ be any fixed collection of balls in $\mathbf{R}^n$. We want to know whether, for suitable f,

$$f(x) = \lim_{\mathrm{diam}(B) \to 0} \frac{1}{|B|} \int_B f(x-y)\, dy \quad \text{a.e.}, \tag{30}$$

where B ranges over the family $\mathcal{B}$. In particular, we are interested in the properties of the corresponding maximal function

$$M_{\mathcal{B}}f(x) = \sup_{B \in \mathcal{B}} \frac{1}{|B|} \int_B |f(x-y)|\,dy. \tag{31}$$

The question as to whether (30) holds for all f in an L^p space is closely connected with the behavior of the maximal function (31) on this space (see Chapter 1, §3.1 and Chapter 10, §2). The theory of the usual maximal operator M corresponds to the case where $\mathcal{B}$ is the collection of all balls centered at the origin, or a slight generalization, in which the centers of the balls are allowed to be displaced from the origin by a distance comparable to their diameters. We now concern ourselves with the study of general $\mathcal{B}$, whose members may be very far from the origin compared to their diameters.

It turns out that the behavior of $M_{\mathcal{B}}$ is closely related to that of $M_{\bar{\mathcal{B}}}$; here $\bar{\mathcal{B}}$ is the family of balls each of which contains some member of $\mathcal{B}$, that is, $\bar{\mathcal{B}} = \{B : B \supset B_0 \text{ for some } B_0 \in \mathcal{B}\}$. $\bar{\mathcal{B}}$ is called the *completion* of $\mathcal{B}$. We also consider the set

$$\mathcal{B}(r) = \bigcup \{B \in \bar{\mathcal{B}} : \text{ radius}(B) = r\}. \tag{32}$$

The main result is then as follows:

THEOREM 4. (a) *If there is a* $1 < p_0 < \infty$ *so that* $\|M_{\mathcal{B}}f\|_{p_0} \leq A\|f\|_{p_0}$, *for all* $f \in L^{p_0}(\mathbf{R}^n)$, *then*

$$|\mathcal{B}(r)| \leq cr^n, \quad \text{for } 0 < r < \infty \text{ and some fixed } c > 0. \tag{33}$$

(b) *Conversely, suppose* (33) *holds. Then* $M_{\mathcal{B}}$ *is of weak-type* $(1,1)$ *and is bounded on all* $L^p(\mathbf{R}^n)$, $1 < p \leq \infty$.

Observe that (33) is equivalent to the following: There exists a fixed N so that all the balls in $\mathcal{B}$ of at most a given radius r lie in a union of N balls of radius r. Before proceeding with the proof, we give two quick examples.

(i) In $\mathbf{R}^1$, let $I(s)$ denote the interval of length s, situated in the positive half-line, given by $(h(s), s + h(s))$. Let $h(0) = 0$, and assume that h is continuous and increasing with s. If we set $\mathcal{B} = \{I(s)\}_{s>0}$, then it is easily seen that $|\mathcal{B}(r)| = 2r + h(r)$. This gives a positive result in Theorem 4 exactly when $h(s) \leq cs$.

(ii) Surprisingly, if we replace the continuous family above by a suitable subfamily, we can obtain denumerable collections of intervals, with centers far from the origin, for which a positive result holds. We need only choose a sequence $\{s_n\}$ tending to 0 rapidly enough so that (among other things) $h(s_{n+k}) \leq s_n$, for some fixed k ($s_n = 2^{-n^2}$, $h_n = 2^{-n^2+n}$ is a good example), and take $\mathcal{B} = \{I(s_n)\}_n$. A further example is in §5.10 below.

After consideration of these examples, the following amplification of our observation may become clear. We first remark that (33) is equivalent to the existence of a fixed N so that $\mathcal{B}(r)$ has at most N connected components, each of diameter essentially r. Take $r < r'$. Then $\mathcal{B}(r) \subset \mathcal{B}(r')$ and each component of $\mathcal{B}(r)$ is contained in some component of $\mathcal{B}(r')$. Now if each of our components shrinks to a point as $r \to 0$ (as in (i)), then statement (b) is an easy consequence of the usual maximal theorem. Consequently, part (b) of the theorem gives us something new only when (infinitely often) components "vanish" and others "bifurcate" as the radius varies (as in (ii)).

3.1.1 We turn to the proof of Theorem 4, and dispose first of the necessity of condition (33). Here the argument is elementary and we begin by showing that if the inequality $\|M_{\mathcal{B}}f\|_{p_0} \leq A\|f\|_{p_0}$ holds for some p_0, $1 < p_0 < \infty$, then a similar inequality holds when $\mathcal{B}$ is replaced by $\bar{\mathcal{B}}$.

To see this, let $m_B = |B|^{-1}\chi_B$, where χ_B is the characteristic function of the ball B. Then if $\bar{B}$ is another ball so that $\bar{B} \supset B$, and B_r denotes the ball of radius r centered at the origin, we have

$$m_{B_{2r}} * m_B \geq 2^{-n} m_{\bar{B}}, \quad \text{where } r = \text{radius}(\bar{B}). \tag{34}$$

Indeed, any ball of radius $2r$ about some point $x \in \bar{B}$ contains B and $\int m_B = 1$; so

$$m_{B_{2r}} * m_B(x) = |B_{2r}|^{-1} = 2^{-n}|\bar{B}|^{-1} = 2^{-n} m_{\bar{B}}(x), \quad \text{for } x \in \bar{B}.$$

Outside $\bar{B}$, $m_{\bar{B}}$ vanishes, so (34) holds.

Now convolve both sides of (34) with $|f|$, letting B range over $\mathcal{B}$, $\bar{B}$ range over $\bar{\mathcal{B}}$. The result is

$$M_{\bar{\mathcal{B}}}(f) \leq 2^n M(M_{\mathcal{B}}(f)),$$

where M is the standard maximal operator, and hence by Theorem 1 in Chapter 1, we get

$$\|M_{\bar{\mathcal{B}}}f\|_{p_0} \leq A\|M_{\mathcal{B}}f\|_{p_0} \leq A\|f\|_{p_0}, \tag{35}$$

as claimed.

Next, we test the conclusion just obtained against $f = \chi_{B_{2r}}$, the characteristic function of the ball of radius $2r$ about the origin. We assert, for this f, that

$$\mathcal{B}(r) \subset \{x : M_{\bar{\mathcal{B}}}f(x) \geq 1\}. \tag{36}$$

To prove (36) note that if $x \in \mathcal{B}(r)$, there is a ball $B(x_0, r) \in \bar{\mathcal{B}}$ with $x \in B(x_0, r)$. Since $B(x_0, r) \in \bar{\mathcal{B}}$, we have

$$M_{\bar{\mathcal{B}}}f(x) \geq |B(x_0,r)|^{-1} \int_{B(x_0,r)} \chi_{B_{2r}}(x-y)\,dy = 1,$$

because $|x-y| \leq 2r$, if $x, y \in B(x_0, r)$. With (36) proved, then, by (35),

$$|\mathcal{B}(r)| \leq \|M_{\bar{\mathcal{B}}}f\|_{p_0}^{p_0} \leq A'\|f\|_{p_0}^{p_0} = A'|B_{2r}| = cr^n,$$

and the necessity of condition (33) is established.

3.1.2 To prove the sufficiency of the condition (33), we begin by associating to the collection of balls $\mathcal{B}$ a corresponding set $\mathcal{B}' \subset \mathbf{R}^{n+1}_+$. By definition, $(x,t) \in \mathcal{B}'$ if $B(x,t) \in \mathcal{B}$; in this way the point $(x,t) \in \mathbf{R}^{n+1}_+$ represents the ball in $\mathbf{R}^n$ of center x and radius t. With this definition, note that if $(y,t) \in \mathcal{B}'$, then $(y+y', t+t') \in \bar{\mathcal{B}}'$ whenever $|y'| < |t'|$; the result is that the cone of aperture 1, with vertex at any point of $\mathcal{B}'$, lies in $\bar{\mathcal{B}}'$.

We shall now assume, as we may, that $\mathcal{B} = \bar{\mathcal{B}}$ in the proof of the sufficiency of (33). For each F defined on $\mathbf{R}^{n+1}_+$, we have its nontangential maximal function $F^*(x) = \sup_{|y|<t} |F(x-y,t)|$. We also consider the maximal function corresponding to the wider approach region $\mathcal{B}'$, namely $F^*_{\mathcal{B}}(x) = \sup_{(y,t)\in\mathcal{B}'} |F(x-y,t)|$.

Our claim is that in this general setting, under condition (33), one has

$$\int_{\mathbf{R}^n} F^*_{\mathcal{B}}(x)\,dx \le c \int_{\mathbf{R}^n} F^*(x)\,dx. \tag{37}$$

In fact, if we assume that the right side of (37) is finite, then $F \in \mathcal{N}$, and we may apply the atomic decomposition given by Theorem 3. So it suffices to verify (37) when F is an atom. After translation, we may assume that F is supported in $T(B)$, where B is the ball of radius r centered at the origin.

Now $F^*_{\mathcal{B}} \le |B|^{-1}$. Next, if $F^*_{\mathcal{B}}(x) \ne 0$, then there is a $(y,t) \in \mathcal{B}'$ with $(x-y,t) \in T(B)$; that is, $|x-y| < r-t$. If we set $y' = x-y$, $t' = r-t$, we have that

$$(y+y', t+t') = (x,r) \in \bar{\mathcal{B}}' = \mathcal{B}'$$

or, in other words, that $x \in \mathcal{B}(r)$ (see Figure 4). Thus we have shown that $F^*_{\mathcal{B}}$ is supported in $\mathcal{B}(r)$.

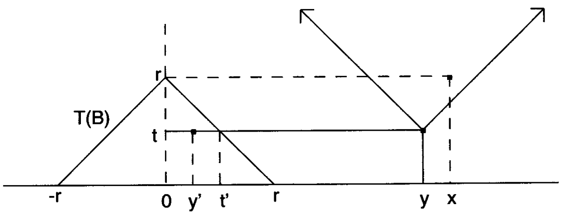

Figure 4. The proof of (37).

Therefore, $\int_{\mathbf{R}^n} F^*_{\mathcal{B}}(x)\,dx \le |B|^{-1}\int_{\mathcal{B}(r)} dx \le c$, and (37) holds for atoms, and so it also holds generally.

Next, (37) immediately leads to a generalization of itself:

$$\int_{\mathbf{R}^n} (F^*_{\mathcal{B}}(x))^p\,dx \le c \int_{\mathbf{R}^n} (F^*(x))^p\,dx, \quad \text{whenever } 0 < p < \infty, \tag{38}$$

which is obtained by replacing F by $|F|^p$ in (37).

From (37) it is also easy to prove the distribution function inequality:

$$|\{x : F^*_{\mathcal{B}}(x) > \alpha\}| \le c|\{x : F^*(x) > \alpha\}|, \quad \text{for all } \alpha > 0. \tag{39}$$

In fact, in (37) we need only replace F by the characteristic function of the set where $|F| > \alpha$ and (39) is obtained.

We now return to the maximal operator $M_{\mathcal{B}}$. For any $f \in L^p(\mathbf{R}^n)$, we define

$$F(x,t) = |B(x,t)|^{-1} \int_{B(x,t)} |f|.$$

Now for any (y,t), we have

$$F(x-y,t) = |B(y,t)|^{-1} \int_{B(y,t)} |f(x-u)|\,du,$$

thus

$$F^*_{\mathcal{B}}(x) = \sup_{(y,t)\in\mathcal{B}'} |F(x-y,t)| = \sup_{B\in\mathcal{B}} |B|^{-1} \int_B |f(x-u)|\,du = M_{\mathcal{B}}f(x).$$

Also note that $F^*(x) \le 2^n Mf(x)$, so the results for the standard maximal operator, together with (38) and (39), show that

$$\begin{aligned} &|\{x : M_{\mathcal{B}}f(x) > \alpha\}| \le \frac{c_1}{\alpha}\int_{\mathbf{R}^n} |f|, \quad \text{for all } \alpha > 0, \text{ and} \\ &\int_{\mathbf{R}^n} (M_{\mathcal{B}}f)^p \le c_p \int_{\mathbf{R}^n} |f|^p, \quad \text{whenever } 1 < p < \infty, \end{aligned} \tag{40}$$

which proves the theorem.

3.2 Multipliers for the normal approach. We keep to the above setting and recall that $Mf(x) = \sup_{0<t<\infty} F(x,t)$ when $F(x,t) = |B(x,t)|^{-1}\int_{B(x,t)} |f|$. A question that occurs naturally is to find the unbounded positive functions $\mu(x,t)$ on $\mathbf{R}^{n+1}_+$ (if any exist), for which we can also control $\sup_{0<t<\infty} \mu(x,t)\,F(x,t)$, or more precisely, those for which

$$\int_{\mathbf{R}^n} \sup_{0<t<\infty} [\mu(x,t)\,F(x,t)]^p\,dx \le A \int_{\mathbf{R}^n} |f(x)|^p\,dx. \tag{41}$$

If we test (41) with $f = \chi_B$, where B is an arbitrary ball, and use the fact that $F(x,t) \geq c^{-1}$ when $x \in B$, and $t \leq \text{radius}(B)$, we get the following necessary condition for the multiplier μ:

$$\int_B \sup_{0<t<r} \mu(x,t)^p \, dx \leq c|B|, \tag{42}$$

where r is the radius of B. The condition is also sufficient.

PROPOSITION. *Suppose μ is continuous and positive in $\mathbf{R}^{n+1}_+$. If (42) holds for some p, $1 < p < \infty$, then (41) holds for the same p.*

A natural class of multipliers μ of the above kind is given in §5.13 below. Let us also note that the normal approach has here an advantage over the nontangential approach: For the latter there are no unbounded multipliers (as is easily verified), making the positive result for the normal approach a little surprising.

The proof of the proposition is an almost immediate consequence of the atomic decomposition given in §2.6. In fact, disregarding momentarily the technical question of measurability, a more general inequality holds, namely,

$$\int_{\mathbf{R}^n} \sup_{0<t<\infty} [\mu(x,t)\,|F(x,t)|]^p \, dx \leq c \int_{\mathbf{R}^n} F^*(x)^p \, dx, \tag{43}$$

whenever F is given on $\mathbf{R}^{n+1}_+$ and μ satisfies (42). The proof of (43) for general p reduces to that for $p = 1$, upon replacing $|F|^p$ by F.

Now, in the case $p = 1$, the finiteness of the right side of (43) means that $F \in \mathcal{N}$, and it suffices to check (43) for atoms when $p = 1$. Suppose then that F is an atom supported in $T(B)$, so $\operatorname{supp} F \subset B \times (0,r)$, where r is the radius of the given ball B. Since $|F| \leq |B|^{-1}$, the left side of (43) is majorized by

$$|B|^{-1} \int_B \sup_{0<t<r} \mu(x,t) \, dx \leq c,$$

verifying (43) for $p = 1$, and hence for all p. However in our case $F^*(x) \leq 2^n Mf(x)$, so combining the above with the L^p boundedness of the maximal operator gives (41) for all $p > 1$, save for the matter of measurability postponed until now.

The question alluded to above relates to the measurability of

$$F^+_\mu(x) = \sup_{t>0} \mu(x,t)|F(x,t)|.$$

The situation is as follows. Since we took F to be Borel-measurable, it turns out that F^+_μ is Lebesgue-measurable in x (although not necessarily Borel-measurable!). Further details are in §5.8 below.

4. Singular approximations of the identity

We return to the study of approximations of the identity begun in §2.1 above. Starting with a fixed integrable Φ with $\int \Phi = 1$, and taking $\Phi_t = t^{-n}\Phi(x/t)$, we again ask: In what sense does $f * \Phi_t \to f$ as $t \to 0$? This is connected with the behavior of the maximal operator M_Φ, defined by

$$M_\Phi f(x) = \sup_{0<t<\infty} |f * \Phi_t(x)|. \tag{44}$$

Several remarks may help clarify the issues we want to address.

(i) With only the assumption that Φ is integrable, we can conclude that $f * \Phi_t \to f$ in L^p, whenever $f \in L^p$, $1 \leq p < \infty$. However,

$$\lim_{t\to 0} f * \Phi_t(x)$$

may fail to exist for almost every x, and $M_\Phi f$ may by infinite almost everywhere when $f \in L^p$, for any $p < \infty$ (see §5.16 below). In fact, for the L^1 theory of M_Φ to hold, it is necessary that $\sup_{t>0} |\Phi_t(x)|$ be integrable on the unit sphere.†

(ii) When Φ is "regular", in that it decreases at a sufficient uniform rate at infinity, then by the majorization in §2.1, we have $M_\Phi \leq cM$, taking care of matters in that case.

(iii) However, there are interesting situations occurring in practice that cannot be covered by (ii), because the Φ in question has a "singular" (i.e., slow) decrease in certain directions. A simple example of this arises in $\mathbf{R}^2$, with $\Phi(x_1,x_2) = \pi^{-2}(1+x_1^2)^{-1}(1+x_2^2)^{-1}$, which corresponds to the "Poisson integral" for bi-harmonic functions on the product of two half-planes; here the directions along the coordinate axes are the most singular. This is a very special case of other naturally occurring examples, such as Poisson integrals for symmetric spaces. (For more details about these, see §5.17 and §5.18.)

In view of the above, we need a further condition on Φ besides integrability. What is suggested by these considerations, and put in its simplest form, is the assumption we shall make:

$$\text{For each } x,\ \Phi(rx) \text{ is decreasing in } r, \quad 0 < r < \infty. \tag{45}$$

PROPOSITION 1. *Suppose Φ is nonnegative, integrable, and satisfies (45). Then M_Φ is bounded from $L^p(\mathbf{R}^n)$ to itself, for $1 < p \leq \infty$.*

† This can be proved by using the maximal principle in §2 of Chapter 10 and, more particularly, the variant in §3.4 of that chapter.

It is not known whether the weak-type L^1 result holds in this generality. However, if we strengthen the integrability assumption on Φ somewhat (adding a Dini-type condition), then one can obtain the desired result in this case also. Recall that a positive function η defined on $[0,1]$ is a Dini modulus of continuity (see §6.5 in Chapter 1) if $\eta(0)=0$, η is nondecreasing, and $\int_0^1 \eta(s)s^{-1}\,ds<\infty$.

PROPOSITION 2. *Suppose that, in addition to* (45), Φ *satisfies*

$$\int_{\mathbf{R}^n} |\Phi(x-y)-\Phi(x)|\,dx \le \eta(|y|). \qquad (i)$$

Suppose also that Φ *has compact support or, more generally, that*

$$\int_{|x|\ge R} \Phi(x)\,dx \le \eta(R^{-1}), \quad \text{whenever } R>1, \qquad (ii)$$

for some Dini modulus η. *Then* M_Φ *is of weak-type* $(1,1)$.

An immediate corollary of these propositions will help elucidate their thrust. Let $\Omega(x)$ be a fixed positive function of x that depends only on the direction of x; i.e., $\Omega(rx)=\Omega(x)$, $r>0$. Define

$$(M_\Omega f)(x) = \sup_{r>0} \frac{1}{r^n}\int_{|y|\le r} |f(x-y)|\,\Omega(y)\,dy;$$

M_Ω is essentially the M_Φ defined above, when $\Phi(x)=\Omega(x)\cdot\chi_{\{|x|\le 1\}}$. The directions in which Ω is relatively large are the singular directions controlled by M_Ω.

COROLLARY. (a) *If* $\Omega\in L^1$, *then* M_Ω *is bounded from* L^p *to itself,* $1<p\le\infty$.

(b) *If* Ω *satisfies an* L^1 *Dini condition, then* M_Ω *is of weak type* $(1,1)$.

For a more refined result in this direction, see §5.19 below.

4.1 Here we shall prove Proposition 1 by the "method of rotations", which reduces matters to the one-dimensional theory.‡

For any unit vector $\xi\in\mathbf{R}^n$, $|\xi|=1$, we denote by $M^{(\xi)}f$ the standard maximal function in the direction ξ:

$$(M^{(\xi)}f)(x) = \sup_{r>0}\frac{1}{2r}\int_{-r}^{r} |f(x-t\xi)|\,dt. \qquad (46)$$

‡ For an introduction to this method, and for further motivation, see *Fourier Analysis*, Chapter 6, and *Singular Integrals*, Chapter 2, §6.5.

Notice that

$$\|M^{(\xi)}f\|_{L^p(\mathbf{R}^n)} \le A_p\|f\|_{L^p(\mathbf{R}^n)}, \quad \text{for } 1<p\le\infty, \qquad (47)$$

where A_p is the bound for the one-dimensional maximal function M. Indeed, (47) is rotation-invariant, so we may assume that ξ is in the direction of the x_1 axis. We apply the one-dimensional maximal inequality to $f(x)=f(x_1,x_2,\ldots,x_n)$ as a function of x_1, keeping $x_2,\ldots,x_n$ fixed. Raising the inequality to the p^{th} power and integrating with respect to $x_2,\ldots,x_n$ gives (47).

Next we claim, under the additional hypothesis that Φ is even, that

$$(M_\Phi f)(x) \le n\int_{|\xi|=1} (M^{(\xi)}f)(x)\int_0^\infty \Phi(r\xi)\,r^{n-1}\,dr\,d\sigma(\xi). \qquad (48)$$

To prove this inequality, it suffices to show that $(f*\Phi_t)(x)$ is majorized by the right side of (48) for each $t>0$. Hence, by a change of variables, it suffices to show it for $(f*\Phi)(x)$. Changing to polar coordinates gives

$$f*\Phi(x) = \frac{1}{2}\int_{|\xi|=1}\int_{-\infty}^{\infty} f(x-r\xi)\,\Phi(r\xi)\,|r|^{n-1}\,dr\,d\sigma(\xi).$$

For fixed ξ, the even function $r\mapsto \Phi(r\xi)|r|^{n-1}$ has $-\int_{|r|}^{\infty} s^{n-1}\,d_s\Phi(s\xi)=\Psi_\xi(r)$ as a decreasing even majorant. Then by the argument in §2.1 (see (17)), for fixed ξ

$$\int_{-\infty}^{\infty} f(x-r\xi)\,\Phi(r\xi)\,|r|^{n-1}\,dr \le M^{(\xi)}f(x)\cdot\int_{-\infty}^{\infty}\Psi_\xi(r)\,dr.$$

However,

$$\int_{-\infty}^{\infty}\Psi_\xi(r)\,dr = 2\int_0^\infty r^n\,d\Phi(r\xi) = 2n\int_0^\infty r^{n-1}\,\Phi(r\xi)\,dr,$$

and an integration in ξ then gives (48).

If Φ is not even, we replace it by $\Phi(x)+\Phi(-x)$, obtaining (48) with the constant doubled. In any case, if we insert (47) in (48) and use Minkowski's inequality for integrals, we obtain the boundedness of M_Φ on L^p, $p>1$, proving Proposition 1.

We note that this argument cannot be used to prove a weak-type L^1 inequality, because such inequalities are not subadditive.†

† See Chapter 5, §5.12 of *Fourier Analysis*.

4.2 We turn to the proof of Proposition 2. Taking $f \geq 0$ (as we may) and using the decreasing character of Φ given by (45), we need only consider a lacunary subsequence of t:

$$\sup_{0<t<\infty} f * \Phi_t(x) \leq 2^n \sup_j f * \Phi_{2^j}(x).$$

Next we claim that, under our assumptions,

$$\int_{|x|\geq 2|y|} \sup_j |\Phi_{2^j}(x-y) - \Phi_{2^j}(x)|\, dx \leq A, \quad \text{whenever } y \neq 0. \tag{49}$$

In fact, $\sup_j |\Phi_{2^j}(x-y) - \Phi_{2^j}(x)| \leq \sum_j |\Phi_{2^j}(x-y) - \Phi_{2^j}(x)|$, so fixing $y \neq 0$, the integral in (49) can be majorized by two terms: the first where the sum is taken over those j where $|y| \leq 2^j$, and the second where the sum is taken over those j where $|y| > 2^j$. Thus the integral in the first sum is majorized by

$$\sum_{2^j \geq |y|} \int_{\mathbf{R}^n} |\Phi_{2^j}(x-y) - \Phi_{2^j}(x)|\, dx = \sum_{2^j \geq |y|} \int_{\mathbf{R}^n} |\Phi(x - 2^{-j}y) - \Phi(x)|\, dx.$$

By assumption (i), this sum is dominated by

$$\sum_{2^j \geq |y|} \eta(2^{-j}|y|) \leq 2\int_0^1 \eta(s)\frac{ds}{s} = A < \infty.$$

The integral in the second sum is dominated by

$$\begin{aligned} 2\sum_{2^j<|y|} \int_{|x|\geq|y|} \Phi_{2^j}(x)\, dx &= 2\sum_{2^j<|y|} \int_{|x|\geq 2^{-j}|y|} \Phi(x)\, dx \\ &\leq 2\sum_{2^j<|y|} \eta(2^j|y|^{-1}) \leq 4\int_0^1 \eta(s)\frac{ds}{s} = A < \infty, \end{aligned}$$

proving (49).

We now consider the mapping T from L^p to $L^p(l^\infty)$ given by

$$T(f) = \{f * \Phi_{2^j}\}_{j=-\infty}^{\infty} = f * K,$$

where $K = \{\Phi_{2^j}\}_{j=-\infty}^{\infty}$. Here we are dealing with f that take their values in the Banach space $B_1 = \mathbf{R}^1$, and Tf that take their values in $B_2 = \ell^\infty(\mathbf{Z})$; thus K takes its values in $\mathcal{B}(B_1, B_2)$. We have

$$\int_{|x|\geq 2|y|} |K(x-y) - K(x)|\, dx = \int_{|x|\geq 2|y|} \sup_j |\Phi_{2^j}(x-y) - \Phi_{2^j}(x)|\, dx;$$

because of (49), we can now invoke Theorem 3 of Chapter 1 (in its Banach space version) and deduce that T is of weak-type (1,1). Thus the mapping $f \mapsto \sup_j |f * \Phi_{2^j}|$ also has this property, and the proposition is proved.

4.2.1 For later purposes it is useful to restate the essence of the above conclusion for Φ that are not necessarily decreasing in the sense of (45).

COROLLARY. *Suppose Φ is an integrable function with the property that*

$$\int |\Phi(x-y) - \Phi(x)|\, dx \leq \eta(|y|),$$

and

$$\int_{|x|\geq R} |\Phi(x)|\, dx \leq \eta(R^{-1}), \quad R \geq 1,$$

for some Dini modulus η. Then the maximal operator

$$f \mapsto \sup_j |(f * \Phi_{2^j})(x)|$$

is of weak-type $(1,1)$, and is bounded on L^p if $1 < p \leq \infty$.

Note first that if Φ satisfies our assumptions, then so does $|\Phi|$. The weak-type estimate and the L^p boundedness also follow from Theorem 3 of Chapter 1.

5. *Further results*

A. Vector-valued inequalities

5.1 We indicate why no analogue of (13) holds for the vector-valued maximal operator when our sequences take values in ℓ^1. Simply divide $[0,1]$ into N equal intervals $I_1, \ldots, I_N$, take $f_j = \chi_{I_j}$, and $f = (f_1, \ldots, f_N, 0, \ldots, 0, \ldots)$. Then $(\overline{M}_1 f)(x)$ is essentially $\log N$ for all $x \in [0,1]$, while $\|f\|_p = 1$ is independent of N, for all p, $1 \leq p \leq \infty$.

5.2 Another example of the unboundedness of $\overline{M}$ on L^∞ (see §1.1) may be given as follows. Take $f = \{f_j\}$, a vector-valued function, where $f_j = \chi_{(2^{-j}, 2^{1-j})}$, $j = 1, 2, \ldots$. Then $|f| = \chi_{(0,1)}$ is in L^∞ with compact support. Now $Mf_j(x) \geq 1/8$ whenever $0 < x < 2^{1-j}$; hence $(\overline{M}f)^2(x) \geq \sum_{1\geq 2^{1-j}\geq x} 1/64$ for $0 < x < 1$, and $\overline{M}f(x) \geq c(\log 1/|x|)^{1/2}$ for these x.

Note that the proof of Theorem 1 (in particular (12)) leads to the estimate $A_p = O(p^{1/2})$, as $p \to \infty$, for $\overline{M} = \overline{M}_2$. The example just given shows that this bound is best possible. More generally, for $\overline{M}_q$ (see §1.3.1), we have the (best possible) bound $A_p = O(p^{1/q})$, if $q > 1$.

There is also the following substitute conclusion for L^∞, which follows directly from the above estimates on A_p as $p \to \infty$: There is a $c_q > 0$ so that, whenever $f = \{f_j\}$ has compact support and $|f| \leq 1$, then $e^{c_q(\overline{M}_q f)^q}$ is integrable on every compact set (see the argument in Chapter 4, §1.3).

5.3 Let O be an open subset of $\mathbf{R}^n$ whose complement is nonempty, and write $d(x)$ for the distance of x from ${}^c O$. Then the Marcinkiewicz integral

$$I_q(x) = \int_{\mathbf{R}^n} \frac{[d(y)]^{n(q-1)}}{(|x-y| + d(y))^{nq}}\, dy$$

is comparable to $[(\overline{M}_q f)(x)]^q$ when $f = \{f_j\} = \{\chi_{Q_j}\}$ and $\{Q_j\}$ is a Whitney decomposition of O. See C. Fefferman and Stein [1971]. The Marcinkiewicz integral has a long history; see, e.g., Zygmund [1959], Calderón and Zygmund [1961], Carleson [1966], Zygmund [1969].

5.4 There is a analogue of Theorem 1 for vector-valued singular integrals. For simplicity, we state it in the context described in §6.2 of Chapter 1. Suppose $K = \{K_j\}$, with $\|\widehat{K_j}\|_{L^\infty} \leq A$ and

$$\int_{|x| \geq c|y|} |K_j(x-y) - K_j(x)|\, dx \leq A,$$

uniformly in j. If we define $T(f) = \{T_j(f_j)\} = \{f_j * K_j\}$, then we have

$$\|\, |Tf|_q \,\|_{L^p(\mathbf{R}^n)} \leq A_{p,q} \|\, |f|_q \,\|_{L^p(\mathbf{R}^n)},$$

whenever $1 < p < \infty$ and $1 < q < \infty$.

When $p = q$, the result is a direct consequence of the scalar-valued case. If $1 < p \leq q$, it follows from §6.4 of Chapter 1, and for $p > q$ a duality argument applies. See García-Cuerva and Rubio de Francia [1985]; an alternative approach uses the result in §6.15 of Chapter 5 below. The case when the K_j are all the same is treated in Chapter 1, §8.25(a).

5.5 The weighted maximal inequality (9) raises the issue of the validity of the more general version

$$\int_{\mathbf{R}^n} [Mf(x)]^p \omega_1(x)\, dx \leq A \int_{\mathbf{R}^n} |f(x)|^p \omega_2(x)\, dx,$$

for nonnegative ω_1 and ω_2. There are two questions that can be asked in this connection. First, what is the condition on ω_2 so that there exists an ω_1 that is strictly positive a.e. for which the above inequality holds? Second, what is the condition on ω_1 so that there exists an ω_2 that is finite a.e. for which the inequality is valid? The conditions are, respectively,

$$|R|^{-1/p'} \int_{|x| \leq R} [\omega_2(x)]^{-1/(p-1)}\, dx \leq A \qquad \text{for all } R \geq 1,$$

and

$$\int_{\mathbf{R}^n} \omega_1(x)\, (1 + |x|)^{-np}\, dx < \infty.$$

See Rubio de Francia [1981], Young [1982], Gatto and C. Guitérrez [1983].

B. The space $\mathcal{N}$ and Carleson measures

5.6 Let $\mathcal{N}_0$ be the closure in $\mathcal{N}$ of the continuous functions with compact support in $\bar{\mathbf{R}}^{n+1}_+$. We observe that $\mathcal{N}_0$ is exactly the set of all $F \in \mathcal{N}$ that: (i) are continuous on $\mathbf{R}^{n+1}_+$, and (ii) have nontangential limits a.e. in $\mathbf{R}^n$.

First note that

$$|F(x,t)| \leq \min_{y \in B(x,t)} |F^*(y)| \leq \frac{1}{|B(x,t)|} \int_{B(x,t)} F^*(y)\, dy \leq ct^{-n} \|F\|_{\mathcal{N}}.$$

Therefore convergence in $\mathcal{N}$ implies uniform convergence in $\mathbf{R}^{n+1}_+$ away from the boundary; hence every member of $\mathcal{N}_0$ is continuous on $\mathbf{R}^{n+1}_+$. The existence of nontangential limits follows by the usual arguments since, by definition, such limits exist on a dense subspace of $\mathcal{N}_0$.

For the converse, let $F \in \mathcal{N}$ satisfy (i) and (ii) and set $F_\varepsilon(x,t) = F(x, t+\varepsilon)$. F_ε is continuous on $\bar{\mathbf{R}}^{n+1}_+$ and can be approximated (in $\mathcal{N}$) by continuous functions with compact support in $\bar{\mathbf{R}}^{n+1}_+$. Now $(F_\varepsilon - F)^*(x) \leq 2F^*(x)$ for all x, while $(F_\varepsilon - F)^*(x) \to 0$, as $\varepsilon \to 0$, for any $x \in \mathbf{R}^n$ at which F has a nontangential limit. Applying the dominated convergence theorem gives $\int_{\mathbf{R}^n} (F_\varepsilon - F)^*(x)\, dx \to 0$. Therefore $F_\varepsilon \to F$ in $\mathcal{N}$.

5.7 That "duality" between $\mathcal{N}$ and the Carleson measures is, broadly speaking, equivalent to the atomic decomposition in $\mathcal{N}$ may be seen as follows. First, if we have the atomic decomposition, then the duality inequality (a) of Theorem 2 is an immediate consequence. Conversely, one may argue heuristically as follows. Consider the space $\mathcal{N}_0$ arising in §5.6, and let $\tilde{\mathcal{N}} \subset \mathcal{N}_0$ denote the subspace of elements that have atomic decompositions (of the form (27) and (28)) whose atoms a_k are *continuous* on $\bar{\mathbf{R}}^{n+1}_+$. Then $\mathcal{N}_0$ and $\tilde{\mathcal{N}}$ have the same dual space (namely $\overline{\mathcal{C}}$, as defined in §2.5.2), and thus $\mathcal{N}_0 = \tilde{\mathcal{N}}$. Similar heuristics show that the duality of H^1 and BMO (in Chapter 4) is equivalent to the atomic decomposition of H^1 given in Chapter 3.

5.8 Suppose $F(x,t)$ is given on $\mathbf{R}^{n+1}_+$ and $F^+(x) = \sup_t |F(x,t)|$.

(a) Note that $\{x : F^+(x) > \alpha\}$ is the projection π of the set $\{(x,t) : |F(x,t)| > \alpha\}$.

(b) If $E \subset \mathbf{R}^{n+1}$ is Lebesgue-measurable, then $\pi(E) \subset \mathbf{R}^n$ is not, in general, Lebesgue-measurable.

(c) If $E \subset \mathbf{R}^{n+1}$ is a Borel set, then $\pi(E)$ is not necessarily a Borel set in $\mathbf{R}^n$.

(d) However, if E is a Borel set in $\mathbf{R}^{n+1}$, then $\pi(E)$ is a Souslin (analytic) set in $\mathbf{R}^n$, and hence is Lebesgue-measurable.

These results illustrate that, at least as far as measurability is concerned, the nontangential function F^* is better behaved than the vertical function F^+. For (c) and (d) see Hausdorff [1937] and Saks [1937].

5.9 Let $d\mu$ be a nonnegative measure on $\mathbf{R}^{n+1}_+$ so that

$$\mu(T(B)) \le c|B|^\gamma, \qquad \text{for some } \gamma \ge 1,$$

and all balls $B \subset \mathbf{R}^n$. Then:

(a) $\left(\int_{\mathbf{R}^{n+1}_+} |F(x,t)|^\gamma \, d\mu(x,t)\right)^{1/\gamma} \le c\|F\|_{\mathcal{N}}$, for $F \in \mathcal{N}$.

(b) $\left(\int_{\mathbf{R}^{n+1}_+} |F(x,t)|^q \, d\mu(x,t)\right)^{1/q} \le c\|F^*\|_{L^p(\mathbf{R}^n)}$, when $0 < p < \infty$ and $q = \gamma p$.

(c) If $F(x,t) = (f * \Phi_t)(x)$ is as in §2.4, then

$$\left(\int_{\mathbf{R}^{n+1}_+} |F(x,t)|^q \, d\mu(x,t)\right)^{1/q} \le c\|f\|_{L^p(\mathbf{R}^n)}$$

when $1 < p \le \infty$ and $q = \gamma p$.

Conclusion (c) is in Duren [1969]. To prove (a), one uses the atomic decomposition; (b) and (c) then follow.

C. Applications

5.10 An illustrative example of Theorem 4 is as follows. In $\mathbf{R}^1$, let

$$(M^{(\gamma)}f)(x) = \sup_{j\in\mathbf{N}} |I_j|^{-1} \int_{I_j} |f(x-y)| \, dy,$$

where I_j is the interval of length 2^{-j} displaced by $\gamma 2^{-j}$, $\gamma \ge 1$; that is,

$$I_j = [\gamma 2^{-j}, (\gamma+1)2^{-j}].$$

Then $M^{(\gamma)}$ has a weak-type (1,1) bound that is $O(\log\gamma)$ as $\gamma \to \infty$; also, the L^p bound of $M^{(\gamma)}$ is $O(\log\gamma)^{1/p}$, for $p > 1$.

However, if $M^{(\gamma)}$ is replaced by its continuous analogue,

$$\sup_{\delta\in\mathbf{R}^+} \delta^{-1} \int_{\gamma\delta}^{(\gamma+1)\delta} |f(x-y)| \, dy,$$

then the best weak-type bound is $O(\gamma)$ as $\gamma \to \infty$ and the best L^p bound is $O(\gamma^{1/p})$.

Incidentally, another proof for the estimates on $M^{(\gamma)}$ can be obtained from the arguments in §4.2 by showing that

$$\int_{|x|\ge 2|y|} \sup_j |\chi_j(x-y) - \chi_j(x)| \, dx = O(\log\gamma), \qquad \gamma \to \infty,$$

if $\chi_j(x) = 2^j\chi(2^jx - \gamma)$ and χ is the characteristic function of the unit interval $[0,1]$. An application of these estimates to Poisson integrals on symmetric spaces can be found in Sjögren [1986]; see also Carlsson, Sjögren, and Strömberg [1985].

5.11 When the basic condition $|\mathcal{B}(r)| \le cr^n$ of §3.1 fails, the appropriate modification is as follows. Let

$$\eta(r) = \inf_{\rho^n \ge r} \frac{\rho^n}{|\mathcal{B}(\rho)|},$$

and define

$$(M_pf)(x) = \sup_{B\in\tilde{\mathcal{B}}} \frac{\eta(|B|)^{1/p}}{|B|} \int_B |f(x-y)| \, dy.$$

One then has that M_p is bounded from $L^p(\mathbf{R}^n)$ to itself, if $p > 1$. For results of this kind, see Nagel and Stein [1984].

An example in $\mathbf{R}^1$ is given by

$$(M_{\alpha,\beta}f)(x) = \sup_{0<\delta\le 1} \delta^{\beta-\alpha} \int_\delta^{\delta+\delta^\alpha} |f(x-y)| \, dy,$$

with $\alpha \ge 1$. Then $M_{\alpha,\beta}$ is bounded from L^p to itself if $p \ge \frac{\alpha}{\beta+1}$ and $p > 1$, but not if $p < \frac{\alpha}{\beta+1}$; to see the second assertion, note that if $f \ge 0$ and $f(x) = |x|^{-\gamma}$ near 0, then $(M_{\alpha,\beta}f)(x) \ge cx^{\beta-\alpha\gamma}$ when $0 < x \le 1$. Note that $M_{\alpha,\beta}f \le c[M(|f|^p)]^{1/p}$‡ if $p = \frac{\alpha}{\beta+1}$, so the usual theory shows that $M_{\alpha,\beta}$ is of weak-type (p,p); however, this is not the case for the L^p boundedness of $M_{\alpha,\beta}$.

5.12 An application of §5.11 deals with the "tangential approach" for $f * \Phi_t$ when f belongs to the Sobolev space L^p_α.† The result is as follows.

Let $f \in L^p_\alpha$ and let Φ be a function on $\mathbf{R}^n$ with

$$|\Phi(x)| \le A(1+|x|)^{-n-1}.$$

Then, with $F(x,t) = (f * \Phi_t)(x)$ and

$$F^{*,\beta}(x) = \sup_{|y-x|\le t^\beta} |F(y,t)|,$$

we have

$$\|F^{*,\beta}\|_{L^p(\mathbf{R}^n)} \le A_{p,\alpha}\|f\|_{L^p_\alpha(\mathbf{R}^n)}$$

whenever $\beta = 1 - \frac{\alpha p}{n}$, $\beta > 0$, $1 < p$, and $\alpha \ge 0$.

See Nagel, Rudin, and Shapiro [1982] for the original proof, which involved ideas from capacity theory.

‡ Here M is the standard maximal operator.

† See Chapter 6, §5.21 for the definition of L^p_α.

5.13 Suppose $Tf = f * K$ is a translation-invariant singular integral on $\mathbf{R}^n$, whose kernel K satisfies $\|\widehat{K}\|_\infty \leq A$ and

$$|K(x-y) - K(x)| \leq A\frac{|y|^\varepsilon}{|x|^{n+\varepsilon}}, \qquad \text{whenever } |x| \geq 2|y|,$$

for some fixed $\varepsilon > 0$. Let $K^t(x) = K(x) \cdot \chi_{|x|\leq t}(x)$ be the truncated kernel. Then, whenever a is a bounded function, we have for $p < \infty$

$$\int_B \sup_{0<t<r} |(a * K^t)(x)|^p \, dx \leq c_p |B| \cdot \|a\|_{L^\infty}^p,$$

for all balls B having radius r, and all $r > 0$. Thus $\mu(x,t) = |(a * K^t)(x)|$ is a multiplier that satisfies the hypotheses of the proposition in §3.2.

See Coifman, Y. Meyer, and Stein [1985], where more involved examples of this kind are also given. Notice that, by §7.3 in Chapter 1, $\sup_{t>0} |(a*K^t)(x)| < \infty$ a.e.; but this function may be unbounded near every $x \in \mathbf{R}^n$.

5.14 The results in §1 and §2 of this chapter extend with little change to the general real-variable setting described in Chapter 1, §1. For this purpose, we recall the family of balls $\{B(x,\delta)\}$ and the fixed measure $d\mu$ postulated there. If

$$(Mf)(x) = \frac{1}{\mu(B(x,\delta))} \int_{B(x,\delta)} |F(y)| \, d\mu(y)$$

is the maximal function, then the vector-valued analogues $\overline{M}(f)$ (and $\overline{M}_q(f)$) are defined as in (1).

For the tent spaces, we write $F^*(x) = \sup_{y \in B(x,t)} |F(y,t)|$, whenever F is given on $\mathbf{R}^{n+1}_+$; tents are defined by

$$T(B(x,\delta)) = \{(y,t) \in \mathbf{R}^{n+1}_+ : y \in B(x,\delta) \text{ and } 0 < t \leq \delta\}.$$

With these definitions, the results of theorems 1 and 2 continue to hold; the proofs require only minor modifications. These observations are applied in Sueiro [1986].

D. Singular maximal functions

5.15 Let $\{\Phi^{(j)}\}$ be a sequence of functions that satisfy the hypotheses of §4.2.1 uniformly in j, namely,

$$\int_{\mathbf{R}^n} |\Phi^{(j)}(x-y) - \Phi^{(j)}(x)| \, dx \leq \eta(|y|),$$

and

$$\int_{|x|\geq R} |\Phi^{(j)}(x)| \, dx \leq \eta(R^{-1}),$$

for some Dini modulus η.

(a) The maximal operator

$$f \mapsto \sup_j |(f * \Phi^{(j)}_{2^j})(x)|$$

is of weak-type $(1,1)$ and is bounded on L^p, if $1 < p \leq \infty$. Here $\Phi^{(j)}_{2^j}(x) = 2^{-nj}\Phi^{(j)}(2^{-j} \cdot x)$.

(b) One may also formulate a singular integral analogue. Suppose $\Phi^{(j)}$ satisfy the above properties, and additionally the cancellation condition

$$\int_{\mathbf{R}^n} \Phi^{(j)}(x) \, dx = 0, \qquad \text{all } j.$$

Then the series $\sum_j \Phi^{(j)}_{2^j}$ converges to a distribution K. Moreover, the operator $Tf = f * K$ is of weak-type (1,1) and is bounded on L^p, $1 < p < \infty$.

In fact, one can show first that

$$\sum_j |\widehat{\Phi^{(j)}}(2^j\xi)| \leq A,$$

which proves the convergence of the series defining K and the L^2 boundedness of T. Moreover, the reasoning of §4.2 shows that

$$\int_{|x|\geq 2|y|} |K(x-y) - K(x)| \, dx \leq A,$$

and so the theory of Chapter 1 is applicable.

5.16 The weak conditions imposed on Φ in §4.2.1 are enough to guarantee control of the maximal operator

$$f \mapsto \sup_j |(f * \Phi_{2^j})(x)|,$$

but are far from sufficient to allow similar conclusions for the continuous analogue

$$f \mapsto \sup_{t \in \mathbf{R}^+} |(f * \Phi_t)(x)|.$$

To see this, fix $\varepsilon > 0$, and let Φ be defined on $\mathbf{R}^1$ by

$$\Phi(x) = (1 - |x|^2)^{\varepsilon - 1} \qquad \text{for } |x| < 1,$$

and $\Phi(x) = 0$ for $|x| \geq 1$. Then Φ satisfies the aforementioned hypotheses with $\eta(u) = cu^\varepsilon$. However, one can easily find an $f \in L^p$, with $p < 1/\varepsilon$, so that

$$\sup_{t>0} |(f * \Phi_t)(x)| = \infty \qquad \text{for all } x \in \mathbf{R}^1.$$

Indeed, let $\delta \geq \varepsilon$ with $\delta p < 1$, and take $f(x) = |x|^{-\delta}$ for $|x| \leq 1$, $f(x) = 0$ for $|x| \geq 1$. Then $f \in L^p(\mathbf{R}^1)$, while $(f * \Phi_t)(x) = \infty$ when $t = |x|$.

Some positive results may be found in Chapter 11, §4.10.

5.17 Suppose P is a polynomial on $\mathbf{R}^n$ so that

$$\int_{\mathbf{R}^n} |P(x)|^{-a} < \infty$$

for some $a > 0$. Then $\Phi(x) = |P(x)|^{-a}$ satisfies the hypotheses of §4.2.1; as a result, we can assert that

$$f \mapsto \sup_{j \in \mathbf{N}} |(f * \Phi_{2^j})(x)|$$

is of weak-type (1,1) and is bounded on $L^p(\mathbf{R}^n)$, $1 < p \leq \infty$. If, in addition, $r \mapsto P(rx)$ is increasing in r, $0 < r < \infty$, for each $x \in \mathbf{R}^n$, then the same conclusions hold for

$$f \mapsto \sup_{t \in \mathbf{R}^+} |(f * \Phi_t)(x)|.$$

To prove these assertions, let

$$\Phi^{(s)}(x) = |P(x)|^{-a(1-s)}(1 + |x|)^{-cs},$$

where c is a (large) positive constant. When $\mathrm{Re}(s) = 1$, it follows that

$$\int_{\mathbf{R}^n} |\Phi^{(s)}(x)|\, dx < \infty \qquad \text{and} \qquad \int_{\mathbf{R}^n} |\nabla_x \Phi^{(s)}(x)|\, dx < \infty,$$

as long as $c \geq n + \mathrm{degree}(P)$. It can also be shown that

$$\int_{\mathbf{R}^n} |\Phi^{(s)}(x)|\, dx < \infty \qquad \text{when } \mathrm{Re}(s) < 0$$

and $\mathrm{Re}(s)$ is sufficiently small. By complex convexity, $\Phi^{(0)} = \Phi$ then satisfies the hypotheses in §4.2.1, with $\eta(u) = cu^\varepsilon$, for some $\varepsilon > 0$.

For reasoning of this type, see Stein [1976c], [1983b].

5.18 The result sketched in §5.17 represents the key idea in proving "restricted" convergence, together with L^1 and L^p maximal inequalities, for Poisson integrals in general Riemannian symmetric spaces.

The simplest examples to describe arise for tube domains over cones. Three particular instances are:

(i) The cone $\{y \in \mathbf{R}^n : y_j > 0, \text{all } j\}$; the tube domain corresponds to an n-fold product of $\mathbf{R}^2_+$. Here $\Phi(x) = c \prod_{j=1}^{n} (1 + x^2)^{-1}$.

(ii) The circular cone $\{y \in \mathbf{R}^n : y_n > (y_1^2 + \cdots + y_{n-1}^2)^{1/2}\}$; then

$$\Phi(x) = c[(x_1^2 + x_2^2 + \cdots + x_{n-1}^2 - x_n^2 + 1)^2 + 4x_n^2]^{-n/2}.$$

(iii) When there is an integer m with $n = m(m+1)/2$, we can represent points $x \in \mathbf{R}^n$ by real, symmetric $m \times m$ matrices, and the cone consists of the positive-definite matrices. In this case

$$\Phi(x) = c|\det(x + iI)|^{-n-1}.$$

The general result requires a formulation of §5.17 in which $\mathbf{R}^n$ (with its isotropic dilations) is replaced by a homogeneous group (see Chapter 13, §7.11(b)). Further details are in Stein [1983]. The examples described above were originally treated by different methods; see Marcinkiewicz and Zygmund [1939b], Stein and N. J. Weiss [1969]. For background about Poisson integrals on tube domains, see Chapter 3 of *Fourier Analysis* and N. J. Weiss [1972]; in the more general setting of symmetric spaces, see Korányi [1972].

5.19 We consider the maximal operator

$$(M_\Omega f)(x) = \sup_{r>0} r^{-n} \int_{|y| \leq r} |f(x-y)|\, \Omega(y)\, dy,$$

where Ω is a nonnegative function that is homogeneous of degree 0. When Ω satisfies an L^1 Dini condition, we saw in §4 that M_Ω is of weak-type (1,1) (and thus is bounded on L^p, $p > 1$). Here are two extensions of this result.

(a) The conclusion still holds if Ω, considered as a function on the unit sphere $\mathbf{S}^{n-1}$, has *finite* L^1 *entropy* in the sense that

$$\Omega \leq \sum c_j |B_j|^{-1} \chi_{B_j},$$

where $B_j \subset \mathbf{S}^{n-1}$ are geodesic balls, $|B_j|$ are their induced volumes, and

$$\sum c_j [1 + \log^+(1/c_j)] < \infty.$$

One can also prove that if Ω satisfies an L^1 Dini condition, then it has finite L^1 entropy. R. Fefferman [1978].

(b) The conclusion also holds if $\Omega \log^+ \Omega \in L^1(\mathbf{S}^{n-1})$, in particular if $\Omega \in L^p(\mathbf{S}^{n-1})$, $p > 1$. See Christ and Rubio de Francia [1988]. Again, one can show that if Ω enjoys an L^1 Dini condition, then $\Omega \log^+ \Omega \in L^1(\mathbf{S}^{n-1})$. The question left unanswered by these results is whether M_Ω is weak (1,1) bounded when Ω is merely in $L^1(\mathbf{S}^{n-1})$.

E. Multi-parameter maximal functions

5.20 The "product theory" of $\mathbf{R}^n$, which involves multi-parameter scalings, leads one to consider the *strong maximal operator*

$$(M_S f)(x) = \sup_R |R|^{-1} \int_R |f(x-y)|\, dy,$$

where the supremum is taken over all rectangles R centered at the origin, with sides parallel to the axes.

(a) Let $(M_jf)(x) = (M^{(e_j)}f)(x)$ be the standard maximal function in the direction e_j (here $e_1, \ldots, e_n$ is the usual basis of $\mathbf{R}^n$), as in §4.1. Then

$$M_S(f) \leq M_1(M_2(\cdots(M_n(f))\cdots)),$$

and as a result

$$\|M_S(f)\|_{L^p(\mathbf{R}^n)} \leq A_p^n \|f\|_{L^p(\mathbf{R}^n)}, \qquad 1 < p \leq \infty.$$

(b) There is no L^1 theory for M_S (when $n \geq 2$); see Chapter 10, §2.3.

(c) The substitute result for L^p with p near 1 can be stated in terms of the class $L(\log L)^{n-1}$; in particular, if f is locally in this class, then

$$\lim_{\operatorname{diam}(R)\to 0} |R|^{-1} \int_R f(x-y)\,dy = f(x) \qquad \text{a.e.}\,.$$

Conclusions (a) and (c) go back to Jessen, Marcinkiewicz, and Zygmund [1935]. Negative results stronger than (b) are in Saks [1935]. Further discussion of M_S (called $M_{\mathcal{R}_0}$ there) can be found in Chapter 10, §2.3.

5.21 The standard maximal operator of Chapter 1 is closely connected with the Vitali covering lemma in §3.1 of that chapter. The following covering lemma plays the same role for M_S.

Let $\{R_\alpha\}$ be a collection of rectangles in $\mathbf{R}^n$, contained in the unit ball B, with sides parallel to the axes. Then there is a finite subcollection $\{R_j\}_{j=1}^N$ so that

$$\left|\bigcup_{j=1}^N R_j\right| \geq c_n \left|\bigcup_\alpha R_\alpha\right|,$$

and

$$\int_B \exp\Big(\sum_{j=1}^N \chi_{R_j}\Big)\,dx \leq c_n'.$$

See Córdoba and R. Fefferman [1975].

5.22 If Φ is a function on $\mathbf{R}^n$, then for each n-tuple $J = (j_1, \ldots, j_n) \in \mathbf{Z}^n$, we define

$$\Phi_{2^J}(x) = 2^{-j_1-\cdots-j_n}\Phi(2^{-j_1}x_1, \ldots, 2^{-j_n}x_n).$$

One may ask, in analogy with §6.1 of Chapter 1 and §4.2.1 of the present chapter, for sufficient conditions on Φ so that the mapping

$$f \mapsto \sup_{J\in\mathbf{Z}^n} |(f * \Phi_{2^J})(x)|$$

is bounded on L^p, $1 < p \leq \infty$.

(a) One notes first that if

$$|\Phi(x)| \leq A \prod_{k=1}^n (1 + |x_k|)^{-1-\varepsilon}$$

for some $\varepsilon > 0$, then

$$\sup_{J\in\mathbf{Z}^n} |(f * \Phi_{2^J})(x)| \leq c(M_Sf)(x).$$

It should be observed that here a similar conclusion holds in the non-dyadic case, that is, for

$$\sup_{t\in(\mathbf{R}^+)^n} |(f * \Phi_t)(x)|,$$

where $t = (t_1, \ldots, t_n)$ and

$$\Phi_t(x) = (t_1 \cdots t_n)^{-1}\Phi(x_1/t_1, \cdots, x_n/t_n).$$

(b) Another, less evident, condition for the L^p boundedness of

$$f \mapsto \sup_{J\in\mathbf{Z}^n} |(f * \Phi_{2^J})(x)|$$

when $p > 1$, is that Φ has compact support and satisfies

$$\int_{\mathbf{R}^n} |\Phi(x-y) - \Phi(x)|\,dx \leq A|y|^\varepsilon,$$

for some $\varepsilon > 0$. See Ricci and Stein [1992], in which multi-parameter analogues of §5.15 may also be found. For some related results, see Nagel and Wainger [1977], Duoandikoetxea [1986], Carbery and Seeger [1993].

(c) A further sufficient condition on Φ is that $\Phi(x) = [P(x)]^{-a}$, where P is a strictly positive polynomial and a is such that

$$\int_{\mathbf{R}^n} [P(x)]^{-a}\,dx < \infty.$$

To prove this, one may use the idea sketched in §5.17 together with the arguments in Sjögren [1986]. This paper should also be consulted for applications of such results to prove "unrestricted" convergence of Poisson integrals on symmetric spaces.

5.23 For the product theory of singular integrals one should consult R. Fefferman and Stein [1982], Journé [1985], R. Fefferman [1986] and [1987], Pipher [1986], as well as the references in §5.22(b).

Notes

§*1*. The vector-valued maximal theorem can be found in C. Fefferman and Stein [1971]; an earlier partial result is in Stein [1970a].

§*2*. The definition of Carleson measures, and the inequality (24), go back to Carleson [1962]. The tent space $\mathcal{N}$ and its atomic decomposition appear in Coifman, Y. Meyer, and Stein [1985], but the duality inequality (Theorem 2) is implicit in C. Fefferman and Stein [1971].

§*3*. The general version of the maximal theorem for arbitrary collections of balls (stated in terms of "approach regions") is in Nagel and Stein [1984]. The multipliers in §3.2 were characterized by Coifman, Y. Meyer, and Stein [1985].

§*4*. The L^p maximal theorem in Proposition 1 is an unpublished observation of Coifman and G. Weiss. The reasoning proving Proposition 2 originates with Zó [1976] and has since been developed to fit various circumstances.

CHAPTER III

Hardy Spaces

The study of Hardy spaces, which originated during the 1910's and 1920's in the setting of Fourier series and complex analysis in one variable, has over time been transformed into a rich and multifaceted theory, providing basic insights into such topics as maximal functions, singular integrals, and L^p spaces. Here we want to emphasize four aspects of this theory.

1. *Extension of L^p*. To put the matter simply but somewhat imprecisely, the elements of H^p are (tempered) distributions for which an appropriate meaning can be given to their pointwise values almost everywhere and that, in a suitably refined sense, belong to $L^p(\mathbf{R}^n)$. When $p > 1$, the actual definition of H^p makes it equivalent to L^p, but when $p \leq 1$, these spaces are much better suited to a host of questions in harmonic analysis than are the L^p spaces.

2. *Equivalence of definitions.* The main thrust of what is proved in this chapter is that a variety of distinct approaches, based on differing definitions, all lead to the same notion of H^p. Among these are: the several *maximal* definitions, of which the simplest is that a distribution f belongs to H^p if, for some $\Phi \in \mathcal{S}$ with $\int \Phi = 1$, the maximal function

$$(M_\Phi f)(x) = \sup_{t>0} |(f * \Phi_t)(x)|$$

is in L^p; the *atomic* decomposition, which allows one to express H^p distributions f as sums of very simple constituents (conversely, such sums are obviously in H^p); and the fact that, broadly speaking, H^p consists of that part of L^p that is stable under the action of singular integrals.

3. *Nature of H^p*. In contrast with the case $p > 1$, the question as to whether a distribution f belongs to H^p, $p \leq 1$, is not only a matter of the *size* of f but also involves some delicate cancellation properties. Thus in the maximal definition given above, the operator M_Φ cannot be replaced by the standard maximal operator M (appearing in chapters 1 and 2), because the latter involves only the absolute value of f; nor can Φ be replaced by a function that is not sufficiently smooth. The cancellation properties also enter (in a very direct way) into the definition

of H^p atoms given below. It is these subtleties that are responsible for the arduous nature of some of the arguments in this chapter, but the rewards for the patient reader are the elegant and far-reaching assertions that one is able to prove.

4. *Applicability.* Among the necessary properties that determine whether a distribution belongs to some H^p are that it be "bounded" in a suitably weak (average) sense, and that it be (at worst) singular of finite order in an appropriate pointwise sense. In the converse direction is the fact that large classes of generalized functions appearing in diverse parts of analysis are (locally) equal to H^p distributions. Examples of this are: singularities represented by meromorphic functions, arbitrary distributions carried on smooth submanifolds, and the Lagrangian distributions that occur in the theory of Fourier integral operators. These assertions may be found in §5.18–§5.20.

Two more remarks about H^p theory may be in order. After its initial flowering in the setting of one complex variable, the requirements of analysis on $\mathbf{R}^n$ led (in the 1960's) to an extension of the theory in which the main stress was placed on harmonic and subharmonic functions; it was then that the stability under singular integrals (Theorem 4) was proved in its first form.[†] However, the point of view that developed later (which we adopt here) does not depend in any essential way on analytic or harmonic functions. Its more flexible approach allows for substantial further generalizations.

Finally, we point to the only H^p space, $p \leq 1$, that is also a Banach space—namely H^1. Its special role and rich properties form much of the subject matter treated in the next chapter.

1. *Maximal characterization of H^p*

As we said above, it is our intention to define H^p as a space of distributions on $\mathbf{R}^n$ in such a way as to give a natural real-variable generalization of the space L^p when $p > 1$. We first make some comments about tempered distributions on $\mathbf{R}^n$.

1.1 One begins with the space $\mathcal{S}$ of *testing functions*: the set of all ϕ on $\mathbf{R}^n$ that are infinitely differentiable and, together with all their derivatives, are rapidly decreasing (i.e., remain bounded when multiplied

[†] For an account of the theory at that stage of development, see *Fourier Analysis*, chapters 2 and 6 and *Singular Integrals*, Chapter 7, as well as the appendix to this chapter.

by arbitrary polynomials). On $\mathcal{S}$ one has a denumerable collection of seminorms $\|\cdot\|_{\alpha,\beta}$ given by

$$\|\phi\|_{\alpha,\beta} = \sup_{x\in\mathbf{R}^n} |x^\alpha \partial_x^\beta \phi(x)|.$$

Here we use conventional notation:

$$x^\alpha = x_1^{\alpha_1}\cdots x_n^{\alpha_n}, \qquad \partial_x^\beta = \frac{\partial^{\beta_1}}{\partial x_1^{\beta_1}}\cdots\frac{\partial^{\beta_n}}{\partial x_n^{\beta_n}};$$

$\alpha = (\alpha_1,\dots,\alpha_n)$ and $\beta = (\beta_1,\dots,\beta_n)$ are n-tuples of natural numbers. A *tempered distribution* is a linear functional on $\mathcal{S}$ that is continuous in the topology on $\mathcal{S}$ induced by this family of seminorms. We shall refer to tempered distributions simply as *distributions*; the set of all distributions (with the weak topology) is denoted by $\mathcal{S}'$.[‡]

Whenever f is a distribution and $\Phi \in \mathcal{S}$, the convolution $f * \Phi$ is a well-defined C^∞ function (possibly slowly increasing at ∞). Of course, the same thing is true for $f * \Phi_t$, where $\Phi_t(x) = t^{-n}\Phi(x/t)$, $t > 0$.

The families Φ_t (with $\Phi \in \mathcal{S}$) will serve as basic approximate identities in our characterization of H^p. However, the fundamental connection with harmonic functions makes it useful to consider also $f * P_t$, where P_t is the Poisson kernel[†]

$$P(x) = P_1(x) = \frac{c_n}{(1+|x|^2)^{(n+1)/2}}, \qquad P_t(x) = t^{-n}P(x/t).$$

Since $f * P_t$ is not meaningful for a general tempered distribution f, H^p theory carries with it a natural restriction on the class of distributions considered: We say that a distribution f is *bounded* if $f * \Phi \in L^\infty(\mathbf{R}^n)$ whenever $\Phi \in \mathcal{S}$. It is not difficult to see that a distribution is bounded exactly when its translates form a bounded set in $\mathcal{S}'$.

Let us point out some useful properties of bounded distributions. First, if f is a bounded distribution and $h \in L^1(\mathbf{R}^n)$, then the convolution $f * h$ can be defined as a distribution. Indeed, let $\phi \in \mathcal{S}$ and write

$$\langle f * h, \phi\rangle = \langle f * \tilde{\phi}, \tilde{h}\rangle = \int_{\mathbf{R}^n} (f * \tilde{\phi})(x)\,\tilde{h}(x)\,dx,$$

where $\tilde{\phi}(x) = \phi(-x)$. Second, it is easily verified that $f * h$ is a bounded distribution and that $f * (h_1 * h_2) = (f * h_1) * h_2$ whenever $h_i \in L^1(\mathbf{R}^n)$.

[‡] A review of some basic points in the theory of tempered distributions may be found in *Fourier Analysis*, Chapter 1, §3.

[†] As defined in Chapter 1, §6.1.

Since P is in L^1, it follows that $f * P_t$ is a well-defined distribution whenever f is bounded. Moreover we claim that, for any fixed t, $f * P_t$ is a bounded C^∞ function. In fact, we can write $P = \phi * h + \psi$ with $\phi, \psi \in \mathcal{S}$ and $h \in L^1$. To see this, notice that $\widehat{P}(\xi) = e^{-2\pi|\xi|}$ is rapidly decreasing and fails to be smooth only at the origin.‡ Taking $h = P$ and $\phi \in \mathcal{S}$ with $\widehat{\phi}(\xi) = 1$ for ξ near 0, we need only set $\widehat{\psi}(\xi) = (1 - \widehat{\phi}(\xi))e^{-2\pi|\xi|}$. Therefore

$$P_t = \phi_t * h_t + \psi_t \qquad \text{and} \qquad f * P_t = f * \psi_t + (f * \phi_t) * h_t,$$

from which our assertion follows. Since $[(\partial_t)^2 + \sum_{j=1}^{n} (\partial_{x_j})^2] P_t(x) = 0$, we also obtain that the function $u(x,t) = (f * P_t)(x)$ is C^∞ in (x,t) and is itself harmonic.

1.2 The maximal characterization. We come now to the formulation of the maximal characterization of H^p.

For any $\Phi \in \mathcal{S}$ and any distribution f, we define $M_\Phi f(x)$ by

$$M_\Phi f(x) = \sup_{t>0} |(f * \Phi_t)(x)|. \tag{1}$$

Next, instead of considering a maximal function based only on a single approximation of the identity, we shall also consider "grand maximal functions" based on collections of such Φ. For this purpose, let $\mathcal{F} = \{\|\cdot\|_{\alpha_i,\beta_i}\}$ be any *finite* collection of seminorms on $\mathcal{S}$. We denote by $\mathcal{S}_\mathcal{F}$ the subset of $\mathcal{S}$ controlled by this collection of seminorms; more precisely, we set

$$\mathcal{S}_\mathcal{F} = \{\Phi \in \mathcal{S} : \|\Phi\|_{\alpha_i,\beta_i} \le 1 \text{ for all } \|\cdot\|_{\alpha,\beta} \in \mathcal{F}\}.$$

We then write

$$\mathcal{M}_\mathcal{F} f(x) = \sup_{\Phi \in \mathcal{S}_\mathcal{F}} M_\Phi f(x). \tag{2}$$

Finally, whenever f is a bounded distribution, let

$$u(x,t) = (f * P_t)(x)$$

be the Poisson integral of f, and let

$$u^*(x) = \sup_{|x-y| \le t} |u(y,t)| \tag{3}$$

denote the nontangential maximal function of u (as in §2 of Chapter 2).

‡ We continue to use the common notation $\widehat{\phi}$ for the Fourier transform of ϕ.

THEOREM 1. *Let f be a distribution and let $0 < p \le \infty$. Then the following conditions are equivalent:*

(i) *There is a $\Phi \in \mathcal{S}$ with $\int \Phi\, dx \ne 0$ so that $M_\Phi f \in L^p(\mathbf{R}^n)$.*

(ii) *There is a collection $\mathcal{F}$ so that $\mathcal{M}_\mathcal{F} f \in L^p(\mathbf{R}^n)$.*

(iii) *The distribution f is bounded and $u^* \in L^p(\mathbf{R}^n)$.*

If any of these three equivalent properties are satisfied, then we say that f belongs to $H^p(\mathbf{R}^n)$. Before coming to the proof of the theorem, we make a series of remarks that may help to clarify the nature of the spaces H^p.

1.2.1 When $1 < p \le \infty$, the conditions are equivalent with f being a function in $L^p(\mathbf{R}^n)$, so that for this range $H^p(\mathbf{R}^n)$ may be identified with $L^p(\mathbf{R}^n)$. In fact, if $f \in L^p(\mathbf{R}^n)$ then, as we have seen in §2.1 of Chapter 2, $M_\Phi f(x) \le cMf(x)$, and therefore $M_\Phi f \in L^p$, by the maximal theorem (in Chapter 1). Conversely, suppose $\int \Phi = 1$ and $M_\Phi f \in L^p$. Then $f * \Phi_{1/n}$ is a bounded sequence in L^p, and by the weak compactness of L^p (as the dual of $L^{p'}$), there is an $f_0 \in L^p$ and a subsequence $f * \Phi_{1/n_j}$ with $f * \Phi_{1/n_j} \to f_0$ weakly. However, $f * \Phi_{1/n_j} \to f$ in the sense of distributions; thus $f = f_0 \in L^p$.

1.2.2 Let us consider the case $p = 1$. By a suitable modification of what was said above it follows that $H^1 \subset L^1$; see §2.3.3 below. However, the converse is false. In the first instance, for a function f to belong to H^1, it must satisfy the moment condition $\int f\, dx = 0$; see §5.4. Besides cancellation conditions of this kind (which implicitly occur at every level), there must also be restrictions on the size of f. Thus if f is positive on a ball B, it can belong to H^1 only if $f \log(1+f)$ is integrable on every ball that is strictly contained in B (see §5.3).

1.2.3 The size and cancellation conditions required of H^1 functions are most neatly expressed in special examples (called *atoms*) that serve also as basic building blocks for H^1 functions. A function a is an H^1 atom (associated to a ball B) if:

(i) a is supported in B,

(ii) $|a| \le |B|^{-1}$ almost everywhere, and

(iii) $\int a\, dx = 0$.

For fixed $\Phi \in \mathcal{S}$, it is easy to show that (see §2.2 below)

$$\|M_\Phi a\|_{L^1} \le c,$$

with c independent of a and B. In particular, a belongs to H^1; and moreover so does the function

$$f = \sum \lambda_k a_k,$$

whenever $\sum |\lambda_k| < \infty$ and the a_k are atoms (associated to possibly differing balls B_k). The fundamental fact that an *arbitrary* H^1 function f can be expressed as such a sum will be taken up after the proof of Theorem 1.

1.2.4 A further remark about the connection between L^p and H^p for p close to 1 illustrates the relation between the assumed smoothness of Φ and the implied moment conditions on elements of H^p.

Suppose f is an integrable function (say with compact support) that satisfies the necessary moment condition $\int f\,dx = 0$. If we assume that $f \in L^q$ for some $q > 1$, then we also have that $f \in H^1$. In fact $M_\Phi(f) \le cM(f)$, where M is the standard maximal operator, thus $M_\Phi(f)$ is locally integrable. However the smoothness of Φ, together with the cancellation condition on f, shows easily that $M_\Phi(f) \le c|x|^{-n-1}$ away from the support of f. Thus $\mathcal{M}_\Phi f \in L^1$ and f belongs to H^1. By the same token, f is in H^p for $n/(n+1) < p \le 1$. Similarly, if we assumed only that $q = 1$, we would still have that $f \in H^p$, when $n/(n+1) < p < 1$.

Finally, to have $f \in H^p$ for $p \le n/(n+1)$ would require further moment conditions on f (see also §5.4 below).

1.3 Proof of Theorem 1. We begin the proof of our theorem by considering some further maximal operators (associated with a given Φ) that provide the crucial tools in what follows. First we define the "non-tangential" version of M_Φ, given by

$$M_\Phi^* f(x) = \sup_{|x-y|<t} |(f * \Phi_t)(y)| = \sup_{|y|<t} |(f * \Phi_t)(x-y)|, \tag{4}$$

and an even larger "tangential" variant M_N^{**} (depending on a parameter N), given by

$$M_N^{**} f(x) = \sup_{y\in\mathbf{R}^n, t>0} |(f * \Phi_t)(x-y)| \left(1 + \frac{|y|}{t}\right)^{-N}. \tag{5}$$

We note the obvious pointwise inequalities

$$M_\Phi f \le M_\Phi^* f \le 2^N M_N^{**} f.$$

LEMMA 1. *If $M_\Phi^* f \in L^p(\mathbf{R}^n)$ and $N > n/p$, then $M_N^{**} f \in L^p(\mathbf{R}^n)$ with*

$$\|M_N^{**} f\|_{L^p} \le C_{N,p} \|M_\Phi^* f\|_{L^p}. \tag{6}$$

Proof. We have already seen that control of the nontangential maximal operator M_Φ^* (defined with respect to the aperture 1) leads to a corresponding control of the nontangential maximal operator defined with respect to any aperture. In fact

$$\int_{\mathbf{R}^n} F_a^*(x)\,dx \le c_n (1+a)^n \int_{\mathbf{R}^n} F^*(x)\,dx \tag{7}$$

if we take $F(x,t) = |(f * \Phi_t)(x)|^p$ and $F_a^*(x) = \sup_{|y|<at} F(x-y,t)$, as in Chapter 2, §2.5.

Now, for $y \in \mathbf{R}^n$, $t > 0$, and $N \ge 0$,

$$|(f * \Phi_t)(x-y)|^p \left(1 + \frac{|y|}{t}\right)^{-Np} \le \sum_{k=0}^{\infty} 2^{(1-k)Np} \cdot F_{2^k}^*(x).$$

Indeed, when $|y| < t$, the term $k = 0$ of the right side already majorizes the left side. Similarly, when $2^{k-1}t < |y| \le 2^k t$, the k^{th} term on the right majorizes the left side. Using (7), we then get (6), with

$$C_{N,p}^p = c_n \sum_{k=0}^{\infty} (1+2^k)^n \cdot 2^{(1-k)Np},$$

which is finite if $Np > n$, proving the lemma.

The next lemma will be used to pass from one approximation of the identity to any other.

LEMMA 2. *Suppose we are given Φ and Ψ in $\mathcal{S}$, with $\int_{\mathbf{R}^n} \Phi\,dx = 1$. Then there is a sequence $\{\eta^{(k)}\}$, $\eta^{(k)} \in \mathcal{S}$, so that*

$$\Psi = \sum_{k=0}^{\infty} \eta^{(k)} * \Phi_{2^{-k}} \tag{8}$$

with $\eta^{(k)} \to 0$ rapidly, in the sense that whenever $\|\cdot\|_{\alpha,\beta}$ is a seminorm, and $M \ge 0$ is fixed, then

$$\|\eta^{(k)}\|_{\alpha,\beta} = O(2^{-kM}) \quad \text{as } k \to \infty.$$

Decompositions of this kind can be obtained either by using the Fourier transform or by dealing directly with the convolutions in question. The former approach, which is somewhat simpler, is the one used in what follows; for the latter, which is also useful in extensions to different settings, see §5.26 below. From the Fourier transform point of view, Lemma 2 can be thought of as an expression of Wiener's principle: The linear span of the set of all translates of a given family of functions is dense in L^1, provided the family satisfies the obvious necessary condition that there are no points at which their Fourier transforms simultaneously vanish. In our case, the family is $\{\Phi_t\}_{t>0}$.

Proof. Fix a $\widehat{\phi} \in C^\infty$ so that $\widehat{\phi}(\xi) = 1$ for $|\xi| \le 1$ and $\widehat{\phi}(\xi) = 0$ for $|\xi| \ge 2$. Then since $1 = \lim_{k\to\infty} \widehat{\phi}(2^{-k}\xi)$, we have that

$$1 = \sum_{k=0}^{\infty} \widehat{\psi}_k(\xi), \tag{9}$$

where $\widehat{\psi}_0(\xi) = \widehat{\phi}(\xi)$, and

$$\widehat{\psi}_k(\xi) = \widehat{\phi}(2^{-k}\xi) - \widehat{\phi}(2^{1-k}\xi) \quad \text{for } k \ge 1.$$

Note that $\widehat{\psi}_k(\xi)$ is supported in $2^{k-1} \le |\xi| \le 2^{k+1}$, $k = 1, 2, \ldots$, and also that $|\partial_\xi^\alpha \widehat{\psi}_k(\xi)| \le c_\alpha 2^{-k\alpha}$.

Because of (9)

$$\widehat{\Psi}(\xi) = \sum_{k=0}^{\infty} \widehat{\psi}_k(\xi)\widehat{\Psi}(\xi).$$

Now the assumption $\int \Phi\, dx = 1$ means that $\widehat{\Phi}(0) = 1$. Let us assume momentarily that $|\widehat{\Phi}(\xi)| \ge 1/2$, for $|\xi| \le 2$. Then we write

$$\widehat{\Psi}(\xi) = \sum_{k=0}^{\infty} \frac{\widehat{\psi}_k(\xi)}{\widehat{\Phi}(2^{-k}\xi)} \widehat{\Psi}(\xi) \cdot \widehat{\Phi}(2^{-k}\xi); \tag{10}$$

and we set

$$\widehat{\eta}^{(k)}(\xi) = \frac{\widehat{\psi}_k(\xi)}{\widehat{\Phi}(2^{-k}\xi)} \widehat{\Psi}(\xi).$$

Roughly speaking, $\eta^{(k)}$ is the part of the given $\Psi \in \mathcal{S}$ whose frequencies are of the order 2^k. It is easily seen that $\eta^{(k)} \in \mathcal{S}$, and that the polynomial decay in the derivatives of $\widehat{\Psi}$ becomes exponential in k, i.e., for any seminorm $\|\cdot\|_{\alpha,\beta}$ and any $M \ge 0$, $\|\eta^{(k)}\|_{\alpha,\beta} = O(2^{-kM})$, as $k \to \infty$.

Finally, the assumption $|\widehat{\Phi}(\xi)| \ge 1/2$ for $|\xi| \le 2$ can be dropped, since we always have $|\widehat{\Phi}(\xi)| \ge 1/2$ for $|\xi| \le 2 \cdot 2^{-k_0}$ when k_0 is sufficiently large; we then need only (in effect) relabel the terms in (10), taking

$$\widehat{\eta}^{(k)}(\xi) = \frac{\widehat{\psi}_{k-k_0}(\xi)}{\widehat{\Phi}(2^{-k}\xi)} \widehat{\Psi}(\xi) \quad \text{for } k \ge k_0,$$

and $\widehat{\eta}^{(k)}(\xi) = 0$ for $k < k_0$. Put another way, we simply dilate $\widehat{\Phi}$ (so that $|\widehat{\Phi}(\xi)| \ge 1/2$ when $|\xi| \le 2$) before beginning the construction.

Let us note that the above argument allows us to observe the following: If we choose a finite collection $\mathcal{F}_0$ of seminorms, and $M \ge 0$ is given, then there exists another (slightly enlarged) finite set $\mathcal{F}$ of seminorms so that whenever $\Psi \in \mathcal{S}_\mathcal{F}$, we have that $\|\eta^{(k)}\|_{\alpha,\beta} \le c2^{-kM}$, whenever $\|\cdot\|_{\alpha,\beta} \in \mathcal{F}_0$. Here c depends on $\mathcal{F}_0$, Φ, ψ, and M, but not on the particular $\Psi \in \mathcal{S}_F$. In deriving $\mathcal{F}$ from $\mathcal{F}_0$, we need to admit extra powers of ξ; this amounts to making Ψ somewhat more smooth than we need the $\eta^{(k)}$ to be.

1.4 We are now in a position to prove that control of the nontangential maximal operator M_Φ^* also allows control of a grand maximal operator $\mathcal{M}_\mathcal{F}$ for appropriate $\mathcal{F}$, and more precisely that the estimate

$$\|\mathcal{M}_\mathcal{F} f\|_{L^p} \le c\|M_\Phi^* f\|_{L^p} \tag{11}$$

holds.

Expanding a given $\Psi \in \mathcal{S}$ by (8) gives

$$(M_\Psi f)(x) = \sup_{t>0} |(f * \Psi_t)(x)| \le \sup_{t>0} \sum_{k=0}^{\infty} |(f * \Phi_{2^{-k}t} * \eta_t^{(k)})(x)|,$$

and by using the definition of M_N^{**} in (5), we have

$$\begin{aligned}(M_\Psi f)(x) &\le \sup_{t>0} \sum_{k=0}^{\infty} \int |(f * \Phi_{2^{-k}t})(x-y)| \cdot t^{-n} |\eta^{(k)}(y/t)|\, dy \\ &\le \sup_{t>0} t^{-n} \sum_{k=0}^{\infty} \int M_N^{**} f(x) \cdot \left(1 + \frac{|y|}{2^{-k}t}\right)^N \cdot |\eta^{(k)}(y/t)|\, dy.\end{aligned} \tag{12}$$

However,

$$t^{-n} \int \left(1 + \frac{2^k|y|}{t}\right)^N \cdot |\eta^{(k)}(y/t)|\, dy = \int (1 + 2^k|y|)^N \cdot |\eta^{(k)}(y)|\, dy \le c2^{-k}$$

if $\|\eta^{(k)}\|_{\alpha,\beta} \le c2^{-k(N+1)}$ for a suitable collection of seminorms and, as we have seen, this happens if Ψ belongs to an appropriately chosen $\mathcal{S}_\mathcal{F}$. Thus

$$\mathcal{M}_\mathcal{F} f(x) = \sup_{\Psi \in \mathcal{S}_\mathcal{F}} M_\Psi f(x) \le c M_N^{**} f(x),$$

for all $x \in \mathbf{R}^n$; taking $N > n/p$ (as in (6)) yields (11).

1.5 The main step. We come now to the turning point in the proof that property (i) (the control of $M_\Phi(f)$ for one $\Phi \in \mathcal{S}$ with $\int \Phi\, dx \ne 0$) implies a corresponding control for a grand maximal function. In view of (11) we need to show that

$$\|M_\Phi^* f\|_p \le c\|M_\Phi f\|_p. \tag{13}$$

We prove (13) by showing that, for any $q > 0$,

$$M_\Phi^* f(x) \le c[M(M_\Phi f)^q(x)]^{1/q},$$

for a substantial set of $x \in \mathbf{R}^n$.† More precisely, we argue as follows. First let us assume that

$$\|M_\Phi^* f\|_{L^p} < \infty;$$

† Here M denotes the usual Hardy-Littlewood maximal operator. We remark that when Φ is the Poisson kernel, this domination occurs for every x; see 5.15 below.

this restriction will be removed below in §1.5.1. Let $\mathcal{F}$ be as in §1.4 and, for any fixed positive number $\lambda > 0$, let

$$F = F_\lambda = \{x : \mathcal{M}_\mathcal{F} f(x) \le \lambda M_\Phi^* f(x)\}.$$

We claim first that

$$\int_{\mathbf{R}^n} (M_\Phi^* f)^p\, dx \le 2 \int_F (M_\Phi^* f)^p\, dx \tag{14}$$

if λ is sufficiently large (independently of f!). To see this, observe that

$$\int_{^cF} (M_\Phi^* f)^p\, dx \le \lambda^{-p} \int_{^cF} (\mathcal{M}_\mathcal{F} f)^p\, dx \le c^p \lambda^{-p} \int_{\mathbf{R}^n} (M_\Phi^* f)^p\, dx,$$

where the last inequality follows from (11). Thus, if we take $\lambda^p \ge 2c^p$, we verify the claim (14) (under the assumption that the left side is finite).

Next we show that, for any $q > 0$,

$$M_\Phi^* f(x) \le c[M(M_\Phi f)^q(x)]^{1/q}, \quad \text{for } x \in F. \tag{15}$$

To see this it will be helpful to use the notation

$$f(x,t) = (f * \Phi_t)(x) \quad \text{and} \quad f^*(x) = M_\Phi^* f(x) = \sup_{|x-y|<t} |f(y,t)|.$$

Now for any x, there exists (y,t) with $|x-y| < t$ and $|f(y,t)| \ge f^*(x)/2$. Choose r small (we shall determine r momentarily) and consider the ball centered at y of radius rt, i.e., the points x' so that $|x'-y| < rt$. We have that

$$|f(x',t) - f(y,t)| \le rt \sup_{|z-y|<rt} |\nabla_z f(z,t)|$$

for $x' \in B(y,rt)$. However,

$$\frac{\partial}{\partial z_i} f(z,t) = \frac{1}{t} f * \Phi_t^i(z), \quad \text{where} \quad \Phi^i = \frac{\partial \Phi}{\partial z_i},$$

and the set of functions of the form $\Phi^i(x+h)$, $|h| \le 1+r$, $i = 1, \ldots, n$, is a compact set in $\mathcal{S}$; hence $\|\Phi^i(x+h)\|_{\alpha,\beta} \le c$ for the seminorms $\|\cdot\|_{\alpha,\beta} \in \mathcal{F}$. Thus

$$|f(x',t) - f(y,t)| \le cr\mathcal{M}_\mathcal{F} f(x) \le cr\lambda M_\Phi^* f(x), \quad \text{if } x \in F.$$

That is, $|f(x',t) - f(y,t)| \le c\lambda r f^*(x)$, while $|f(y,t)| \ge f^*(x)/2$. So if we take r so small that $c\lambda r \le 1/4$, we have

$$|f(x',t)| \ge \frac{1}{4} f^*(x) = \frac{1}{4} M_\Phi^* f(x), \quad \text{for all } x' \in B(y,rt).$$

Taking the q^{th} power and averaging over the ball centered at x of radius $(1+r)t$ gives

$$\begin{aligned} |M_\Phi^* f(x)|^q &\le \left(\frac{1+r}{r}\right)^n \frac{4^q}{|B(x,(1+r)t)|} \int_{B(x,(1+r)t)} |f(x',t)|^q\, dx' \\ &\le cM[(M_\Phi f)^q](x), \end{aligned}$$

which is (15). Finally, using the maximal theorem (for M) with $q < p$ leads to

$$\int_F (M_\Phi^* f(x))^p\, dx \le c \int_{\mathbf{R}^n} (M(M_\Phi^q f)(x))^{p/q}\, dx \le \bar{c} \int_{\mathbf{R}^n} M_\Phi^p f(x)\, dx,$$

which, combined with (14), proves (13).

1.5.1 We return to the point left open above, namely that the finiteness of $\|M_\Phi f\|_{L^p}$ implies that of $\|M_\Phi^* f\|_{L^p}$. Usually the passage from *a priori* inequalities like (13) is straightforward, but here matters are a little tricky.

For each L (large) and ε, $0 < \varepsilon \le 1$, we define a modification of M_Φ^*, given by‡

$$M_\Phi^{\varepsilon,L}(f)(x) = \sup_{|x-y|<t<\varepsilon^{-1}} |f * \Phi_t(y)| \cdot \frac{t^L}{(\varepsilon + t + \varepsilon|y|)^L}.$$

Our first observation is that when f is any distribution and Φ is a given element of $\mathcal{S}$, then since $f * \Phi(x)$ grows at most polynomially in x, there is an L large enough so that, for each $\varepsilon > 0$, $M_\Phi^{\varepsilon,L}(f) \in L^p$. We now fix this L. Next, an examination of the argument in §1.4 shows that it also proves

$$\|\sup_{\Psi \in \mathcal{S}_\mathcal{F}} M_\Psi^{\varepsilon,L}(f)\|_{L^p} \le c_L \|M_\Phi^{\varepsilon,L}(f)\|_{L^p}, \tag{16}$$

where the bound c_L is independent of ε but may depend on L. In fact, if we go back to the key inequality (12), we see that, instead of merely inserting the factor $(1 + 2^k|y|/t)^N$, we must instead insert the factor

$$\frac{t^L \cdot (\varepsilon + 2^{-k}t + \varepsilon|x-y|)^L}{(\varepsilon + t + \varepsilon|x|)^{-L} \cdot (2^{-k}t)^{-L}} \cdot \left(1 + \frac{2^k|y|}{t}\right)^N,$$

which is bounded by

$$c2^{kL} \cdot \left(1 + \frac{|y|}{t}\right)^L \cdot \left(1 + \frac{2^k|y|}{t}\right)^N.$$

‡ Because of the decay factor $\dfrac{t^L}{(\varepsilon + t + \varepsilon|y|)^L}$, $M_\Phi^{\varepsilon,L}$ does *not* commute with translations.

The argument then proceeds as before, proving (16). From this we can also obtain that $\|M_\Phi^{\varepsilon,L}(f)\|_{L^p} \le \bar{c}_L \|M_\Phi(f)\|_{L^p}$; finally, if we let $\varepsilon \to 0$, we get

$$\|M_\Phi^* f\|_{L^p} \le \bar{c}_L \|M_\Phi f\|_{L^p}. \tag{17}$$

This shows that $M_\Phi^* f \in L^p$, if $M_\Phi f \in L^p$. Note, however, that this argument by itself does not give a satisfactory bound, since the L chosen above depends on the f in question.

1.6 We have shown the equivalence of conditions (i) and (ii) of Theorem 1 and now turn to the proof that (ii) implies (iii).

First let Φ be any element of $\mathcal{S}$. Then we know (because of (ii)) that $M_\Phi^* f \in L^p$. Note that

$$\begin{aligned}|f * \Phi_t(x)|^p &\le \min_{|x-y|\le t} (M_\Phi^* f(y))^p \\ &\le \frac{c}{t^n} \int_{|x-y|\le t} (M_\Phi^* f(y))^p \, dy \le c t^{-n} \|M_\Phi^* f\|_{L^p}.\end{aligned}$$

Thus $f * \Phi = f * \Phi_1$ is bounded, and since this is true for every $\Phi \in \mathcal{S}$, we have that f is a bounded distribution, according to our previous definition.

Next we claim that, if P is the Poisson kernel, then

$$P(x) = c_n (1 + |x|^2)^{-(n+1)/2} = \sum_{k=0}^{\infty} 2^{-k} \Phi_{2^k}^{(k)}(x), \tag{18}$$

where $\{\Phi^{(k)}\}$ is a bounded collection of functions in $\mathcal{S}$. This, together with the fact that $M_{\Phi^{(k)}}^* \le c\mathcal{M}_{\mathcal{F}}$, with c independent of k, and the observation that $M_{\Phi_a}^*(f)(x) \le M_\Phi^*(f)(x)$ whenever $a \ge 1$, shows that $u^*(x) \in L^p$. To see (18), fix a $\phi \in C_0^\infty$, with $\phi(x) = 1$ for $|x| \le 1/2$, and $\phi(x) = 0$ for $|x| \ge 1$, and write

$$P(x) = \phi(x)P(x) + \sum_{k=1}^{\infty} [\phi(2^{-k-1}x) - \phi(2^{-k}x)]P(x),$$

which gives (18) with $\Phi^{(0)}(x) = \phi(x)P(x)$, and

$$\Phi^{(k)}(x) = c_n[\phi(x/2) - \phi(x)](2^{-2k} + |x|^2)^{-(n+1)/2}, \quad k \ge 1.$$

But since $\phi(x/2) - \phi(x)$ is supported where $1/2 \le |x| \le 2$, the fact that $\{\Phi^{(k)}\}$ is a bounded collection is easily verified, and thus the proof that (ii) implies (iii) is also established.

1.7 Conclusion of proof. Finally, we show that condition (iii) of Theorem 1 implies condition (i) by showing that a suitable integral of Poisson kernels gives a nontrivial test function. The idea is to use the existence of a function η defined on $(1, \infty)$ that is rapidly decreasing at ∞ and satisfies the moment conditions[†]

$$\int_1^\infty \eta(s)\, ds = 1, \quad \text{and} \quad \int_1^\infty s^k \eta(s)\, ds = 0, \quad k = 1, 2, \ldots. \tag{19}$$

Let us set

$$\Phi(x) = \int_1^\infty \eta(s) P_s(x)\, ds. \tag{20}$$

Since

$$(1 + t^2)^{-(n+1)/2} = \sum_{k<R} a_k t^k + O(t^R), \quad 0 \le t < \infty$$

for appropriate binomial coefficients a_k, we have that

$$\begin{aligned}P_s(x) &= \frac{c_n s}{(s^2 + |x|^2)^{(n+1)/2}} \\ &= \sum_{k<R} c_n a_k s |x|^{1-n} \left(\frac{s}{|x|}\right)^k + O(s^{R+1}|x|^{-n-1-R}).\end{aligned}$$

Inserting this in (20) shows that Φ is rapidly decreasing; a similar argument works for any derivative of Φ. Thus $\Phi \in \mathcal{S}$. Clearly

$$\int_{\mathbf{R}^n} \Phi(x)\, dx = \int_1^\infty \eta(s)\, ds = 1.$$

Also

$$M_\Phi f(x) \le \sup_{0<s<\infty} |u(x,s)| \cdot \int_1^\infty |\eta(s)|\, ds,$$

with $u(x,s) = (f * P_t)(x)$. Hence $u^* \in L^p$ implies that condition (i) holds, and the proof of Theorem 1 is complete.[‡]

1.8 Further remarks.

1. It is clear from the proof of the theorem that the finite set of seminorms $\mathcal{F}$ (for the grand maximal function appearing in (ii)) can be chosen independent of f, and to depend only on p (and n). More precisely, suppose N is a positive integer, and let

$$\mathcal{F} = \mathcal{F}_N = \{\|\cdot\|_{\alpha,\beta} : |\alpha| \le N, |\beta| \le N\}.$$

[†] An explicit example is $\eta(s) = (1/\pi s)\mathrm{Im}\{\exp[1 - w(s-1)^{1/4}]\}$, with $w = e^{-i\pi/4}$. See *Singular Integrals*, p. 183.

[‡] Notice that we used only the fact that the "vertical" maximal function $u^+(x) = \sup_t |u(x,t)|$ was in L^p. Thus we have also shown that $u^+ \in L^p$ implies that $u^* \in L^p$.

If N is sufficiently large (in terms of p), the quantities

$$\|M_\Phi f\|_{L^p}, \qquad \|\mathcal{M}_{\mathcal{F}_N} f\|_{L^p}, \qquad \text{and} \quad \|u^*\|_{L^p}$$

are mutually comparable, with bounds independent of f. Any one of these can be taken to be the "norm" on the space $H^p(\mathbf{R}^n)$; of course, this is an actual norm only when $p \geq 1$. We shall nevertheless continue to refer to quantities like $\|M_\Phi f\|_{L^p}$ as the H^p norm of f, and sometimes write them as $\|f\|_{H^p}$. Notice that the p^{th} power of the "norm" $\|\cdot\|_{H^p}^p$ is subadditive when $p \leq 1$ and hence gives a metric on H^p. This metric defines the topology of H^p; further details are in §5.1 below.

Observe also that convergence in the H^p norm is stronger than convergence in the sense of distributions. This follows easily by the same argument that shows that every element of H^p is a bounded distribution; see also remark 4 below.

2. The "least" regularity required for Φ, and the corresponding smallest set of seminorms $\mathcal{F}$, as they depend on p, is suggested in §5.9 and §5.10 below.

3. The significance of the grand maximal function is seen in the fact that it immediately gives us control of suitable smooth averages of our distribution. A precise statement of this is as follows. Let us set $\mathcal{M}f = \mathcal{M}_{\mathcal{F}_N} f$, where (for instance) $\mathcal{F}_N$ is the collection of seminorms appearing in the first remark above, and N is fixed. Then whenever ϕ is a bump function supported in a ball B that is suitably normalized, we have

$$\langle f, \phi\rangle \leq c\mathcal{M}f(\bar{x}), \quad \bar{x} \in B. \tag{21}$$

The requirements on ϕ, besides that it be supported in B, are the inequalities

$$|\partial^\alpha \phi| \leq r^{-n-|\alpha|}, \quad \text{for } |\alpha| \leq N,$$

where r is the radius of B.

That (21) holds follows easily from the definition of $\mathcal{M}$. A similar result holds for the grand maximal function defined in terms of an arbitrary finite collection $\mathcal{F}$ of seminorms, because $\mathcal{F} \subset \mathcal{F}_N$ when N is sufficiently large.

4. It is useful to observe that whenever $f \in H^p$ and $\Phi \in \mathcal{S}$, then the C^∞ function $f * \Phi$ belongs to L^r, for every $r \geq p$. Since $|f * \Phi(x)| \leq M_\Phi^*(x)$, it is immediate that $f*\Phi \in L^p$. In addition, by the very definition of M_Φ^*, we have

$$|(f*\Phi)(x)| \leq (M_\Phi^* f)(y), \quad \text{for every } y \text{ with } |y-x| \leq 1,$$

and hence

$$|f*\Phi(x)|^p \leq \frac{1}{|B_x|}\int_{B_x} (M_\Phi^* f(y))^p\,dy,$$

where B_x is the ball of radius 1 centered at x.

Therefore

$$|(f*\Phi)(x)|^p \leq c\int (M_\Phi^* f)^p\,dy = c\|f\|_{H^p}^p,$$

so $f*\Phi \in L^\infty$. Since $f*\Phi \in L^p \cap L^\infty$, it follows that $f*\Phi \in L^r$.

5. Since we have seen that $H^p = L^p$ when $p > 1$, for the remainder of this chapter our emphasis will be on H^p for $p \leq 1$.

2. *Atomic decomposition for H^p*

We now turn to an equivalent way of looking at H^p for $p \leq 1$. Specifically, we shall decompose elements in H^p as sums of atoms, as was described (for $p=1$) in §1.2.3 above. The key tool that makes this possible is a variant of the Calderón-Zygmund decomposition; the earlier version (in Chapter 1, §4) was, as we have already seen, indispensable to the theory of singular integrals and maximal functions developed in the previous chapters.

2.1 A variant of the Calderón-Zygmund decomposition. Here the role of the standard maximal operator will be played by the grand maximal operator $\mathcal{M} = \mathcal{M}_\mathcal{F}$, where $\mathcal{F}$ is a finite set of seminorms, fixed throughout this discussion. It will also be necessary to have at our disposal a less demanding maximal operator — one that is based on a single fixed approximation to the identity. So we set

$$(\mathcal{M}_0 f)(x) = (M_\Phi f)(x) = \sup_{t>0} |(f*\Phi_t)(x)|,$$

where Φ is a fixed C^∞ function, supported in the unit ball, with $\int \Phi \neq 0$.†

PROPOSITION. *Suppose that $f \in L^1_{\text{loc}}$ with $\mathcal{M}(f) \in L^p$, $0 < p < \infty$, and α is a positive number. Then there is a decomposition $f = g + b$, $b = \sum b_k$, and a collection of cubes $\{Q_k^*\}$, so that*

(i) *The function g is bounded, with $|g(x)| \leq c\alpha$ a.e.*

(ii) *Each function b_k is supported in Q_k^*,*

$$\int_{\mathbf{R}^n} (\mathcal{M}_0 b_k)^p\,dx \leq c\int_{Q_k^*} (\mathcal{M}f)^p\,dx,$$

and $\int b_k\,dx = 0$.

(iii) *The $\{Q_k^*\}$ have the bounded intersection property, and*

$$\bigcup_k Q_k^* = \{x : \mathcal{M}f(x) > \alpha\}.$$

† Our new notation reflects the fact that the specifics of $\mathcal{F}$ and Φ are no longer central to our arguments.

Remark. Although we assume the local integrability of f, for later convenience we present some of the proof using the language of distributions. We have

$$\langle h, \phi\rangle = \int h(x)\,\phi(x)\,dx$$

whenever $h \in L^1_{\text{loc}} \subset \mathcal{S}'$ and $\phi \in \mathcal{S}$.

2.1.1 We let O denote the open set $O = \{x : \mathcal{M}f(x) > \alpha\}$, and we shall need a Whitney decomposition of it (see Chapter 1, §3.2). Since here we are in the setting of $\mathbf{R}^n$ with its usual structure, it is convenient (but not really necessary) to use the usual Whitney decomposition of O into disjoint cubes. Thus we can find closed cubes Q_k whose interiors are mutually disjoint, and whose side lengths are comparable to their distance from ${}^cO$, with $O = \cup_k Q_k$. Next, fix $\tilde{a}$ and a^* with $1 < \tilde{a} < a^*$; if $\tilde{Q}_k = \tilde{a}Q_k$, $Q_k^* = a^*Q_k$,[‡] then $Q_k \subset \tilde{Q}_k \subset Q_k^*$. Also, if we take a^* sufficiently close to 1, $\cup Q_k^* = O$, and the $\{Q_k^*\}$ have the bounded intersection property: every point is contained in at most a fixed number of the $\{Q_k^*\}$. Hence, conclusion (iii) is satisfied.

2.1.2 Fix a positive smooth function ζ that equals 1 in the cube of side length 1 centered at the origin and vanishes outside the concentric cube of side length $\tilde{a}$. We set $\zeta_k(x) = \zeta([x-x_k]/\ell_k)$, where x_k is the center of the cube Q_k and ℓ_k is its side length. Write $\eta_k = \zeta_k/(\sum_j \zeta_j)$. The η_k form a partition of unity for the set O subordinate to the locally finite cover $\{\tilde{Q}_k\}$ of O; that is to say, $\chi_O = \sum \eta_k$ with each η_k supported in the cube $\tilde{Q}_k$. Notice that the η_k satisfy the natural differential inequalities

$$|\partial^\beta \eta_k(x)| \le c_\beta \ell_k^{-|\beta|}$$

and that $\int \eta_k\,dx \simeq |Q_k^*| \simeq \ell_k^n$. For convenience, we also set

$$\tilde{\eta}_k = \eta_k/(\textstyle\int \eta_k\,dx)$$

We now define b_k by $b_k = (f - c_k)\eta_k$, where the constants c_k are determined by the requirement that $\int b_k\,dx = 0$, i.e., $c_k = \langle f, \tilde{\eta}_k\rangle$. We claim that

$$|c_k| \le c\alpha. \tag{22}$$

In fact, if $\bar{x}$ is a point in ${}^cO$ whose distance from Q_k is comparable to ℓ_k, then (22) follows from (21), if we take $\phi = \tilde{\eta}_k$ and B to be the ball centered at x_k of radius $c\ell_k$, with c so large that $B \supset Q_k^*$ and $\bar{x} \in B$. Similar reasoning (or the fact that $Q_k^* \subset O$) shows that

$$|c_k| \le c\,\mathcal{M}f(x), \quad \text{for any } x \in Q_k^*. \tag{23}$$

Now $g(x) = f(x)$ for $x \notin O$, and hence $|g(x)| \le c\alpha$, because $\mathcal{M}f(x) \le \alpha$ there. Also $g(x) = \sum c_k\eta_k$ for $x \in O$, and so $|g(x)| \le c\alpha$ by (22); therefore conclusion (i) is proved.

[‡] This means that the cubes $\tilde{Q}_k$ and Q_k^* are expansions (by factors $\tilde{a}$ and a^* respectively) of the cube Q_k about its center.

2.1.3 What remains to be done is the proof of the L^p inequalities for $\mathcal{M}_0(b_k)$, and we first treat the case when $p > n/(n+1)$. Our aim is to obtain the following two estimates:

$$(\mathcal{M}_0 b_k)(x) \le c\,\mathcal{M}f(x) \quad \text{if } x \in Q_k^*, \tag{24}$$

and

$$(\mathcal{M}_0 b_k)(x) \le c\alpha \frac{\ell_k^{n+1}}{|x - x_k|^{n+1}} \quad \text{if } x \notin Q_k^*. \tag{25}$$

Let $x \in Q_k^*$ and write $(f\eta_k * \Phi_t)(x) = \int f(y)\eta_k(y)\Phi_t(x-y)\,dy$. If $t \le \ell_k$, since Φ is supported in the unit ball, we get $(f\eta_k * \Phi_t)(x) \le c\mathcal{M}f(x)$ by taking $\phi(y) = \eta_k(y)\Phi_t(x-y)$, B to be the ball of radius t centered at x, and $\bar{x} = x$, in the generalized mean-value inequality (21). When $t > \ell_k$, we instead take the radius of the ball B to be $c\ell_k$ with c large enough so that $B \supset Q_k^*$. This shows that

$$\mathcal{M}_0(f\eta_k)(x) = \sup_{t>0} |(f\eta_k * \Phi_t)(x)| \le c\mathcal{M}f(x), \quad x \in Q_k^*.$$

Since $\mathcal{M}_0(c_k\eta_k)(x) \le c\mathcal{M}f(x)$ by (23), we have proved the estimate (24).

To see (25), we use that $\int b_k\,dx = 0$ and write

$$\int b_k(y)\,\Phi_t(x-y)\,dy = \int b_k(y)[\Phi_t(x-y) - \Phi_t(x-x_k)]\,dy = I_1 - I_2;$$

here I_1 is like the previous integral, except b_k is replaced by $f\eta_k$; in I_2, b_k is replaced by $c_k\eta_k$.

Recall that η_k is supported in the cube $\tilde{Q}_k$, and we have taken $\tilde{Q}_k$ to be strictly contained in Q_k^*. Thus if $x \notin Q_k^*$ and $\eta_k(y) \ne 0$, we have $|x-y| \simeq |x - x_k| \ge c\ell_k$, and the support property of Φ requires that $t \ge c|x - x_k|$. Now set $\phi(y) = \eta_k(y)[\Phi_t(x-y) - \Phi_t(x-x_k)]$ and observe that, if the variables are restricted as above, then

$$|\partial^\beta \phi(y)| \le A_\beta \frac{\ell_k}{|x - x_k|^{n+1}} \cdot \ell_k^{-|\beta|}. \tag{26}$$

Thus we may invoke the inequality (21) again, here with B a ball centered at x_k, of radius $c\ell_k$, that reaches out to ${}^cO$. So

$$|I_1| \le c\alpha \frac{\ell_k^{n+1}}{|x - x_k|^{n+1}}, \quad x \notin Q_k^*.$$

A cruder argument, using (26) with $\beta = 0$ and (22), gives a similar estimate for $|I_2|$; combining these two yields (25). Now

$$\int (\mathcal{M}_0 b_k)^p\,dx = \int_{Q_k^*} (\mathcal{M}_0 b_k)^p\,dx + \int_{{}^cQ_k^*} (\mathcal{M}_0 b_k)^p\,dx.$$

For the first integral we use (24), and for the second we use (25). The second is then dominated by

$$c\alpha^p \int_{|x-x_k|\geq c\ell_k} \left(\frac{\ell_k}{|x-x_k|}\right)^{(n+1)p} dx = c'\alpha^p \ell_k^n = c\alpha^p |Q_k^*|,$$

provided $(n+1)p > n$, as we have assumed. But $\alpha^p |Q_k^*| \leq \int_{Q_k^*} (\mathcal{M}f)^p\, dx$, since $\mathcal{M}f(x) \geq \alpha$ for $x \in Q_k^*$. This proves (ii) and completes the proof of the proposition when $p > n/(n+1)$.

2.1.4 The proof for small values of p needs an improvement of the estimate (25) at infinity, and this comes about by modifying the construction of the "bad" functions b_k so that they satisfy additional moment properties.

Let us fix a nonnegative integer d and write $\mathcal{P}_d$ for the finite-dimensional vector space of polynomials on $\mathbf{R}^n$ having degree at most d. We replace the constants c_k used above by polynomials $c_k(x) \in \mathcal{P}_d$ specified by the condition that

$$\langle f - c_k, q\eta_k \rangle = 0, \qquad \text{for all } q \in \mathcal{P}_d. \tag{27}$$

This choice of c_k can be effected as follows. Let $\mathcal{H}$ be the Hilbert space $L^2(Q_k^*, \tilde{\eta}_k\, dx)$ and let $\mathcal{H}_d$ be $\mathcal{P}_d$ considered as a subspace of $\mathcal{H}$. Let $e_1, \ldots, e_N$ be polynomials forming an orthonormal basis of $\mathcal{H}_d$; we have

$$\langle e_i, e_j \tilde{\eta}_k \rangle = \delta_{ij}.$$

Finally, let $P_k : \mathcal{S}' \to \mathcal{H}_d$ be the projection operator

$$(P_k F)(x) = \sum_{j=1}^{N} \langle F, \bar{e}_j \tilde{\eta}_k \rangle \cdot e_j(x);$$

note that P_k is an extension of the usual projection $\mathcal{H} \to \mathcal{H}_d$. We take $c_k = P_k(f)$ and (27) is immediate.

Now the estimates that replace (22) and (23) are

$$|c_k \eta_k| \leq c\alpha \tag{22'}$$

and

$$|c_k(x)\eta_k(x)| \leq c\mathcal{M}f(x), \quad \text{for any } x \in Q_k^*. \tag{23'}$$

To prove these inequalities we first observe that

$$\sup_{x \in Q_k^*} |\partial^\beta q(x)| \leq A_\beta \ell_k^{-|\beta|} \|q\|_{\mathcal{H}}, \qquad q \in \mathcal{P}_d. \tag{28}$$

Indeed, it suffices to prove a similar estimate with $\|q\|_{\mathcal{H}}$ replaced by $(\int_{Q_k} |q|^2/|Q_k|)^{1/2}$. The resulting inequalities are actually translation and dilation invariant (as is the space $\mathcal{P}_d$) so we need only verify (28) in the case when Q_k is the unit cube centered at the origin.

Applying (28) to the e_j, $1 \leq j \leq N$, allows us to establish (22′) and (23′) for $c_k = P_k(f)$ by the same arguments that were used to prove (22) and (23).

With this achieved, we set $b_k = (f - c_k)\eta_k$ and proceed as before. Estimate (24) follows from (22′) and (23′). In the treatment of $(\mathcal{M}_0 b_k)(x)$, $x \notin Q_k^*$, we use that now

$$\langle b_k, q \rangle = 0 \qquad \text{whenever } q \in \mathcal{P}_d$$

and replace $\Phi_t(x-y)$ by $\Phi_t(x-y) - q(y)$, where $q(y)$ is the degree d Taylor polynomial of $y \mapsto \Phi_t(x-y)$, centered at $y = x_k$. Writing $\phi(y) = \eta_k(y)[\Phi_t(x-y) - q(y)]$, we have

$$|\partial^\beta \phi(y)| \leq A_\beta \frac{\ell_k^{d+1-|\beta|}}{|x-x_k|^{n+d+1}}.$$

The result is then

$$(\mathcal{M}_0 b_k)(x) \leq c\alpha \frac{\ell_k^{n+d+1}}{|x-x_k|^{n+d+1}}, \quad \text{if } x \notin Q_k^*. \tag{25'}$$

This gives the desired L^p inequality for $\mathcal{M}_0(b_k)$ if we choose d so large that $(n+d+1)p > n$. The proof of the proposition is therefore concluded.

2.1.5 A corollary. We restate here for emphasis the additional cancellation properties that entered in the proof of the proposition.

COROLLARY. *Given any positive integer d, we can arrange the decomposition in the above proposition so that the functions b_k satisfy the further moment properties*

$$\int b_k q\, dx = 0,$$

for all polynomials q of degree at most d.

2.2 Atomic decomposition. We now come to the atomic decomposition of H^p. An H^p *atom*, $p \leq 1$, is a function a so that

(i) a is supported in a ball B,

(ii) $|a| \leq |B|^{-1/p}$ almost everywhere, and

(iii) $\int x^\beta a(x)\, dx = 0$ for all β with $|\beta| \leq n(p^{-1} - 1)$.

Notice that (i) and (ii) guarantee that $\int |a(x)|^p\,dx \le 1$. Condition (iii) will guarantee the stronger, and for us crucial, inequality

$$\int (\mathcal{M}_0 a(x))^p\,dx \le c, \tag{29}$$

where $\mathcal{M}_0$ is an appropriate maximal operator. Here we verify (29) for $\mathcal{M}_0 f(x) = \sup_{t>0} |f * \Phi_t(x)|$, where Φ is a smooth function supported in the unit ball about the origin with $\int \Phi \neq 0$.

Note first that (by (ii)) $\mathcal{M}_0 a(x) \le c|B|^{-1/p}$. This estimate will be used for $x \in B^*$, where B^* is the ball concentric with B having twice the radius. To deal with $x \notin B^*$ one uses (iii) and writes

$$(a * \Phi_t)(x) = \int a(y)\,\Phi_t(x-y)\,dy = \int a(y)[\Phi_t(x-y) - q_{x,t}(y)]\,dy,$$

where $q_{x,t}(y)$ is the degree d Taylor polynomial of the function $y \mapsto \Phi_t(x-y)$ expanded about $\bar{x}$, the center of B; here d is the smallest integer with $d > n(p^{-1}-1) - 1$ (i.e., $d = \lfloor n(p^{-1}-1)\rfloor$).

By the usual estimate of the remainder term in a Taylor expansion,

$$|\Phi_t(x-y) - q_{x,t}(y)| \le c\frac{|y-\bar{x}|^{d+1}}{t^{n+d+1}}.$$

Since $y \in B$, $x \notin B^*$, and $\Phi(z) = 0$ when $|z| \ge 1$, we have that $t \ge c|x - \bar{x}|$. Hence (letting r be the radius of B),

$$\mathcal{M}_0 a(x) \le c|B|^{-1/p}\left(\frac{r}{|x-\bar{x}|}\right)^{n+d+1}, \qquad x \notin B^*,$$

which proves (29), since $(n+d+1)p > n$.

Thus we have observed that an H^p atom is in fact an element of H^p, and that the H^p "norms" of such atoms are uniformly bounded (in particular, they are independent of the balls B).

We next claim that if $\{a_k\}$ is a collection of H^p atoms and $\{\lambda_k\}$ is a sequence of complex numbers with $\sum |\lambda_k|^p < \infty$, then the series

$$f = \sum_k \lambda_k a_k \tag{30}$$

converges in the sense of distributions, and its sum f belongs to H^p, with $\|f\|_{H^p} \le c(\sum |\lambda_k|^p)^{1/p}$.

Indeed, when the sum (30) is finite, we have

$$\mathcal{M}_0(f) = \mathcal{M}_0(\sum \lambda_k a_k) \le \sum |\lambda_k| \mathcal{M}_0(a_k),$$

while

$$\left(\sum |\lambda_k| \mathcal{M}_0(a_k)\right)^p \le \sum |\lambda_k|^p \mathcal{M}_0(a_k)^p,$$

since $p \le 1$. Integrating these inequalities yields

$$\int [(\mathcal{M}_0 f)(x)]^p\,dx \le \sum |\lambda_k|^p \int [(\mathcal{M}_0 a_k)(x)]^p\,dx \le c\sum_k |\lambda_k|^p,$$

which gives the unconditional convergence† of (30) in the H^p metric. By remark 1 in §1.8, we also have convergence in the sense of distributions.

Our main result is the converse.

THEOREM 2. *Let $p \le 1$. Then every $f \in H^p$ can be written as a sum of H^p atoms, as in* (30) *above, that converges in H^p norm; moreover*

$$\sum_k |\lambda_k|^p \le c\|f\|_{H^p}^p.$$

2.3 Proof. We prove the theorem first under the additional assumption that our distribution f is a locally integrable function. For each integer j, we consider the generalized Calderón-Zygmund decomposition given above in §2.1, where the altitude α is 2^j. It will be useful to adopt the notation that all the elements of that decomposition will carry the superscript j. Thus we write $f = g^j + b^j$, with $b^j = \sum_k b_k^j$; the b_k^j are given by $b_k^j = (f - c_k^j)\eta_k^j$ and are supported in cubes Q_k^{j*}, with $\cup_k Q_k^{j*} = O^j$; finally $O^j = \{x : \mathcal{M}f(x) > 2^j\}$, and $O^{j+1} \subset O^j$.

Observe first that $g^j \to f$ in H^p (and hence also in the sense of distributions) as $j \to \infty$. Indeed, $f - g^j = b^j$, and by conclusions (ii) and (iii) of the proposition in §2.1, we have that

$$\begin{aligned}\|b^j\|_{H^p}^p \simeq \int (\mathcal{M}_0 b^j)^p\,dx &\le \sum_k \int (\mathcal{M}_0 b_k^j)^p\,dx \\ &\le c\int_{\cup Q_k^{j*}} (\mathcal{M}f)^p\,dx = c\int_{\mathcal{M}f > 2^j} (\mathcal{M}f)^p\,dx \to 0.\end{aligned}$$

Next, $|g^j| \le c2^j$ (by conclusion (i) of the proposition), thus $g^j \to 0$ uniformly as $j \to -\infty$. Hence

$$f = \lim_{N\to\infty} \sum_{j=-\infty}^{N} (g^{j+1} - g^j) \tag{32}$$

in the sense of distributions.

† A series is said to be *unconditionally convergent* if its sum is invariant under reordering of the summands.

Notice that $|g^{j+1} - g^j| \le c2^j$, while $g^{j+1} - g^j = b^j - b^{j+1}$, and therefore $g^{j+1} - g^j$ is supported in O^j. So we can rewrite (32) as[‡]

$$f = \sum_j \sum_k (g^{j+1} - g^j)\eta_k^j. \tag{33}$$

Set $\lambda_{j,k} = c2^j |B_k^j|^{1/p}$, and let $a_{j,k} = \lambda_{j,k}^{-1} \cdot (g^{j+1} - g^j)\eta_k^j$; here B_k^j is the smallest ball containing Q_k^{j*}. It is clear from this that $a_{j,k}$ is supported in B_k^j and that $|a_{j,k}| \le |B_k^j|^{-1/p}$ there. Also it is easily seen (see (38) below) that $\sum_{j,k} |\lambda_{j,k}|^p \le c\|f\|_{H^p}^p$. So $f = \sum_{j,k} \lambda_{j,k} a_{j,k}$ would be an atomic decomposition, except that the putative atoms $a_{j,k}$ do not satisfy the crucial cancellation property. To ensure this, we need a more refined decomposition than (33), and to that we now turn.

2.3.1 Recall the projections $P_k = P_k^j$ that were used in §2.1.4. We had $c_k^j = P_k^j(f)$, and, of course, $c_\ell^{j+1} = P_\ell^{j+1}(f)$. Now let $c_{k,\ell}$ denote the polynomial of degree $\le d$ given by

$$c_{k,\ell} = P_\ell^{j+1}[(f - c_\ell^{j+1})\eta_k^j]. \tag{34}$$

There are two things we need to know about the polynomials $c_{k,\ell}$. First, $c_{k,\ell} \ne 0$ only if $Q_k^{j*} \cap Q_\ell^{j+1*} \ne \emptyset$; this follows directly from the definition of P_ℓ^{j+1} (since it involves η_ℓ^{j+1}, which is supported in Q_ℓ^{j+1*}). Observe that if $Q_k^{j*} \cap Q_\ell^{j+1*} \ne \emptyset$, then $\operatorname{diam}(Q_k^{j*}) \ge c \cdot \operatorname{diam}(Q_\ell^{j+1*})$, because $O^{j+1} \subset O^j$. The second fact is that

$$|c_{k,\ell}\,\eta_\ell^{j+1}| \le c2^j. \tag{35}$$

This follows as did the corresponding inequalities (22) and (22′) for c_k, using the support properties of η_k^j just discussed.

We return to the decomposition (32), and write

$$g^{j+1} - g^j = b^j - b^{j+1} = \sum_k (f - c_k^j)\eta_k^j - \sum_\ell (f - c_\ell^{j+1})\eta_\ell^{j+1}.$$

This allows us to set

$$g^{j+1} - g^j = \sum_k A_k^j, \tag{36}$$

where

$$A_k^j = (f - c_k^j)\eta_k^j - \sum_\ell (f - c_\ell^{j+1})\eta_\ell^{j+1}\eta_k^j + \sum_\ell c_{k,\ell}\,\eta_\ell^{j+1}. \tag{37}$$

The reason for (36) is, first, that $\sum_k \eta_k^j = 1$ on the support of η_ℓ^{j+1} ($\subset O^{j+1} \subset O^j$). Next $\sum_k c_{k,\ell} = 0$, since by definition (34), this sum is the projection P_ℓ^{j+1} of $f - c_\ell^{j+1}$, which is zero.

[‡] This expansion is in the spirit of the primitive atomic decomposition for the nontangential space $\mathcal{N}$ given in §2.6 of Chapter 2.

We have therefore proved the identity (36), (37), and we now examine each term A_k^j. We note that:

(i) A_k^j is supported in a ball B_k^j that contains Q_k^{j*} as well as all the Q_ℓ^{j+1*} that intersect Q_k^{j*}. By the remarks after (34), we see that we can take B_k^j so that $|B_k^j| = c|Q_k^j|$.

(ii) $|A_k^j| \le c2^j$. This is because the part of (37) that involves f directly is actually $f\chi_{{}^cO^{j+1}} \cdot \eta_k^j$, and $|f| \le c2^j$ there. For the other parts of (37), one uses the bounds (35) and (22′).

(iii) A_k^j satisfies the moment conditions required for an atom. This is clear for $b_k^j = (f - c_k^j)\eta_k^j$; for $(f - c_\ell^{j+1})\eta_\ell^{j+1}\eta_k^j - c_{k,\ell}\,\eta_\ell^{j+1}$, it is exactly the definition of the polynomials $c_{k,\ell}$.

With $f = \sum_{j,k} A_k^j$ as above, we set $a_k^j = \lambda_{j,k}^{-1} A_k^j$ with $\lambda_{j,k} = c2^j |B_k^j|^{1/p}$. Then $f = \sum_{j,k} \lambda_{j,k} a_k^j$. Note that our assertions imply that the a_k^j are atoms. Also,

$$\begin{aligned}\sum_{j,k} |\lambda_{j,k}|^p &= c\sum_{j,k} 2^{jp}|B_k^j| = c'\sum_{j,k} 2^{jp}|Q_k^j| \\ &= c'\sum_j 2^{jp}|\{\mathcal{M}f > 2^j\}| \le c\int (\mathcal{M}f)^p\,dx \simeq \|f\|_{H^p}^p.\end{aligned} \tag{38}$$

This shows that f has the required atomic decomposition, under the assumption that $f \in H^p \cap L^1_{\text{loc}}$.

2.3.2 To conclude the proof of Theorem 2 it will suffice to prove the following.

Lemma. *$H^p \cap L^1_{\text{loc}}$ is dense in H^p, $0 < p < \infty$.*

Indeed, assuming the lemma, given any $f \in H^p$, we can find a sequence f_m of locally integrable H^p functions so that $f_0 = 0$, $f_m \to f$ in H^p as $m \to \infty$, and $\|f_{m+1} - f_m\|_{H^p}^p \le 2^{-m-1}\|f\|_{H^p}^p$. Thus $f = \sum_{m=0}^{\infty} (f_{m+1} - f_m)$ unconditionally in H^p and in the sense of distributions. Applying the atomic decomposition to each $f_{m+1} - f_m$, we have

$$f = \sum_{m,j,k} \lambda_{j,k}^m a_{j,k}^m,$$

with

$$\sum_{m,j,k} |\lambda_{j,k}^m|^p \le c\sum_m \|f_{m+1} - f_m\|_{H^p}^p \le c'\|f\|_{H^p}^p,$$

giving an atomic decomposition for f.

Turning to the proof of the lemma, we remark that it is already implicit in the variant of the Calderón-Zygmund decomposition presented in §2.1 above.† The idea is that if we decompose a given $f \in H^p$ as $f = g + b$, then

$$\|b\|_{H^p}^p \simeq \|\mathcal{M}_0 b\|_{L^p}^p \le \sum_k \|\mathcal{M}_0 b_k\|_{L^p}^p \le c \int_{\{\mathcal{M}f(x)>\alpha\}} (\mathcal{M}f)^p\,dx,$$

by (ii) and (iii) of that proposition. Thus we would have $\|b\|_{H^p} \to 0$ as $\alpha \to \infty$, while $g \in L^1_{\mathrm{loc}}$, in view of conclusion (i) of the proposition. But there is an obstacle to this approach in that we assumed that f was itself a locally integrable function so as to be able to obtain conclusion (i); by that assumption, it was then legitimate to make separate pointwise estimates for g in the sets ${}^cO$ and O. This argument would not be valid were we dealing with a general distribution f.

The way around this difficulty is to replace conclusion (i) of the proposition by another estimate that is valid without the assumption that $f \in L^1_{\mathrm{loc}}$, to wit:

$$(\mathcal{M}_0 g)(x) \le c\mathcal{M}f(x)\chi_{{}^cO} + c\alpha \sum_k \frac{\ell_k^{n+1}}{(\ell_k + |x - x_k|)^{n+1}} \tag{39}$$

To understand (39), we simply re-examine the proof of the proposition assuming now only that $f \in H^p$; this requires that we replace some of the integrals in that proof with pairings between distributions and test functions. Having done this, one can define the b_k as before: $b_k = (f - c_k)\eta_k$. The b_k are then distributions with compact support. The statement $\int b_k\,dx = 0$ is still meaningful (and valid), and the inequalities (24) and (25) (or (25′)) still hold for $\mathcal{M}_0(b_k)$. These inequalities also give conclusion (ii) of the proposition, and thus $\sum b_k$ converges unconditionally in H^p, and hence as a distribution. With this, the distribution $g = f - b = f - \sum b_k$ is well-defined.

In proving (39) we consider first the case when $x \notin O$. Then since $\mathcal{M}_0(g) \le \mathcal{M}_0(f) + \sum \mathcal{M}_0(b_k)$, the inequality (39) is clear for these x, since $\mathcal{M}_0(f) \le c\mathcal{M}(f)$ and $|x - x_k| \ge c\ell_k$ whenever $x \notin O$ (recall that x_k is the center of the cube Q_k^* and ℓ_k is its side length).

Next fix $x \in O$. Then $x \in Q_m$ for some fixed m. We partition the cubes $\{Q_k^*\}$ into two classes: the "near" cubes, for which $Q_m^* \cap Q_k^* \ne \emptyset$; and the "far" cubes, for which $Q_m^* \cap Q_k^* = \emptyset$. Observe that there are at most a fixed number (independent of m) of "near" cubes.

† For another argument, see §5.14 below.

Now write $g = (f - \sum_{\text{near}} b_k) - \sum_{\text{far}} b_k$, where the first sum is over the near cubes and the second is over the far cubes. For a far cube,

$$(\mathcal{M}_0 b_k)(x) \le c\alpha \frac{\ell_k^{n+1}}{|x - x_k|^{n+1}}$$

by (25), since $x \in Q_m$ and $x \notin Q_k^*$. Since $|x - x_k| \ge c\ell_k$ we have

$$(\mathcal{M}_0 \sum_{\text{far}} b_k)(x) \le c\alpha \sum_{\text{far}} \frac{\ell_k^{n+1}}{(\ell_k + |x - x_k|)^{n+1}}.$$

We next write $f - \sum_{\text{near}} b_k = f - \sum_{\text{near}} f\eta_k - \sum_{\text{near}} c_k\eta_k = h - \sum_{\text{near}} c_k\eta_k$. For the second term, applying (22) and the fact that only a bounded number of summands are involved, we see that

$$(\mathcal{M}_0 \sum_{\text{near}} c_k\eta_k)(x) \le \sum_{\text{near}} |c_k|(\mathcal{M}_0\eta_k)(x) \le c\alpha \frac{\ell_m^{n+1}}{(\ell_m + |x - x_m|)^{n+1}},$$

since $x \in Q_m$ and $\ell_m \ge c|x - x_m|$.

Finally we estimate $(\mathcal{M}_0 h)(x)$ by considering $(h * \Phi_t)(x)$ for $t > 0$. If $t \le c_0 \operatorname{dist}(Q_m, {}^cO)$, where c_0 is an appropriately small constant, then $(h * \Phi_t)(x) = 0$, because $1 - \sum_{\text{near}} \eta_k$ vanishes in Q_m^*, and Φ_t is supported in $B(0,t)$. On the other hand, if $t \ge c_0 \operatorname{dist}(Q_m, {}^cO)$, then we can estimate $(h * \Phi_t)(x)$ as we did in (24). This gives the bound

$$(\mathcal{M}_0 h)(x) \le c\alpha \le c\alpha \frac{\ell_m^{n+1}}{(\ell_m + |x - x_m|)^{n+1}},$$

since $x \in Q_m$.

Altogether then the desired inequality (39) is proved, and with it we get that $\mathcal{M}_0(g) \in L^1$, because

$$\int \mathcal{M}_0(g)\,dx \le \int_{\mathcal{M}f \le \alpha} \mathcal{M}f\,dx + c\alpha \sum_k |\ell_k|^n,$$

and

$$\int_{\mathcal{M}f \le \alpha} \mathcal{M}f\,dx \le \alpha^{1-p} \int (\mathcal{M}f)^p\,dx,$$

while $\sum |\ell_k|^n = \sum |Q_k| = |\{x : \mathcal{M}f(x) > \alpha\}| < \infty$. Thus we have shown that $g \in H^1$ and that H^1 is dense in H^p, $0 < p < \infty$.

2.3.3 To complete the proof of the lemma, we need only show that $H^1 \subset L^1$, as was mentioned in §1.2.2 above. Taking $g \in H^1$ and $\mathcal{M}_0 = M_\Phi$ as above, we have $g \in \mathcal{S}'$ and $M_\Phi g \in L^1$ and wish to show that $g \in L^1$.

Consider the Banach space of continuous functions on $\mathbf{R}^n$ that vanish at infinity and let $A \supset L^1$ be its dual space of finite Borel measures. Taking h_j to be an appropriate subsequence of $g * \Phi_{1/k}$, there is a $d\mu \in A$ with $h_j \to d\mu$ weakly and hence also as distributions; the pointwise inequality $h_j \leq M_\Phi g$ allows us to conclude that $d\mu = h(x)\,dx$ is absolutely continuous with $g = h \in L^1$ and $\|g\|_{L^1} \leq c\|M_\Phi g\|_{L^1} \simeq \|g\|_{H^1}$.

2.4 Further remarks on atomic decompositions.

2.4.1 The subspace of finite linear combinations of H^p atoms is dense in H^p; this is a direct consequence of the norm convergence of the sum (30). Notice that this subspace consists of all bounded functions with compact support that satisfy the moment conditions (iii) of §2.2. Other useful dense subspaces of H^p are described in §5.2 below.

2.4.2 There is a wide degree of flexibility in the definition of atoms: For the atomic decomposition of an element of H^p, we can always choose atoms with any number of additional vanishing moments. In other words, if d is any integer with $d \geq n(p^{-1} - 1)$, we can assume (by the corollary in §2.1.5) that all moments up to order d of our atoms are zero. However, we may not assume *fewer* moment conditions than were set down in §2.2. For instance, when $p = 1$, if one assumes (i) and (ii) but not the moment condition, then the space generated by these "atoms" is in fact all of L^1. See also §5.6 below.

2.4.3 The size condition (ii) $|a| \leq |B|^{-1/p}$ can be replaced by the weaker condition

$$\left(\frac{1}{|B|}\int_B |a|^q\,dx\right)^{1/q} \leq |B|^{-1/p},$$

with $q > 1$ if $p = 1$, and with $q = 1$ if $p < 1$. In fact, using the properties of the standard maximal operator M, it is not difficult to see that the inequality (29) holds in this more general setting. See §5.7 below.

2.4.4 Examination of the support and size properties of the atoms constructed in the proof of Theorem 2 shows that its conclusion also holds for $L^p = H^p$, $1 < p < \infty$, but the condition $\sum|\lambda_k|^p < \infty$ no longer guarantees that a corresponding sum of atoms is in L^p. A useful variant of the atomic decomposition, valid for $1 < p < \infty$, may be found in Chapter 4, §6.18.

3. Singular integrals

One of the principal interests of H^p theory is that it gives a natural extension of the results for maximal functions and singular integrals (originally developed for L^p, $p > 1$) to the range $p \leq 1$. For the maximal function, this extension was carried out when we studied the maximal characterization of H^p in §1. We turn here to the related generalization for singular integrals. The conclusion one arrives at is that, broadly speaking, the L^p boundedness theorems for singular integrals proved in Chapter 1 for $p > 1$ extend to the context of H^p, for all $p > 0$. We take up here two significant special cases of this principle.

3.1 We restrict ourselves to the singular integrals of §5 and §6 of Chapter 1 that are translation-invariant. In this context the results of that chapter can be restated as follows: We begin with an operator T that is bounded from L^2 to itself and commutes with translations. As mentioned in §6 of Chapter 1, such an operator is representable as a multiplication operator via the Fourier transform: There exists a bounded function m (the "multiplier"), so that

$$\widehat{Tf}(\xi) = m(\xi)\hat{f}(\xi), \quad |m(\xi)| \leq A, \tag{40}$$

for $f \in L^2$. By Plancherel's theorem, the boundedness assumption on m is equivalent with

$$\|Tf\|_{L^2} \leq A\|f\|_{L^2}. \tag{41}$$

Let K denote the distribution with $\hat{K} = m$; we then have

$$Tf = f * K \tag{42}$$

(at least) when $f \in \mathcal{S}$.

We now make the *a priori* assumption that the distribution K is given, away from the origin, by a locally integrable function; we denote this function by $K(x)$. We have that

$$(Tf)(x) = \int K(x-y)f(y)\,dy, \tag{43}$$

in the following sense: If f is in L^2 and has compact support, then (43) holds for x outside the support of f. The second quantitative assumption on T (besides (41)) is

$$\int_{|x|\geq 2|y|} |K(x-y) - K(x)|\,dy \leq A, \quad \text{whenever } y \neq 0. \tag{44}$$

We know (by the results of Chapter 1) that such T, initially defined on $L^2(\mathbf{R}^n)$, are extendable to $L^1(\mathbf{R}^n)$ and satisfy a weak-type inequality there. Hence T is also well-defined on $H^1(\mathbf{R}^n) \subset L^1(\mathbf{R}^n)$.

THEOREM 3. *If T satisfies the assumptions* (40)–(44) *above, then it is bounded from $H^1(\mathbf{R}^n)$ to $L^1(\mathbf{R}^n)$; that is,*

$$\|Tf\|_{L^1(\mathbf{R}^n)} \leq A'\|f\|_{H^1(\mathbf{R}^n)}, \tag{45}$$

where the bound A' depends only on the constant A appearing in (41) *and* (44).

Proof. In view of the atomic decomposition (Theorem 2 in §2.2), it suffices to verify the inequality (45) when f is an arbitrary H^1 atom a. Since T is translation-invariant, we may assume that a is supported in a ball centered at the origin. Let us denote this ball by B and let r be its radius. If B^* is the ball concentric with B having twice its radius, then, by (41),

$$\int_{B^*} |Ta|^2\,dx \leq A^2 \int |a|^2\,dx \leq A^2|B|\cdot|B|^{-2} = A^2|B|^{-1}.$$

So by Schwarz's inequality

$$\int_{B^*} |Ta|\,dx \leq A|B^*|\cdot|B|^{-1} = 2^n A.$$

Next, if $x \notin B^*$, we can use the representation (43); thus

$$Ta(x) = \int K(x-y)a(y)\,dy = \int [K(x-y) - K(x)]a(y)\,dy,$$

since $\int a(y)\,dy = 0$ (property (iii) in §2.2). Now we can invoke the inequality (44), because $|x| \geq 2r$ and $|y| \leq r$. The result is that

$$\int_{^cB^*} |Ta|\,dx \leq A\int |a(y)|\,dy \leq A.$$

Combining this with (46), we see that we have verified (45) for atoms, finishing the proof of the theorem.

The above proof and that of the weak-type inequality (12) in Chapter 1 are strikingly close in spirit; see in particular the remarks in §5.3 of that chapter. It is also noteworthy that, under our assumptions, the operator T is actually bounded from H^1 to H^1. To prove this requires either a further argument (see §5.25 below), or, as we shall see, a strengthened version of the hypothesis (44) for the kernel. The latter approach also has the advantage of working for all H^p, $0 < p \leq 1$.

3.2 We shall be considering operators T, with kernels K, as described above, but instead of (44) we shall make a stronger assumption (of the kind already discussed in Chapter 1, §6.1). Namely, we suppose we are given a fixed positive number $\gamma > 0$, and we assume that $K \in C^{\lfloor\gamma\rfloor}(\mathbf{R}^n\backslash 0)$,† together with the bounds

$$\begin{gathered} |\partial_x^\beta K(x)| \leq A|x|^{-n-|\beta|}, \quad \text{for } |\beta| \leq \lfloor\gamma\rfloor, \text{ and} \\ |\partial_x^\beta K(x-y) - \partial_x^\beta K(x)| \leq A\frac{|y|^{\gamma-\lfloor\gamma\rfloor}}{|x|^{n+\gamma}}, \quad \text{for } |\beta| = \lfloor\gamma\rfloor,\ |x| \geq 2|y|. \end{gathered} \tag{48}$$

Clearly, if K satisfies (48), then it also satisfies (44).

We shall next deal with the definition of $Tf = f * K$ when f is an arbitrary element of H^p, for some $p \leq 1$. As we have said, the distribution K is given by the inverse Fourier transform of the multiplier m. However, the convolution of two distributions is in general not well defined, so we must exploit the special properties of both K and f. To do this, fix a $\phi \in C_0^\infty$ with $\phi(x) = 0$ when $|x| \geq 1$, and $\phi(x) = 1$ when $|x| \leq 1/2$. Write $K_0 = K\phi$, $K_\infty = K(1-\phi)$; we have $K = K_0 + K_\infty$.

Since K_0 has compact support, $f * K_0$ is a well defined distribution. We shall be able to define $f * K_\infty$ by writing $\langle f * K_\infty, \Phi\rangle = \langle f * \tilde{\Phi}, \tilde{K}_\infty\rangle$ for $\Phi \in \mathcal{S}$; here $\tilde{\Phi}(x) = \bar{\Phi}(-x)$, $\tilde{K}_\infty(x) = \bar{K}_\infty(-x)$ (the bars denote complex conjugation). Since $f * \tilde{\Phi} \in L^2$ (see §1.8) and (clearly) $\tilde{K}_\infty \in L^2$, Schwarz's inequality shows that $\langle f * \tilde{\Phi}, \tilde{K}_\infty\rangle$ is expressed by a convergent integral, giving us the definition of $f * K_\infty$ as a distribution.

These considerations also show that if $\{f_k\}$ is a sequence that converges in the H^p norm, then $Tf_k = f_k * K$ converges in the topology of tempered distributions.

Having cleared up the definition of $Tf = f * K$ for $f \in H^p$, we can now state the main boundedness result for these operators.

THEOREM 4. *Suppose K is a distribution such that $\widehat{K}$ is a bounded function, and that away from the origin K satisfies the differential inequalities* (48). *If $p \leq 1$ is such that $\gamma > n(p^{-1} - 1)$, then the operator T given by $Tf = f * K$ is bounded from H^p to itself.*

3.3 In view of the atomic decomposition of H^p (as in §2.2) and the fact that, for $p \leq 1$, the quantity $\|\cdot\|_{H^p}^p$ is subadditive, it suffices to show that

$$\|Ta\|_{H^p} \leq A'\|a\|_{H^p} \tag{49}$$

for every H^p atom a, with A' independent of a.

† Here $\lfloor\gamma\rfloor$ denotes the greatest integer n with $n \leq \gamma$.

3.3.1 The proof of (49) is facilitated by using both the translation and dilation "invariance" of the operator T, as well as that of the space H^p.

We begin with translations and the operator τ_h, defined on functions by

$$(\tau_h f)(x) = f(x-h), \qquad h \in \mathbf{R}^n.$$

Of course, τ_h extends to distributions by duality:

$$(\tau_h f)(\phi) = f(\tau_{-h}\phi)$$

for test functions ϕ. Note that $\tau_h(f * K) = \tau_h(f) * K$, and so the operator $Tf = f * K$ commutes with translations; that is,

$$\tau_h(Tf) = T(\tau_h f). \tag{50}$$

We consider next the dilation operator $f \mapsto f_\delta$, which is given by

$$f_\delta(x) = \delta^{-n} f(x/\delta), \qquad \delta > 0$$

on functions. Its extension to distributions f is determined by the identity

$$f_\delta(\phi) = f(\phi_{\delta^{-1}}) \cdot \delta^{-n}.$$

We should also note that

$$(f * K)_\delta = f_\delta * K_\delta.$$

The identity follows from an obvious change of variables, at least when f and K are, say, bounded functions with compact support. A simple limiting argument then extends it to the appropriate distributions. Thus, convolution operators T enjoy the transformation law

$$(Tf)_\delta = T_\delta(f_\delta), \tag{51}$$

where $T_\delta f = f * K_\delta$.

For us, the crucial observation is the whenever K satisfies our hypotheses above (namely (41) and (48)) then, for each $\delta > 0$, K_δ satisfies the same hypotheses, with the *same* bounds. It is in this sense that our operators are translation and dilation invariant.

Indeed, our assertion regarding (48) is a direct consequence of a change of variables, and reflects the homogeneity of the factors $|x|^{-n-|\beta|}$ and $|y|^{\gamma-\lfloor\gamma\rfloor}/|x|^{n+\gamma}$. The invariance of (41) follows from the fact that $\widehat{K_\delta}(\xi) = \widehat{K}(\delta\xi)$, for all $\delta > 0$.

We note also the corresponding invariance of the space H^p, which can be stated as follows. Whenever $f \in H^p$, then both $\tau_h(f)$ and f_δ are in H^p with

$$\|\tau_h(f)\|_{H^p} = \|f\|_{H^p}, \qquad \|f_\delta\|_{H^p} = \delta^{-n+n/p}\|f\|_{H^p}. \tag{52}$$

In fact, according to the definition of H^p (§1.2 and §1.8), we may take

$$\|f\|_{H^p} = \|M_\Phi f\|_{L^p}$$

for some fixed $\Phi \in \mathcal{S}$. Since M_Φ commutes with translations and dilations, this reduces (52) to its analogue for L^p norms.

3.3.2 The proof of Theorem 4 is based on the following lemma. Suppose that Φ is a fixed test function, supported in the unit ball, with $\int \Phi\, dx = 1$.

LEMMA. *Let K be a distribution that satisfies* (41) *and* (48), *and define $K^{(t)}$ by $K^{(t)} = K * \Phi_t$. Then $K^{(t)}$ satisfies the same hypotheses, with bounds that may be chosen independent of t. In particular*

$$|\partial_x^\beta K^{(t)}(x)| \le A'|x|^{-n-|\beta|}$$

whenever $|\beta| \le \lfloor\gamma\rfloor$, and

$$|\partial_x^\beta K^{(t)}(x-y) - \partial_x^\beta K^{(t)}(x)| \le A'\frac{|y|^{\gamma-\lfloor\gamma\rfloor}}{|x|^{n+\gamma}},$$

whenever $|\beta| = \lfloor\gamma\rfloor$ and $|x| \ge 2|y|$. Here A' does not depend on t (but it may depend on Φ).

To prove the lemma, we exploit the identity $(K*\Phi_t)_\delta = K_\delta * \Phi_{t\delta}$ and replace K by $K^{(t^{-1})}$. This reduces the lemma to the case $t = 1$, provided that we take A' to depend only on the bounds appearing in (41) and (48).

Now with $t = 1$, we consider two cases, x small and x large. If x is small, we write

$$K^{(1)}(x) = \int_{\mathbf{R}^n} e^{-2\pi i x\cdot\xi}\widehat{K}(\xi)\,\widehat{\Phi}(\xi)\,d\xi.$$

Then the assumption $|\widehat{K}(\xi)| \le A$ implies that $K^{(1)}(x)$ and $\partial_x^\beta K^{(1)}(x)$ are uniformly bounded. For large x, it is legitimate to use the realization

$$K^{(1)}(x) = \int_{\mathbf{R}^n} K(x-y)\,\Phi(y)\,dy,$$

since Φ is supported in the unit ball. From this it is also clear that $K^{(1)}$ satisfies the desired estimates, and our lemma is proved.

3.3.3 We now return to the proof of (49). Because of the invariance properties (50) and (51) of the operator T, and the corresponding transformation properties of the H^p norm given by (52), it suffices to verify (49) when a is an atom that is supported in the unit ball about the origin.

For such an atom, consider $M_\Phi(Ta) = \sup_{t>0}|Ta * \Phi_t|$. For $|x| \le 2$, we use the fact that $M_\Phi(Ta) \le cM(Ta)$; here M is the standard maximal operator of Chapter 1. Thus

$$\begin{aligned}\int_{|x|\le 2}[M_\Phi(Ta)]^p\,dx &\le c^p\int_{|x|\le 2}[M(Ta)]^p\,dx \\ &\le \bar{c}\|M(Ta)\|_{L^2}^p \le c\|Ta\|_{L^2}^p \le c\|a\|_{L^2}^p \le c.\end{aligned}$$

The next-to-last inequality is a consequence of the L^2 boundedness of T, which we assumed.

For $|x| \geq 2$, we use the fact that

$$(Ta) * \Phi_t = a * K * \Phi_t = a * K^{(t)}.$$

In other words,

$$(Ta * \Phi_t)(x) = \int_{|y| \leq 1} K^{(t)}(x-y)\, a(y)\, dy. \tag{53}$$

By Taylor's theorem,

$$K^{(t)}(x-y) = \sum_{|\beta| \leq \lfloor \gamma \rfloor} (\partial_x^\beta K^{(t)})(x) \cdot \frac{(-y)^\beta}{\beta!} + R^{(t)}(x,y),$$

and

$$|R^{(t)}(x,y)| \leq A \frac{|y|^{\gamma - \lfloor \gamma \rfloor}}{|x|^{n+\gamma}} \qquad \text{for } |x| \geq 2|y|.$$

We insert this in (53), and use the moment conditions (see §2.2) satisfied by the atom. The result is that

$$\sup_{t>0} |(Ta * \Phi_t)(x)| \leq c|x|^{-n-\gamma}.$$

Since $(n+\gamma)p > n$, it is then obvious that

$$\int_{|x| \geq 2} [M_\Phi(Ta)]^p\, dx \leq c',$$

and (49) is therefore established. This completes the proof of Theorem 4.

4. Appendix: Relation with harmonic functions

In this section, we return closer to the historical source of H^p theory and re-examine these spaces from the point of view of harmonic functions. That is, we shall show that the elements of H^p (which were initially considered as distributions on $\mathbf{R}^n$) can also be viewed as the boundary values of certain harmonic functions in the upper half-space $\mathbf{R}^{n+1}_+$; or, alternatively, in terms of systems of conjugate harmonic functions that generalize the notion of analytic functions occurring in the case $n = 1$; or, finally, by way of the area integral of Lusin.

The ideas described below are an outgrowth of the theory of harmonic functions appearing in Chapter 7 of *Singular Integrals* and Chapter 6 of *Fourier Analysis*. In order to keep the present exposition relatively brief, we omit some arguments that are presented in detail in those works.

4.1 Characterization in terms of harmonic functions; a generalized Fatou theorem. An element of H^p is a distribution on $\mathbf{R}^n$ that, according to Theorem 1, satisfies certain equivalent maximal properties. Our first change in viewpoint will be to identify such a distribution with the boundary values of a harmonic function on the upper half-space

$$\mathbf{R}^{n+1}_+ = \{(x,t) : x \in \mathbf{R}^n, t > 0\}.$$

Let u be harmonic on $\mathbf{R}^{n+1}_+$ and, in accordance with our previous definitions, we let u^* denote its nontangential maximal function, $u^*(x) = \sup_{|y-x|<t} |u(y,t)|$.

PROPOSITION 1. *Suppose* $0 < p \leq \infty$. *If* u *is harmonic in* $\mathbf{R}^{n+1}_+$, *then*

$$u^* \in L^p(\mathbf{R}^n), \tag{54}$$

if and only if u *is the Poisson integral of an* $f \in H^p(\mathbf{R}^n)$, *i.e.,*

$$u(x,t) = (f * P_t)(x), \quad \textit{for some } f \in H^p. \tag{55}$$

Moreover, $\|u^*\|_{L^p} \simeq \|f\|_{H^p}$.‡

Proof. Suppose $f \in H^p$. Then $u = f * P_t$ is a well-defined harmonic function with $u^* \in L^p$ and $\|u^*\|_{L^p} \simeq \|f\|_{H^p}$, by Theorem 1 and the discussion that precedes it (in §1).

Conversely, suppose that $u^* \in L^p$. We observe first that $|u(x,t)| \leq ct^{-n/p}$. In fact, by definition, $|u(x,t)| \leq u^*(y)$, whenever $|y-x| < t$, so by integration

$$|u(x,t)| \leq \frac{c_n}{t^n} \int_{|y-x|<t} (u^*(y))^p\, dy \leq c_n t^{-n} \|u^*\|^p_{L^p},$$

which proves our assertion. Therefore u is bounded in any half-space $\{(x,t) : t \geq \varepsilon > 0\}$.

Let $f_\varepsilon(x) = u(x,\varepsilon)$; we claim that $(f_\varepsilon * P_t)(x) = u(x, t+\varepsilon)$. Indeed, both functions are bounded and harmonic on $\mathbf{R}^{n+1}_+$, continuous on the closure $\bar{\mathbf{R}}^{n+1}_+$, and equal on the boundary. By a version of the maximum modulus principle (see *Singular Integrals*, Chapter 7, §1.2), we get the desired equality. Now set $u_\varepsilon(x,t) = u(x, t+\varepsilon)$. Then, since $u^*_\varepsilon(x) \leq u^*(x)$ and $u_\varepsilon = f * P_\varepsilon$, Theorem 1 implies that the family $\{f_\varepsilon\}$ is uniformly bounded in the H^p norm and, by §1.8 in the present chapter, the family is also bounded as a set of tempered distributions.

Therefore (by weak compactness), there exists a sequence $\varepsilon_k \to 0$ and a distribution f, so that $f_{\varepsilon_k} \to f$ in the sense of distributions. We claim that $f \in H^p$ and that $u(x,t) = (f * P_t)(x)$.

To see this, note that by the proof of Theorem 1 (using again the fact that $u^*_\varepsilon(x) \leq u^*(x) \in L^p$), for any $\Phi \in \mathcal{S}$, $\sup_{\varepsilon>0} M_\Phi(f_\varepsilon)(x) \in L^p$. Also, for each x and t,

$$(f_{\varepsilon_k} * \Phi_t)(x) \to (f * \Phi_t)(x), \quad \text{as } k \to \infty.$$

‡ For the case $p > 1$, compare with *Singular Integrals*, Chapter 7, §1.2.

Thus

$$\sup_t |f * \Phi_t(x)| \le \sup_{\varepsilon_k, t} |f_{\varepsilon_k} * \Phi_t(x)| \le \sup_\varepsilon M_\Phi(f_\varepsilon)(x) \in L^p,$$

and $f \in H^p$. The fact that, for each x and t,

$$u(x, t + \varepsilon_k) = f_{\varepsilon_k} * P_t(x) \to f * P_t(x)$$

follows immediately from the discussion in §1, in which $P(x) = P_1(x)$ was decomposed as $P = \phi * h + \psi$, with $\phi, \psi \in \mathcal{S}$ and $h \in L^1(\mathbf{R}^n)$. The proof of the proposition is therefore complete.

4.2 Conjugate harmonic functions. We now come to the characterization of H^p in terms of systems of conjugate harmonic functions. The simplest formulation occurs in the range $1 - 1/n < p < \infty$, and we consider this case in some detail. The general case is described in §5.16 below. Notice that when $n = 1$, we are already dealing with the full range $0 < p < \infty$, and that for general n, our range $1 - 1/n < p < \infty$ contains the crucial case $p = 1$. The appropriate notion of conjugacy for this range can be formulated as follows.

We suppose that we are given $n+1$ functions $u_0, u_1, \dots, u_n$ on $\mathbf{R}^{n+1}_+$ that satisfy the following generalized Cauchy-Riemann equations:

$$\sum_{j=0}^{n} \frac{\partial u_j}{\partial x_j} = 0, \quad \text{and} \quad \frac{\partial u_j}{\partial x_k} = \frac{\partial u_k}{\partial x_j}, \quad 0 \le j, k \le n. \tag{56}$$

Here, and at other times, it will be convenient to set $x_0 = t$. Let us recall some elementary facts about solutions to (56):

(i) Each u_j is necessarily harmonic.

(ii) The equations (56) are equivalent with the existence of a harmonic function H on $\mathbf{R}^{n+1}_+$ so that $u_j = \partial H / \partial x_j$, $j = 0, 1, \dots, n$.

(iii) When $n = 1$, the equations in (56) are the usual Cauchy-Riemann equations, which assert that $F = u_0 + iu_1$ is a holomorphic function of $x_1 + ix_0$.

By analogy with the classical theory of H^p spaces (when $n = 1$), we are led to consider the condition

$$\sup_{t>0} \int_{\mathbf{R}^n} |F(x,t)|^p \, dx < \infty, \tag{57}$$

where $F = (u_0, u_1, \dots, u_n)$.

PROPOSITION 2. *Let $1 - 1/n < p < \infty$ and suppose that u is harmonic on $\mathbf{R}^{n+1}_+$. Then u satisfies the H^p property of Proposition 1 (i.e., $u^* \in L^p$), if and only if there is a system $F = (u_0, u_1, \dots, u_n)$, with $u_0 = u$, satisfying the equations (56) and the L^p boundedness property (57). Moreover, $\|u^*\|^p_{L^p}$ is equivalent with the quantity appearing in (57), and the correspondence $F \mapsto u_0 = u$ is one to one.*

Proof. Besides the Poisson kernel $P_t(x) = t^{-n} P(x/t)$, with $P(x) = c_n(1 + |x|^2)^{(n+1)/2}$, we shall need to consider the *conjugate Poisson kernels* $Q_t^{(j)}$, $j = 1, \dots, n$, defined by

$$Q_t^{(j)}(x) = t^{-n} Q^{(j)}(x/t), \quad \text{where} \quad Q^{(j)}(x) = \frac{c_n x_j}{(1+|x|^2)^{\frac{n+1}{2}}}.$$

For each $t > 0$, the operators $f \mapsto f * P_t$ and $f \mapsto f * Q_t^{(j)}$, initially defined on (say) $L^2(\mathbf{R}^n)$, satisfy the properties required of the singular integrals treated by Theorem 4 in §3. Indeed, each of these operators is of the form $Tf = f * K$, where the distribution K is actually an L^2 function. We have[†]

$$\widehat{P}_t(\xi) = e^{-2\pi|\xi|t}, \qquad \widehat{Q}_t^{(j)}(\xi) = i \frac{\xi_j}{|\xi|} e^{-2\pi|\xi|t}.$$

Thus assumption (40) is clearly satisfied; similarly, it is straightforward to verify that the differential inequalities (48) hold for all $\gamma > 0$. Moreover the bounds are uniform in $t > 0$.

The limiting case $t \to 0$ of $f \mapsto f * P_t$ is of course the identity operator. For the conjugate Poisson kernels, the limiting case gives the Riesz transforms R_j, $j = 1, \dots, n$; here $R_j f = f * K^{(j)}$, where the $K^{(j)}$ are (principal value) distributions, with

$$\widehat{K}^{(j)}(\xi) = i \frac{\xi_j}{|\xi|}, \quad \text{and} \quad K^{(j)}(x) = c_n \frac{x_j}{|x|^{n+1}},$$

so again the transformations R_j satisfy all the assumptions of Theorem 4. By using the Fourier transform, it is easy to see that, when $f \in L^2(\mathbf{R}^n)$, if we define $u = u_0 = f * P_t$, and $u_j = f * Q_t^{(j)}$, $j = 1, \dots, n$, then the system $F = (u_0, \dots, u_n)$ satisfies the generalized Cauchy-Riemann equations (56); in addition we have the identity $f * Q_t^{(j)} = R_j(f) * P_t$.[‡]

Having recalled these facts about Poisson and conjugate Poisson kernels, we return to the proof of Proposition 2. Suppose first that $u^* \in L^p$. Then by the previous proposition, $u = f * P_t$ for some $f \in H^p$. By the atomic decomposition of f (Theorem 2 in §2), there is a sequence $f_k \subset H^p \cap L^2$, with $f_k \to f$ in H^p norm, and with $\|f_k\|_{H^p} \le c\|f\|_{H^p}$. Now set

$$u_k^{(0)} = f_k * P_t, \quad u_k^{(j)} = f_k * Q_t^{(j)}, \quad F_k = (u_k^{(0)}, \dots, u_k^{(n)}).$$

Observe that

$$\|u_k^{(0)}(x,t)\|^p_{L^p(\mathbf{R}^n)} \le \|u_k^{(0)}\|^p_{L^p(\mathbf{R}^n)} \le c\|f_k\|^p_{H^p} \le c\|f\|^p_{H^p},$$

the next to last inequality holding because of Theorem 1. Similarly, using the fact that $u_k^{(j)}(\cdot, t) = R_j(f_k) * P_t$, we get

$$\|u_k^{(j)}\|^p_{L^p} \le c\|R_j(f_k)\|_{H^p} \le c'\|f_k\|_{H^p} \le c\|f\|_{H^p}.$$

[†] See *Fourier Analysis*, Chapter 6, §4.

[‡] For this, see *Singular Integrals*, Chapter 3, §2.3.

Here the second inequality is a consequence of the boundedness of the operator R_j on H^p (by Theorem 4). Altogether then

$$\sup_{t>0}\int_{\mathbf{R}^n} |F_k(x,t)|^p\,dx \le c\|f\|_{H^p}. \tag{58}$$

Now the f_k converge to f in H^p norm, therefore the $R_j(f_k)$ converge to $R_j(f)$ in H^p norm. This gives the pointwise convergence of $f_k * P_t$ to $f * P_t$, and the pointwise convergence of $R_j(f_k) * P_t$ to $R_j(f) * P_t$; in fact, this convergence is uniform on compact sets. The limits $\lim_{k\to\infty} u_k^{(j)}(x,t)$ are then harmonic and satisfy the equations (56); also (57) follows from (58) by Fatou's lemma. Notice that the arguments up to this point are valid for all $p > 0$; it is in the converse direction that the requirement $p > n/(n-1)$ will be decisive.

The key fact is that whenever $F = (u_0, \ldots, u_n)$ satisfies the generalized Cauchy-Riemann equations (56), the function $|F|^q$ is subharmonic for $q \ge (n-1)/n$.[†]

The subharmonicity of $|F|^q$ has the following consequence. Whenever $(n-1)/n \le q$, we have, for each x, t, and $\varepsilon > 0$,

$$|F(x,t+\varepsilon)|^q \le (|F(\cdot,\varepsilon)|^q * P_t)(x). \tag{59}$$

To see this, observe first that $|F(x,t)|$ is bounded in any half-space $t \ge \varepsilon > 0$. In fact, by the subharmonicity of $|F|^p$ (if $p > (n-1)/n$), we have

$$|F(x,t)|^p \le \frac{1}{|\tilde{B}|}\int_{\tilde{B}} |F(y,t)|^p\,dy\,dt,$$

where $\tilde{B}$ is the $(n+1)$ dimensional ball centered at (x,t) of radius t. Thus $|f(x,t)| \le ct^{-n/p}$ follows if we invoke (59). Therefore the right side of (57) is a harmonic function that is bounded on $\mathbf{R}^{n+1}_+$ and continuous on the closure of $\mathbf{R}^{n+1}_+$. The left side of (59) is a subharmonic function, bounded on $\mathbf{R}^{n+1}_+$, and continuous on the closure of $\mathbf{R}^{n+1}_+$. Since their boundary values agree, it follows by a variant of the maximum principle[‡] that (59) holds for all x and t.

Now assumption (57) implies that the functions $|F(\cdot,\varepsilon)|^q$ are uniformly in $L^r(\mathbf{R}^n)$, where $r = p/q > 1$, provided that we choose $(n-1)/n < q < p$. By the weak compactness of L^r, we can find an $h \in L^r$ and a sequence $\varepsilon_k \to 0$, so that $|F(\cdot,\varepsilon_k)|^q \to h$ weakly in L^r, as $k \to \infty$. As a consequence,

$$|F(x,t)|^q \le (h * P_t)(x).$$

Notice also that $h \ge 0$ and that

$$\|h\|_{L^r}^r \le \sup_{t>0}\int_{\mathbf{R}^n} |F(x,t)|^p\,dx,$$

[†] For a proof of this, which we do not give here, see either *Fourier Analysis*, Chapter 6, §4, or *Singular Integrals*, Chapter 7, §3.

[‡] See the alternative arguments in *Fourier Analysis*, Chapter 2, §4, and *Singular Integrals*, Chapter 7, §3.2.

The nontangential domination of the Poisson integral by the standard maximal operator M (as in Chapter 2) then gives $u^* = u_0^* \le |F|^* \le c\cdot[M(h)]^{1/q}$. Finally,

$$\int_{\mathbf{R}^n} (u^*)^p\,dx \le c' \int (Mh)^{p/q} = c' \int (Mh)^r\,dx \le c \int h^r\,dx,$$

by Theorem 1 in Chapter 1, since $r > 1$. The final point to observe is that if $F = (u_0, \ldots, u_n)$ satisfies (56) and (57), and $u_0 \equiv 0$, then $f \equiv 0$. In fact, equation (56) implies that $F(x,t)$ is independent of t and the majorization $|F(x,t)| \le ct^{-n/p}$ gives the desired result. This concludes the proof of the proposition.

4.3 Characterization by the Riesz transforms. We shall now see that the property $f \in H^p$ can be characterized by appropriate singular integrals in a way that has some analogy with the earlier maximal characterization. What follows is closely connected to the characterization by conjugate harmonic functions, and so here (as in §4.2 above), we consider the case $1 - 1/n < p < \infty$. For the general case, see §5.16 below.

We intend to prove a precise version of the following heuristic statement: A distribution f belongs to H^p if and only if $f \in L^p$ and $R_j(f) \in L^p$, for $j = 1, \ldots, n$. There are two issues raised here that need to be clarified, in order to make our assertion valid. First we must limit the class of distributions f to those for which the $R_j(f)$ are well defined (as distributions). Second, we must make sense of the statements that f and $R_j(f)$ belong to L^p, when $p < 1$.

Addressing the first point, we shall always require that our distributions f are *restricted at infinity*, in the sense that, for all sufficiently large $r < \infty$, $F * \Phi \in L^r(\mathbf{R}^n)$ for all $\Phi \in \mathcal{S}$. Note that, by the fourth remark in §1.8, any $f \in H^p$ (with $p < \infty$) has this property, taking $r \ge p$.[†]

With the assumption that f is restricted at infinity, we can define distributions $R_j(f)$ as in §3.2. Indeed, writing $R_j(f) = f * K$, we can split the kernel $K = K_0 + K_\infty$, where K_0 is a distribution supported near the origin, and K_∞ is a bounded function that is $O(|x|^{-n})$ at infinity. Now $f * K_0$ makes sense for any distribution f, while $f * K_\infty$ is defined by $\langle f * K_\infty, \Phi\rangle = \langle f * \tilde{\Phi}, \tilde{K}\rangle$, using the fact that $\tilde{K} \in L^{r'}$, where $1/r + 1/r' = 1$.

To make sense of the assertion $f \in L^p$, $R_j(f) \in L^p$, we fix a $\phi \in \mathcal{S}$ with $\int \phi\,dx = 1$, set $\phi_\varepsilon(x) = \varepsilon^{-n}\phi(x/\varepsilon)$, and write

$$\|f * \phi_\varepsilon\|_{L^p} + \sum_j \|R_j(f) * \phi_\varepsilon\|_{L^p} \le A, \quad \text{all } \varepsilon > 0. \tag{60}$$

Proposition 3. *Suppose $1 - 1/n < p < \infty$ and let f be a distribution that is restricted at infinity. Then $f \in H^p$ if and only if f and the $R_j(f)$ belong to L^p in the sense* (60).

Proof. Suppose first that (60) holds. Set $F_\varepsilon = (u_0^\varepsilon, u_1^\varepsilon, \ldots, u_n^\varepsilon)$, where $u_0^\varepsilon = f_\varepsilon * P_t$ and $u_j^\varepsilon = f_\varepsilon * Q_t^{(j)}$, $j = 1, \ldots, n$, with $f_\varepsilon = f * \phi_\varepsilon$. Now u_0^ε is

[†] This property is stronger than the fact that H^p distributions are "bounded", which we used when defining the Poisson integral of such a distribution in §1.1.

a harmonic function in $\mathbf{R}^{n+1}_+$, continuous and bounded in the closure. This is because our assumptions imply that f_ε and all its partial derivatives belong to L^r, for r sufficiently large. The same holds for the u_j^ε, and hence for F_ε. The harmonic majorization argument used above to prove (59) also shows that

$$|F_\varepsilon(x,t)|^q \le (|F_\varepsilon(\cdot,0)|^q * P_t)(x).$$

If we integrate this with respect to x, and use the fact that

$$F_\varepsilon(x,0) = (f * \phi_\varepsilon, R_1(f) * \phi_\varepsilon, \dots, R_n(f) * \phi_\varepsilon),$$

we get (by (60))

$$\int_{\mathbf{R}^n} |F_\varepsilon(x,t)|^p\,dx \le A.$$

Observe that $F_\varepsilon(x,t) \to F(x,t)$ for each x and $t > 0$, as $\varepsilon \to 0$. Thus, by Fatou's lemma,

$$\sup_{t>0} \int_{\mathbf{R}^n} |F(x,t)|^p\,dx \le A,$$

and Proposition 2 allows us to conclude that $f \in H^p$.

To prove the converse, recall that if $f \in H^p$, then it is restricted at infinity; in addition, the $R_j(f)$ also belong to H^p. Now $|(f * \phi_\varepsilon)(x)| \le M_\phi f(x)$, similarly $|R_j(f) * \phi_\varepsilon(x)| \le M_\phi(R_j f)(x)$; hence Theorem 1 shows that (60) holds, completing the proof of the proposition.

4.4 Characterization by Area Integral. One idea that played a key role in the development of the theory of H^p spaces was the principle that nontangential control of a harmonic function is essentially equivalent with control of an appropriate square function—more precisely, the generalized "area integral" of Lusin. An outgrowth of this idea is the following additional characterization of H^p.

Besides the nontangential maximal function, we consider the square function $S(u)$ defined by

$$Su(x) = \left(\int_{\Gamma(x)} |\nabla u(y,t)|^2\, t^{1-n}\,dy\,dt\right)^{1/2}, \quad x \in \mathbf{R}^n,$$

where $\Gamma(x)$ is the cone $\{(y,t) : |y-x| < t\}$ with vertex at $x \in \mathbf{R}^n$, and $|\nabla u|^2 = |\partial u/\partial t|^2 + \sum_{j=1}^{n} |\partial u/\partial x_j|^2$.

PROPOSITION 4. *Suppose $0 < p < \infty$, and let u be a harmonic function on $\mathbf{R}^{n+1}_+$ that vanishes at infinity in the sense that $u(x,t) \to 0$ as $t \to \infty$. Then $u \in H^p$ (i.e., $u^* \in L^p$) if and only if $Su \in L^p$. Also $\|u^*\|_{L^p} \simeq \|Su\|_{L^p}$.*

The proposition may be viewed as the "global" version of a "local" analogue that states that, for a harmonic function u, the set of $x \in \mathbf{R}^n$ where $u^*(x)$ is finite is, up to a set of measure zero, the same as the set of x for which $Su(x)$ is finite. The detailed proof of the latter equivalence may be found in *Singular Integrals*, Chapter 7, §2. The proof of the present proposition closely parallels that of the local version; hence we shall only sketch the main modifications that we need. These require us to make certain quantitative estimates, where qualitative statements had previously sufficed.

4.4.1 The first useful observation is that the aperture of the cone used in the definition of u^* may be chosen differently from the aperture used to define $S(u)$. In fact, let $\Gamma_a(x)$ be the cone of aperture a with vertex at $x \in \mathbf{R}^n$ and write

$$u_a^*(x) = \sup_{\Gamma(x)} |u(y,t)| = \sup_{|y-x|<at} |u(y,t)|,$$

$$(S_a u)(x) = \left(\int_{\Gamma_a(x)} |\nabla u|^2\, t^{1-n}\,dy\,dt\right)^{1/2}.$$

Then it is an immediate consequence of Chapter 2, §2.5 (with $F = |u|^p$) that $\|u_a^*\|_{L^p} \simeq \|u_b^*\|_{L^p}$, for any $0 < a, b < \infty$. Similarly, $\|S_a(u)\|_{L^p} \simeq \|S_b(u)\|_{L^p}$. The proof of this second assertion is in the same spirit as that for the nontangential maximal functions, although the details are quite different.[‡]

4.4.2 Another preliminary remark deals with the case $1 < p < \infty$. This case is essentially a consequence of the theory of singular integrals in Chapter 1, §8.23. Indeed, it is seen there that, if $f \in L^p$ and $u(x,t) = f * P_t(x)$, then $\|Su\|_{L^p} \simeq \|f\|_{L^p}$. This immediately shows that $\|Su\|_{L^p} \le A\|u^*\|_{L^p}$.

For the converse, one uses the maximal theorem of Chapter 1 and §2.1 of Chapter 2. The result is $\|u^*\|_{L^p} \le A\|Su\|_{L^p}$, under the assumption that u is the Poisson integral of an L^p function. To remove this *a priori* assumption one sets

$$u_{\varepsilon,N}(x,t) = u(x,t+\varepsilon) - u(x,t+N)$$

and observes that $u_{\varepsilon,N}(x,t) \le c_{\varepsilon,N}(Su)(x)$. Thus $\sup_{t>0} \int |u_{\varepsilon,N}(x,t)|^p\,dx < \infty$, consequently $u_{\varepsilon,N}$ is the Poisson integral of an L^p function[†] and, by what has already been proved, $\|u^*_{\varepsilon,N}\|_{L^p} \le \|S(u_{\varepsilon,N})\|_{L^p}$. However,

$$(Su_{\varepsilon,N})(x) \le (Su_\varepsilon)(x) + (Su_N)(x) \le c(Su)(x),$$

therefore $\|u^*_{\varepsilon,N}\|_{L^p} \le c\|Su\|_{L^p}$; an application of Fatou's lemma shows that $\limsup_{\substack{\varepsilon\to 0\\ N\to\infty}} \|u^*_{\varepsilon,N}\|_{L^p} \ge \|u^*\|_{L^p}$, concluding the proof.

4.4.3 We shall also need certain integration lemmas. Suppose F is a closed set in $\mathbf{R}^n$. Let $\mathcal{R}_a(F)$ denote the nontangential region built over F with cones of aperture $a > 0$;[‡] i.e.,

$$\mathcal{R}_a(F) = \bigcup_{x\in F} \Gamma_a(x).$$

Let $A = A(x,t)$ be an arbitrary nonnegative measurable function on $\mathbf{R}^{n+1}_+$. Then

$$\int_F \left\{\int_{\Gamma(x)} A(y,t)\,dy\,dt\right\} dx \le c_a \int_{\mathcal{R}_a(F)} A(y,t)\,t^n\,dt\,dy. \tag{60}$$

[‡] See §4.4.3 below and Chapter 4, §6.13.

[†] See *Singular Integrals*, Chapter 7, §1.2.

[‡] Recall that when $O = {}^cF$, and $a = 1$, then ${}^c\mathcal{R}_a(F) = T(O)$ is the "tent" over O; see §2.2 of Chapter 2.

This is an immediate consequence of Fubini's theorem and the observation that
$$\int_\Gamma \chi\left(\frac{x-y}{at}\right)dx \le c_a t^n,$$
where χ is the characteristic function of the unit ball. The converse to the above requires the notion of global γ-density, as defined in Chapter 2, §2.5.

Indeed, we can assert that there is a γ sufficiently close to 1 so that
$$\int_{\mathcal{R}_a(F^*)} A(y,t)\, t^n\, dt\, dy \le c_{a,\gamma}\int_F\left\{\int_{\Gamma(x)} A(y,t)\, dy\, dt\right\} dx, \qquad (61)$$
where F^* is the set of points of global γ-density with respect to F.

To see (61), one needs that
$$\int_F \chi\left(\frac{x-y}{t}\right) dx \ge c_{a,\gamma}\, t^n, \quad \text{if } (y,t)\in\mathcal{R}_a(F^*). \qquad (62)$$
But for a point (y,t) of that form, there exists an $\bar{x}\in F^*$ so that $|y-\bar{x}|<at$. Now it is obvious that
$$|B(\bar{x},at)\cap {}^cB(y,at)| \le c|B(\bar{x},at)|,$$
with $c<1$. So
$$|F\cap B(y,at)| \ge |F\cap B(\bar{x},at)| - |B(\bar{x},at)\cap {}^cB(y,at)| \ge (\gamma - c)\cdot|B(\bar{x},at)|.$$
Therefore we get (62), if $\gamma > c$, and thus (61) is proved.

4.4.4 The last preliminary facts needed are the inequalities
$$(t|\nabla u|)_b^*(x) \le c_{a,b}\, u_a^*(x)$$
and
$$(t|\nabla u|)_b^*(x) \le c_{a,b}(S_a u)(x),$$
for harmonic functions u, where $(\cdot)_b^*$ denotes the nontangential maximal function taken over cones of aperture b; we also require that $a>b$. These inequalities are easy consequences of the mean-value property of harmonic functions. See *Singular Integrals*, Chapter 7, §2.2.2 for the variant using truncated cones.

4.4.5 To prove the proposition one uses Green's theorem in the form
$$\int_{\mathcal{R}}(f\Delta g - g\Delta f) = \int_{b\mathcal{R}}\left(f\frac{\partial g}{\partial n} - g\frac{\partial f}{\partial n}\right) d\sigma,$$
where $\partial/\partial n$ denotes differentiation with respect to the outward unit normal on $b\mathcal{R}$ and $d\sigma$ is surface measure on $b\mathcal{R}$.

To begin with, let $u^*\in L^p$, and take $f=t$, $g=|u|^2/2$, with $F=\{x: u_a^*(x)>\alpha\}$. Using (60), the result is that, whenever $b<a$,
$$|\{x:(S_b u)(x)>\alpha\}| \le c\left\{\frac{1}{\alpha^2}\int_{u_a^*\le\alpha}(u_a^*)^2\, dx + |\{x: u_a^*(x)>\alpha\}|\right\}.$$
Multiplying by α^{p-1} and integrating in α then shows
$$\int (S_b u(x))^p\, dx \le c_p\int (u_a^*(x))^p\, dx,$$
for $0<p<2$.

The proof of the converse is analogous. We choose b_1 so that $b<b_1<a$. We set $F=\{x: S_a u(x)\le\alpha\}$, F^* the points of global γ-density with respect to F, and $\mathcal{R}=\mathcal{R}_{b_1}(F^*)$. Then by the use of (61), one can show that
$$|\{x: u_b^*(x)>c'\alpha\}| \le c\left\{\frac{1}{\alpha^2}\int_{S_a(u)\le\alpha} S_a^2(u)(x)\, dx + |\{x: S_a(u)(x)>\alpha\}|\right\},$$
and again an integration in α proves
$$\int (u_b^*(x))^p\, dx \le c_p'\int (S_a u(x))^p\, dx$$
for $0<p<2$. Given the remarks in §4.4.1, §4.4.3, and §4.4.4, the proofs of the above assertions follow closely the "local" analogues given in *Singular Integrals*, Chapter 7, §2.3 and §2.4. Details can be found in C. Fefferman and Stein [1972].

5. *Further results*

A. General properties of H^p spaces

5.1 A number of the standard properties of L^p spaces (when $p>1$) are also valid for H^p spaces (when $p\le 1$), in suitably modified form.

(a) *Completeness:* H^p is complete in the metric $d(f,g)=\|f-g\|_{H^p}^p$, $p\le 1$.

(b) *Weak compactness of unit ball:* If $f_n\in H^p$ with $\|f_n\|_{H^p}\le A$, then there exists an $f\in H^p$ and a subsequence f_{n_k} so that $f_{n_k}\to f$ in the sense of distributions.

(c) *Approximation in norm:* Let $f\in H^p$, and let $\Phi\in\mathcal{S}$ with $\int\Phi\, dx=1$. Then $\|f*\Phi_t\|_{H^p}\le c\|f\|_{H^p}$ and $f*\Phi_t\to f$ in H^p "norm" as $t\to 0$.

(d) *Definability almost everywhere:* To each distribution $f\in H^p$, there is associated a function $f_0(x)$, defined a.e. on $\mathbf{R}^n$, so that for each $\Phi\in\mathcal{S}$ with $\int\Phi\, dx=1$, we have $(f*\Phi_t)(x)\to f_0(x)$ for a.e. x, as $t\to 0$. Note however that the function $f_0(x)$ may vanish (a.e.) without f being zero! In fact, we may take $f=d\mu$, a totally singular measure as in §5.4(b) below.

To prove (c), observe that if $\Phi,\Psi\in\mathcal{S}$, then any finite collection of seminorms is bounded on the set $\{\Phi*\Psi_s: 0<s\le 1\}$. To prove (d), one uses the fact that $L^1\cap H^p$ is dense in H^p (see also §5.14 below).

5.2 The following are additional dense subspaces of H^p, $0 < p < \infty$.

(a) $\{f \in \mathcal{S} : \int x^\alpha f(x)\,dx = 0 \text{ for all } \alpha \in \mathbf{N}^n\}$.

(b) The set of all bounded, compactly supported functions for which $\int_{\mathbf{R}^n} x^\alpha f(x)\,dx = 0$, for all α with $|\alpha| < N$. Here N is any fixed integer with $N > n(p^{-1} - 1)$.

For (b), see §2.4. The assertion (a) is a consequence of these considerations. See also *Singular Integrals*, Chapter 7, §3.3.3.

5.3 If a distribution $f \in H^1$ is nonnegative on some open set O, then it agrees with a function in the class $L \log L$ on any compact $K \subset O$; i.e., $|f| \log(1 + |f|)$ is integrable on K.

Note that in this case, $(M_\Phi f)(x) \geq cM(f\eta)(x)$ for $x \in K$, where η is a suitable cut-off function and Φ is supported in the unit ball. One then invokes Chapter 1, §8.14.

5.4 Suppose $p \leq 1$ and $f \in H^p$.

(a) The Fourier transform $\widehat{f}$ is continuous on $\mathbf{R}^n$ and

$$|\widehat{f}(\xi)| \leq A|\xi|^{n(p^{-1}-1)} \cdot \|f\|_{H^p}, \quad \xi \in \mathbf{R}^n.$$

(b) Near the origin, this can be refined to

$$\frac{\widehat{f}(\xi)}{|\xi|^{n(p^{-1}-1)}} \to 0, \qquad \text{as } \xi \to 0.$$

(c) This shows that the moment conditions in §3.2 are necessary. Indeed, if $f \in H^1(\mathbf{R}^1)$ then necessarily $\int_{\mathbf{R}^n} f(x)\,dx = 0$. More generally, if $f \in L^1(\mathbf{R}^n) \cap H^p(\mathbf{R}^n)$ then $\int_{\mathbf{R}^n} x^\alpha f(x)\,dx = 0$, whenever $|\alpha| \leq n(p^{-1} - 1)$ and the function $x^\alpha f(x)$ is in L^1.

(d) A variant of (a) is a generalization of an inequality of Hardy and Littlewood for $2 \geq p > 1$, namely that for $f \in H^p$,

$$\left(\int_{\mathbf{R}^n} |\widehat{f}(\xi)|^p |\xi|^{(p-2)n}\,d\xi \right)^{1/p} \leq c_p \|f\|_{H^p}, \qquad p \leq 1.$$

To prove (a), use the fact that $f * \Phi_t \in L^1$ and

$$\|f * \Phi_t\|_{L^1} \leq At^{-n(p^{-1}-1)} \|f\|_{H^p};$$

see §1.4. The refined assertion (b) then follows by verifying it on a dense subspace of H^p, such as in §5.2. The inequality (d) can be proved directly from the atomic decomposition.

5.5 Various examples illustrate the nature of distributions that are in H^p.

(a) A bounded, compactly supported function f belongs to H^p if and only if it satisfies the moment conditions $\int x^\alpha f(x)\,dx = 0$ for all $|\alpha| \leq n(p^{-1} - 1)$.

(b) When $p < 1$, an L^1 function (or, more generally, a finite measure) having compact support belongs to H^p exactly when it satisfies the above moment conditions.

(c) A finite linear combination F of derivatives of translates of the Dirac delta function belongs to H^p exactly when: (i) the maximum order of differentiation r satisfies $r < n(p^{-1}-1)$; and (ii) $F(x^\alpha) = 0$ whenever $|\alpha| \leq n(p^{-1}-1)$.

In particular, if δ_a is the Dirac function centered at a, then $\delta_0 - \delta_a \in H^p$ for $n/(n+1) < p < 1$; also $(\partial/\partial x_j)[\delta_0 - \delta_a] \in H^p$ for $n/(n+2) < p < n/(n+1)$, etc.

(d) Let Ω be a bounded (say) smooth domain in $\mathbf{R}^n$, and let f be a nonnegative function defined in Ω with $f(x) \approx d(x)^{-N}$; here $d(x)$ denotes the distance from x to ${}^c\Omega$. Then f is the restriction to Ω of a distribution $F \in H^p(\mathbf{R}^n)$ if and only if $Np < n$ (i.e., exactly when $f \in L^p(\Omega)$).

(e) A general distribution (even one that is "bounded" in the sense of §1.1) may not agree with an H^p function on any open set. On $\mathbf{R}^1$, consider the distribution f given by the lacunary series $f = \sum_{k=0}^{\infty} a_k e^{i2^k x}$, where $\{a_k\}$ is (say) bounded, but $\sum |a_k|^2 = \infty$. Then, if $\phi \in C^\infty$ is bounded and not identically zero, we have that $\phi f \notin H^p$ for all $p > 0$.

The proof of (b) follows the argument in §2.2, except that the estimate in Chapter 1, §8.15 is used for the ball B^*. The assertion (d) can be deduced from results in Miyachi [1990]. To show (e), one uses a basic feature of non-L^2 lacunary series, namely that they are summable on at most sets of measure zero; for this see Zygmund [1959].

5.6 Suppose we drop the cancellation property $\int a(x)\,dx = 0$ when considering H^1 atoms in §2.2. Then the f that are obtainable as sums (30) are exactly all $f \in L^1(\mathbf{R}^n)$.

To see this, consider $f \in L^1(\mathbf{R}^1)$, and for each k define

$$f_k(x) = 2^k \int_{\ell \cdot 2^{-k}}^{(\ell+1)\cdot 2^{-k}} f(u)\,du, \qquad x \in (\ell 2^{-k}, (\ell+1)2^{-k});$$

f_k is the "dyadic martingale" sequence associated to f. Clearly $f_k \to f$ in L^1 norm, as $k \to \infty$. Thus there exist $k_j \to \infty$ so that $\sum_j \|f_{k_j} - f_{k_{j-1}}\|_{L^1} < \infty$. Now write $f = f_{k_0} + \sum_{j=1}^{\infty} (f_{k_j} - f_{k_{j-1}})$. Since $f_{k_j} - f_{k_{j-1}}$ is constant on each dyadic interval of length 2^{-k_j}, the resulting sum gives an "atomic" decomposition of f, whose putative atoms fail to have the cancellation property.

5.7 In the definition of H^p atoms given in §2.2, we may replace the size condition $|a| \le |B|^{-1/p}$ by the more general inequality

$$\frac{1}{|B|}\int_B |a(x)|^q\,dx \le |B|^{-q/p},$$

for a fixed $q > 1$.

Indeed, by the maximal theorem, $\int_{B^*}(Ma)^q\,dx \le c\int_B |a|^q\,dx$, and thus by Hölder's inequality (29) follows as before. Moreover, when $p < 1$, the size condition can be weakened further: it suffices to take

$$\frac{1}{|B|}\int_B |a(x)|\,dx \le |B|^{-1/p},$$

and use Chapter 1, §8.15.

That such atoms can be decomposed as sums of our standard atoms follows from Theorem 2. A direct proof of this fact can be found in García-Cuerva and Rubio de Francia [1985].

Another variant arises if we replace the support and size condition on a by (say)

$$|a(x)| \le |B|^{-1/p}\min\{1, |B|^r/|x - x_0|^{nr}\}$$

where $rp > 1$ and x_0 is the center of B. This allows a to have support outside of B, but restricts it to be suitably small at infinity. Such generalized atoms (and their variants) are called *molecules* and are useful in certain applications; see, e.g., Taibleson and G. Weiss [1980], Chao, Gilbert, and Tomas [1981].

5.8 The Lipchitz spaces Λ_γ can be paired with H^p if $\gamma = n(p^{-1} - 1)$ and $0 < p < 1$. Thus if $g \in \Lambda_\gamma$† then $\ell(f) = \int_{\mathbf{R}^n} f(x)\,g(x)\,dx$, initially defined for $f \in L^1 \cap H^p$, has a bounded extension to H^p, with $|\ell(f)| \le c\|f\|_{H^p}\|g\|_{\Lambda_\gamma}$.

There is also a converse statement, but it involves the homogeneous versions of the Λ_γ, See Duren, Romberg, and Shields [1969] for $n = 1$ and, e.g., Taibleson and G. Weiss [1980] for $n > 1$.

5.9 We discuss the minimal conditions on Φ so that whenever $f \in H^p$, $p \le 1$, we have that $M_\Phi(f) \in L^p$ with $\|M_\Phi f\|_{L^p} \le c\|f\|_{H^p}$.

(a) For Φ that have compact support, it suffices to have $\Phi \in \Lambda_\gamma$ for some $\gamma > n(p^{-1} - 1)$.

(b) For Φ not having compact support but that vanish at infinity and satisfy $|\partial_x^\gamma \Phi(x)| \le A(1 + |x|)^{-N}$ for $|\gamma| = \lfloor n(p^{-1} - 1)\rfloor + 1$, it suffices to have $Np > n$.

These statements are simple consequences of the atomic decomposition.

† See Chapter 6, §5.3, for the definition and basic properties of Λ_γ.

5.10 That the results in §5.9 are essentially sharp is illustrated by two types of Φ's that arise in a variety of questions in harmonic analysis.

(a) First consider $\Phi(x) = (1 - |x|^2)^\alpha$ for $|x| < 1$, with $\Phi(x) = 0$ for $|x| \ge 1$. If $\alpha < n(p^{-1} - 1)$ (or $\alpha = n(p^{-1} - 1)$), there exists an $f \in H^p$, $p \le 1$, so that $M_\Phi(f) = \infty$ a.e. (or $M_\Phi(f) \notin L^p(\mathbf{R}^n)$). However, there is a positive "weak-type" result when $\alpha = n(p^{-1} - 1)$ and $p \le 1$; in this case

$$|\{x : (M_\Phi f)(x) > \gamma\}| \le c\gamma^{-p}\|f\|_{H^p}^p, \qquad \text{for all } \gamma > 0.$$

(b) Next consider $\Phi(x) = |x|^{-(n/2)-\delta}J_{(n/2)+\delta}(2\pi|x|)$, where J_k is the Bessel function of order k. Since $(d/du)^\ell J_k(u) = O(u^{-1/2})$ as $u \to \infty$, comparison with §5.9(b) leads us to consider two cases:

$$p < \frac{2n}{n+1+2\delta}, \qquad \text{and} \qquad p = \frac{2n}{n+1+2\delta},$$

with $p \le 1$. In the first, there exists $f \in H^p$ so that $M_\Phi(f) = \infty$ almost everywhere. In the second, when $p < 1$, we have the analogous weak-type inequality, but $M_\Phi(f) \in L^p$ is in general false; when $p = 1$, $M_\Phi(f) = \infty$ a.e. for appropriate f, when $\delta = (n-1)/2$.

For these results, see Stein, Taibleson, and G. Weiss [1981], Stein [1983a]. In case (a), M_Φ is related to the spherical maximal operator (see Chapter 11); in case (b), we are dealing with Bochner-Riesz summability (see Chapter 9). It is a curious and paradoxical fact that, in the context of these results, we have at present a more complete picture of the situation for H^p when $p \le 1$ that we do for L^p when $p > 1$.

5.11 Suppose f is a (not necessarily tempered) distribution, and $\Phi \in C^\infty$ is compactly supported with $\int_{\mathbf{R}^n}\Phi\,dx \ne 0$. If $M_\Phi(f) \in L^p$ for some $p > 0$ then f is tempered, and hence $f \in H^p$. See Uchiyama [1986].

B. Harmonic and analytic functions

5.12 The classical Hardy spaces of holomorphic functions in the upper half-plane, denoted by $\mathcal{H}^p(\mathbf{R}^2_+)$, are defined as follows. For $0 < p < \infty$, the space $\mathcal{H}^p(\mathbf{R}^2_+)$ consists of the functions F, holomorphic on

$$\mathbf{R}^2_+ = \{z = x + iy \in \mathbf{C} : y > 0\},$$

that satisfy

$$\left(\sup_{y>0}\int_{\mathbf{R}^1}|F(x+iy)|^p\,dx\right)^{1/p} = \|F\|_{\mathcal{H}^p(\mathbf{R}^2_+)} < \infty.$$

The following hold for $F \in \mathcal{H}^p(\mathbf{R}^2_+)$.

(a) $F^*(x) = \sup_{|u|<y}|F(x-u+iy)| \in L^p(\mathbf{R}^1)$ with $\|F^*\|_{L^p} \le c_p\|F\|_{\mathcal{H}^p(\mathbf{R}^2_+)}$.

(b) $\lim_{y\to 0} F(x+iy) = F_0(x)$ exists for a.e. $x \in \mathbf{R}^1$.

(c) $\|F\|_{\mathcal{H}^p(\mathbf{R}^2_+)} = \|F_0\|_{L^p(\mathbf{R}^1)}$.

(d) $\lim_{y\to 0} F(x+iy) = F^b$ exists also in the sense of tempered distributions. The connection with the (real) Hardy spaces treated in this chapter is that $F^b \in H^p(\mathbf{R}^1)$ and $\|F^b\|_{H^p(\mathbf{R}^1)} \approx \|F\|_{\mathcal{H}^p(\mathbf{R}^2_+)}$.

(e) Moreover, every $f \in H^p(\mathbf{R}^1)$ can be written as $f = F_1^b + \overline{F_2^b}$, with $F_i \in \mathcal{H}^p(\mathbf{R}^2_+)$.

Assertions (a), (b), and (c) are well-known facts about the Hardy space $\mathcal{H}^p(\mathbf{R}^2_+)$. See, e.g., Zygmund [1959], Duren [1970], Garnett [1981], and also *Fourier Analysis*, chapters 2 and 3. Property (d) is a consequence of §4.1. The proof of (e) depends on the description of $H^p(\mathbf{R}^1)$ in terms of the Cauchy-Riemann equations given in §4.2. Indeed, if u_0 is the Poisson integral of f, then we may take $F_1 = F_2 = (u_0 + iu_1)/2$.

5.13 The generalization of §5.12 to higher dimensions requires that we pass to several complex variables and tube domains over cones. In fact, let Γ be an open cone in $\mathbf{R}^n$; its dual cone Γ^* is defined by

$$\Gamma^* = \{\xi \in \mathbf{R}^n : \langle x, \xi\rangle > 0 \text{ for all } x \in \Gamma\}.$$

We assume that Γ^* is nonempty (i.e., Γ is "regular"). The tube domain T_Γ is given by $\{z = x+iy \in \mathbf{C}^n : y \in \Gamma\}$. We define $\mathcal{H}^p(T_\Gamma)$ to be the class of holomorphic functions F on T_Γ for which

$$\sup_{y\in\Gamma}\left(\int_{\mathbf{R}^n} |F(x+iy)|^p\,dx\right)^{1/p} = \|F\|_{\mathcal{H}^p(T_\Gamma)} < \infty.$$

In analogy with the above we can assert the following. Fix a (proper) subcone Γ_0 of Γ. Then

(a) $F^*(x) = \sup_{|u|\le|y|,\, y\in\Gamma_0} |F(x-u+iy)| \in L^p(\mathbf{R}^n)$ with

$$\|F^*\|_{L^p(\mathbf{R}^n)} \le c_p \|F\|_{\mathcal{H}^p(T_\Gamma)}.$$

(b) $\lim_{y\to 0,\, y\in\Gamma_0} F(x+iy) = F_0(x)$ exists for a.e. $x \in \mathbf{R}^n$.

(c) $\|F\|_{\mathcal{H}^p(T_\Gamma)} = \|F_0\|_{L^p(\mathbf{R}^n)}$.

(d) $\lim_{y\to 0,\, y\in\Gamma_0} F(x+iy) = F^b$ exists in the sense of distributions. Also, $F^b \in H^p(\mathbf{R}^n)$ with $\|F^b\|_{H^p(\mathbf{R}^n)} \approx \|F\|_{\mathcal{H}^p(T_\Gamma)}$.

An analogue of §5.12(e) is as follows. Suppose $\{\Gamma_i\}$ is a finite collection of cones so that $\mathbf{R}^n\backslash\{0\} = \bigcup_i \Gamma_i^*$. Then each $f \in H^p(\mathbf{R}^n)$ can be written as $f = \sum_i F_i^b$, with $F_i \in \mathcal{H}^p(T_{\Gamma_i})$.

Conclusions (a), (b), and (c) are in *Fourier Analysis*, Chapter 3; (d) can be deduced from §4.1. For the decomposition $f = \sum_i F_i^b$, see Carleson [1976]. The proof is based on the fact that one can write $1 = \sum \chi_i(\xi)$, where χ_i is supported in Γ_i, is homogeneous of degree 0, and is smooth away from the origin. If the operator S_i is defined by $\widehat{S_i f}(\xi) = \chi_i(\xi)\cdot\widehat{f}(\xi)$, then, by Chapter 1, §8.19 and by §3.2 of the present chapter, each S_i is a bounded operator from $H^p(\mathbf{R}^n)$ to itself.

5.14 Another proof that $L^1 \cap H^p$ is dense in H^p, $p \le 1$, can be given as follows. If $f \in H^p$ and $u = f * P_t$ is the Poisson integral of f, then (by Theorem 1) $u^* \in L^p(\mathbf{R}^n)$. It follows by Calderón's theorem‡ that u has nontangential limits almost everywhere. Now let $f_\varepsilon = f * P_\varepsilon$, and let u_ε be the Poisson integral of f_ε; as a result $\int_{\mathbf{R}^n}[(u-u_\varepsilon)^*]^p\,dx \to 0$ as $\varepsilon \to 0$, by the dominated convergence theorem. Thus $f_\varepsilon \to f$ in H^p norm, while $f_\varepsilon \in L^1 \cap H^p$, because of Remark 4 in §1.8.

5.15 (a) The following generalized mean-value inequality for harmonic functions is useful in H^p theory: For each q, $0 < q < \infty$, there is a constant c_q so that

$$|u(x)|^q \le c_q \frac{1}{|B|}\int_B |u(y)|^q\,dy,$$

whenever u is harmonic in the ball B, and x is the center of B.

(b) As a consequence, if u is harmonic in $\mathbf{R}^{n+1}_+$, then

$$u^*(x) \le c[M([u^+]^q)(x)]^{1/q}, \qquad \text{all } q > 0.$$

In particular, if Φ is the Poisson kernel, taking $u = f * \Phi$ gives

$$(M^*_\Phi f)(x) \le c[M([M_\Phi f]^q)(x)]^{1/q}.$$

(c) A further consequence is that the condition $u^* \in L^p$ in the Fatou-like theorem in §4.1 can be relaxed to $u^+ \in L^p$. Note however that the even weaker condition $\sup_{t>0}\int_{\mathbf{R}^n} |u(x,t)|^p\,dx < \infty$ does *not* suffice for $p \le 1$.

For (a) see Hardy and Littlewood [1932]; (b) is in C. Fefferman and Stein [1972]. Counterexamples relevant to (c) (for $n = 1$) may be found in Littlewood [1931].

5.16 The theory of conjugate harmonic functions (in §4.2 and §4.3), which described H^p for $p > (n-1)/n$, can be extended to all $p > 0$ as follows. When $n > 1$ and $r \ge 1$ is an integer, consider the tensor product of r copies of $\mathbf{R}^{n+1}$, i.e., $\bigotimes^r \mathbf{R}^{n+1}$. It is a vector space with dimension $(n+1)^r$ and its elements are called *tensors* of rank r; these tensors inherit the usual inner product from $\mathbf{R}^{n+1}$. Let $e_0, \dots, e_n$ be the standard basis of $\mathbf{R}^{n+1}$. If $F \in \bigotimes^r \mathbf{R}^{n+1}$, we say that trace$(F) = 0$ is F is orthogonal to all elements of the form $(\sum_{i=0}^n e_i \otimes e_i) \otimes G$, $G \in \bigotimes^{r-2} \mathbf{R}^{n+1}$. We say that F is symmetric if it is invariant under the induced action of the full permutation group on $n+1$ letters. If $F : \mathbf{R}^{n+1} \to \bigotimes^r \mathbf{R}^{n+1}$, we define

$$(\nabla F)(x) = \sum_{j=0}^n \frac{\partial F}{\partial x_j} \otimes e_j;$$

‡ See *Fourier Analysis*, Chapter 2, or *Singular Integrals*, Chapter 7

∇F is a tensor of rank $r+1$.

Now consider those tensor-valued functions F that are symmetric and have trace zero. The analogue of the Cauchy-Riemann equations (56) is the requirement that ∇F also be symmetric and have trace zero (as a tensor of rank $r+1$).

Starting with a real-valued harmonic function u on $\mathbf{R}^{n+1}_+$, we attach an F as above, so that $\langle F, \bigotimes^r e_0\rangle = u$. Then the analogue of Proposition 2 holds with $p > (n-1)/(n+r-1)$. For the analogue of Proposition 3 (again for $p > (n-1)/(n+r-1)$), one replaces the operators $I, R_1, \ldots, R_n$ by the monomials of degree $\le r$ in the Riesz transforms $R_1, \ldots, R_n$.

For further details, see Stein and G. Weiss [1968] and C. Fefferman and Stein [1972]. The fact that restrictions of the kind $p > (n-1)/n$ are needed for the proposition in §4.2 has been shown by Wolff [1992]. This paper also contains other counterexamples of substantial interest in the theory of harmonic functions.

C. Localized H^p and applications

5.17 For a number of applications it is necessary to be able to localize elements in $H^p(\mathbf{R}^n)$. An immediate obstacle is that if η is a cut-off function (i.e., an element of C_0^∞) and $f \in H^p$, then $\eta \cdot f$ is not necessarily in H^p, because the (global) moment conditions described in §5.4 may be violated. A way to get around this, and other such difficulties, is to consider the (larger) space H^p_{loc}. To define it, let $M^{(1)}_\Phi$ be the truncated version of M_Φ, namely

$$(M^{(1)}_\Phi f)(x) = \sup_{0<t\le 1} |f * \Phi_t(x)|;$$

the definitions of $M^{(1)}_{\mathcal{F}}$ and $u^{(1)}_*$ are similar. One can prove that the inclusions

$$M^{(1)}_\Phi(f) \in L^p(\mathbf{R}^n), \qquad M^{(1)}_{\mathcal{F}}(f) \in L^p(\mathbf{R}^n), \qquad \text{and} \qquad u^{(1)}_*(f) \in L^p(\mathbf{R}^n)$$

are all equivalent; if they hold we say that $f \in H^p_{\text{loc}}(\mathbf{R}^n)$.

Among the obvious assertions that can be made are that $H^p_{\text{loc}} = L^p$ for $1 < p \le \infty$ and $H^p_{\text{loc}} \supset H^p$ for all $p > 0$. Here are some other facts about H^p_{loc}:

(i) If $f \in H^p_{\text{loc}}$ and $\eta \in C_0^\infty$ then $\eta \cdot f \in H^p_{\text{loc}}$.

(ii) More generally, if $\alpha : \mathbf{R}^n \to \mathbf{R}^n$ is a diffeomorphism of a neighborhood of the support of η, then $\eta \cdot (f \circ \alpha)$ also belongs to H^p_{loc}.

(iii) When $p \le 1$, each $f \in H^p_{\text{loc}}$ has an atomic decomposition similar to that given by Theorem 2; the only change is that for atoms associated to balls of radius ≥ 1, no moment condition is required.

(iv) Let $\Phi \in \mathcal{S}$ and suppose that $\widehat{\Phi}(\xi) = 1 + O(|\xi|^N)$ for some large (fixed) N. For $f \in H^p_{\text{loc}}$ write

$$f = f_1 + f_2 = (f - f * \Phi) + (f * \Phi).$$

Then $f_1 \in H^p$.

(v) As a consequence, if $f \in H^p_{\text{loc}}$ has compact support, then $f \in H^p$ modulo a C_0^∞ function.

For further details, see Goldberg [1979].

5.18 Substantial classes of distributions that appear in various problems in analysis turn out to belong (locally) to some H^p. These represent the most common "singularities" that arise in practice. A general way of describing such distributions is as follows.

Suppose we are given a distribution f that for simplicity we take to have compact support. Assume that, associated with it, there is a closed "singular set" S so that, outside of S, the distribution f is given by a locally integrable function $f(x)$ whose size is controlled in terms of the distance of x from S. More precisely, we assume that:

(i) For $x \notin S$, we have $|f(x)| \le c d(x)^{-N}$; here $d(x)$ is the distance of x from S and $N > 0$ is fixed.

We also suppose that S has measure zero in the following strong sense:

(ii) Let $S_\delta = \{x \in \mathbf{R}^n : d(x) \le \delta\}$. Then there is an $r > 0$ so that $|S_\delta| \le c\delta^r$ as $\delta \to 0$.

Notice that there are *countable* sets S that do not enjoy this condition.

On the basis of (i) and (ii), we can conclude that $f \in H^p_{\text{loc}}$ for some $p > 0$.

To prove this, fix a $\Phi \in C^\infty$ that is supported in the unit ball, with $\int \Phi\, dx = 1$. Then if $x \notin S$,

$$|(f * \Phi_t)(x)| \le c d(x)^{-N} \qquad \text{when} \quad t \le \frac{d(x)}{2}.$$

However $|f * \Phi_t(x)| \le A t^{-m}$, if f is a distribution of order m. Thus

$$(M_\Phi f)(x) \le c d(x)^{-\max(N,m)} \quad \text{a.e.},$$

and as a result $f \in H^p_{\text{loc}}$, with $p < r/\max(N,m)$. This because assumption (ii) implies

$$\int_{d(x)\le 1} d(x)^{-q}\, dx < \infty$$

if $q < r$.

An obvious instance of the above is any distribution that is supported on a smooth submanifold of $\mathbf{R}^n$. Deeper examples are given in the next two sections.

5.19 (a) Let P be a polynomial on $\mathbf{R}^n$. Starting from $\operatorname{Re}(s) \ge 0$, the distribution-valued function

$$s \mapsto f_s = |P|^s$$

has a meromorphic continuation to the complex s-plane. The distributions f arising in this way ($f = f_s$ if s is a regular point, f taken to be the residue of f_s if s is a singular point) have the property that $\eta \cdot f \in H^p_{\text{loc}}$, for some p, whenever $\eta \in C_0^\infty$ is a cut-off function.

(b) A variant of this is as follows. Suppose F is real analytic and f is a distribution with $f = 1/F(x)$ outside the set $S = \{x : F(x) = 0\}$. Then $\eta f \in H^p_{\rm loc}$ for some p.

For the meromorphic continuation of the distributions f_s, see Bernstein and Gelfand [1969], Atiyah [1970]. If $S = \{x : P(x) = 0\}$, then it is known, whenever x ranges over a compact set, that $|P(x)| \geq cd(x)^\mu$, for some $\mu \geq 1$; see Lojasiewicz [1959]. Since $\int_{|x|\leq c} |P(x)|^{-r}\,dx < \infty$ for $r < 1/\text{degree}(P)$, and since $|P(x)| \leq cd(x)$ for x in a compact set, it follows that $|S_\delta| \leq c\delta^r$, and §5.18 is applicable.

The distributions in (b) arise in the "division problem". See Hörmander [1958], Lojasiewicz [1959]. The argument in these references also proves that $|F(x)| \geq cd(x)^\mu$. Since, by the Weierstrass preparation theorem, it can be shown that $\int_{|x|\leq c} |F(x)|^{-r}\,dx < \infty$ for some $r > 0$, §5.18 is again applicable.

5.20 Fourier integral distributions fall within the scope of §5.18, and hence they belong to $H^p_{\rm loc}$ for some $p > 0$. A precise formulation of this assertion may be found in Chapter 9, §6.18.

D. Miscellaneous

5.21 The $L^p \to L^q$ inequalities of Hardy, Littlewood, and Sobolev for fractional integration (see Chapter 1, §8.21) extend to the H^p spaces. More precisely, let

$$(I_\alpha f)(x) = \frac{1}{\gamma_\alpha}\int_{\mathbf{R}^n} f(x-y)\,|y|^{\alpha-n}\,dy, \qquad \text{with} \quad \gamma_\alpha = \frac{\pi^{n/2}\,2^\alpha\,\Gamma(\alpha/2)}{\Gamma((n-\alpha)/2)}.$$

The I_α, which can initially be defined for $0 < \alpha < n$ and for (say) f bounded with compact support, can be analytically continued to $0 < \text{Re}(\alpha) \leq n/p$ if, in addition, f satisfies the moment conditions $\int x^\gamma f(x)\,dx = 0$ for $|\gamma| \leq n(p^{-1} - 1)$; here $p \leq 1$. Moreover,

$$\|I_\alpha f\|_{H^q} \leq A_{\alpha,p}\|f\|_{H^p}, \qquad \frac{1}{q} = \frac{1}{p} - \frac{\alpha}{n},$$

as long as $0 < p < q < \infty$. Incidentally, the above range of exponents can be extended in two ways:

(i) It includes $q = \infty$ when $p \leq 1$ (and $\alpha = np$).

(ii) If $\text{Re}(\alpha) = 0$ then $I_\alpha : H^p \to H^p$ for $0 < p < \infty$.

The result for $p \leq 1$ (including its extension to $q = \infty$) follows from its verification on atoms. Conclusion (ii) is a direct consequence of Theorem 4 in §3. See also Hardy and Littlewood [1928] for $n = 1$, Stein and G. Weiss [1960], Krantz [1982a].

5.22 Suppose Ω is a Lipschitz domain in $\mathbf{R}^n$. If $f \in H^p(\mathbf{R}^n)$, $p \leq 1$, and f vanishes in ${}^c\overline{\Omega}$, then f has an atomic decomposition whose atoms are supported in Ω.

The case when Ω is an interval in $\mathbf{R}^1$ was treated in Coifman and G. Weiss [1977a]; the general case is in D. Chang, Krantz, and Stein [1992].

5.23 When F_1 and F_2 belong to the classical (holomorphic) Hardy spaces $\mathcal{H}^p$ and $\mathcal{H}^q$ respectively, then $F_1 \cdot F_2 \in \mathcal{H}^r$, with $r^{-1} = p^{-1} + q^{-1}$. As a result, if $f \in L^p(\mathbf{R}^1)$, $g \in L^q(\mathbf{R}^1)$, with $1 < p < \infty$ and $p^{-1} + q^{-1} = 1$, then $fH(g) + gH(f) \in H^1(\mathbf{R}^1)$; here H is the Hilbert transform. The following generalizations have proved to be of interest in nonlinear methods related to "compensated compactness".

(a) If $f \in L^p(\mathbf{R}^n)$, $g \in L^q(\mathbf{R}^n)$, $1 < p < \infty$, $p^{-1} + q^{-1} = 1$, T is a translation-invariant singular integral of the kind treated in §3.2, and $\tilde{T}$ is the transpose of T, then

$$fT(g) - \tilde{T}(f)g \in H^1(\mathbf{R}^n).$$

(b) If $f_i \in L_1^{p_i}(\mathbf{R}^n)$, $i = 1, \ldots, n$, with $1 < p_i < \infty$, $\sum_1^n p_i^{-1} = 1$, then

$$\det\left\{\frac{\partial f_i}{\partial x_j}\right\} \in H^1(\mathbf{R}^n).$$

(c) If $f_i \in L^p(\mathbf{R}^n)$, $g_i \in L^q(\mathbf{R}^n)$, $1 < p < \infty$, $p^{-1} + q^{-1} = 1$, with $\sum_{i=1}^n \frac{\partial f_i}{\partial x_i} = 0$, and $\frac{\partial g_i}{\partial x_j} = \frac{\partial g_j}{\partial x_i}$ in the sense of distributions, then

$$\sum_1^n f_i g_i \in H^1(\mathbf{R}^n).$$

Result (a) is in Coifman, Rochberg, and G. Weiss [1976]; for (b) and (c) see Coifman, Lions, Meyer, and Semmes [1989], also S. Müller [1990]; generalizations are in Coifman and Grafakos [1992].

5.24 A substantial class of singular integral operators that are of weak-type (1,1) (by the theory of Chapter 1) are, in view of §3, also bounded on H^p. That certain operators satisfy these estimates simultaneously is not accidental and is a consequence of the following general principle.

Suppose T is a bounded operator from $L^q(\mathbf{R}^n)$ to itself, for some $1 < q < \infty$. Assume also that for some $0 < p < 1$, T (initially defined on $H^p \cap L^q$) extends to a bounded operator from H^p to L^p. Then T is of weak-type (1,1).

The idea of the proof is that, whenever $f \in L^1(\mathbf{R}^n)$ and $f = g + \sum b_k$ is a Calderón-Zygmund decomposition (as in §4 of Chapter 1), then the b_k are in H^p, with $n/(n+1) < p < 1$. This result may be viewed as one aspect of the "real" interpolation theory for H^p spaces. For this theory, see Igari [1963] and C. Fefferman, Rivière, and Sagher [1974]. The weak-type conclusion is in Folland and Stein [1982]. The "complex" interpolation theory of H^p spaces is discussed in the next chapter.

5.25 One can strengthen the conclusion of Theorem 3: Under the same hypotheses, T is actually a bounded mapping from $H^1(\mathbf{R}^n)$ to itself.

To prove this, one notes first that, because of Theorem 4, the Riesz transforms R_j, $j = 1, \ldots, n$, satisfy this stronger property. Next one checks that $R_jT = TR_j$ (when acting on atoms) and, finally, invokes the criterion given by the proposition in §4.3.

5.26 Let Φ and Ψ belong to $\mathcal{S}$, with $\int \Phi\, dx = 1$. Suppose that N is a large fixed integer and write $\Phi^{(N+1)}$ for the $N+1$-fold convolution of Φ with itself. Let $\zeta \in C^\infty(\mathbf{R}^1)$ with $\zeta(t) = \dfrac{t^{N-1}}{(N-1)!}$ for t near 0, and $\zeta(t) = 0$ for t near 1. Then the identity

$$\Psi = \int_0^1 [\Phi_t^{(N+1)} * \Psi] \cdot \frac{d^N \zeta}{dt^N}\, dt + (-1)^N \int_0^1 \left(\frac{d^N \Phi_t^{(N+1)}}{dt^N} * \Psi \right) \cdot \zeta(t)\, dt$$

can be taken as a substitute for (8). See Chapter 4 of Folland and Stein [1982].

5.27 The theory of H^p spaces has been extended to the product setting described in §5.20 of Chapter 2. For this the reader should consult Gundy and Stein [1979], S.-Y. A. Chang and R. Fefferman [1980], [1982], [1985], and Merryfield [1985].

Notes

§*1 and* §*3*. For H^p theory in the context of one complex variable, see Zygmund [1959], chapters 7 and 14. Further developments of the one-dimensional theory may be found in Duren [1970], Garnett [1981]. Accounts of the several variable theory up to 1970 are in *Fourier Analysis* and *Singular Integrals*, which also contain references to the literature.

The results in §1, the variant of the Calderón-Zygmund decomposition in §2.1, and the assertions in §3 (with different proofs) are in C. Fefferman and Stein [1972]. A suggestive earlier result is the theorem of Burkholder, Gundy, and Silverstein [1971] that gives a maximal characterization of $H^p(\mathbf{R}^1)$ in terms of Poisson integrals, and is proved using Brownian motion. These considerations were in turn based on the inequalities relating square functions and maximal functions of martingales, as developed in Burkholder and Gundy [1970].

§*2*. The atomic decomposition for H^1 is implicit in the duality of that space with BMO. Its actual formulation and proof (for all H^p, $p \le 1$) was first given (for $\mathbf{R}^1$) by Coifman [1974]; the n-dimensional version is due to Latter [1978]. The proof we give here is based on these papers and some ideas in Macias and Segovia [1979b].

CHAPTER IV

H^1 and BMO

The aim of this chapter is the further study of the Hardy space H^1 in conjunction with its dual space of the functions of bounded mean oscillation. To achieve this we will develop a number of interconnected ideas. These can be summarized as follows.

(i) *Duality of H^1 and BMO.* The fact (due to C. Fefferman) that the space BMO is the dual of H^1 is a central conclusion of this chapter. This theorem crystallized a variety of earlier insights regarding the general principle that H^1 is a natural substitute for L^1, with BMO playing a similar role with respect to L^∞.

(ii) *Size properties of BMO.* These are subsumed in the assertion of John and Nirenberg that contains the following remarkable fact: the L^1 inequalities that define BMO functions imply similar L^p inequalities for every $p < \infty$ and, in fact, imply corresponding exponential bounds.

(iii) *The "sharp" function.* This nonlinear operator (written as $f \mapsto f^\sharp$) is a modification of the maximal function that is particularly adapted to the present context. Indeed, we have $f^\sharp \in L^\infty$ exactly when $f \in$ BMO. There are two further points we want to highlight here. The first is a "duality inequality" involving $f^\sharp$ that already contains the assertion (i), and which shows that L^p control of $f^\sharp$ (for $p < \infty$) implies a corresponding property of f. The second is that the sharp function allows us to express a *pointwise* domination of Tf in terms of f, whenever T is a singular integral operator.

(iv) *Relations with square functions.* Through the use of square functions, one is led to a characterization of BMO in terms of Carleson measures. The resulting equivalence, which is interesting in its own right, has several applications. We mention two: the problem of establishing L^2 inequalities for general Calderón-Zygmund operators (as taken up in Chapter 7), and the existence of a quasi-orthogonal development of BMO functions (given in §4.5 below) that, incidentally, is akin to the types of expansions arising in the theory of wavelets.

(v) *Interpolation of operators.* It turns out that the spaces BMO and H^1 are often indispensable when using the technique of interpolation of analytic families of operators. The reason for this is, at bottom, quite simple: by their nature, singular integral operators are not compatible with L^∞ or L^1 estimates; instead, here one can usefully substitute BMO or H^1. In affecting this idea, the sharp function is the needed tool.

Our presentation of much of the above will follow two independent approaches. Proceeding so will give us a better view of the whole subject and will incidentally allow us to see more than one proof of some of the key results. The first way used passes via the duality inequality, which in turn can be quickly derived from the atomic decomposition described in the previous chapter. The second and alternate path is more elementary. It is based on properties of the "dyadic" maximal operator M^Δ and relative distributional inequalities. The last two notions will serve us again in the next chapter.

1. The space of functions of bounded mean oscillation

1.1 In the same sense that the Hardy space $H^1(\mathbf{R}^n)$ is in many ways a substitute for $L^1(\mathbf{R}^n)$, it will turn out that the space BMO($\mathbf{R}^n$) (the functions of "bounded mean oscillation") is the corresponding natural substitute for the space $L^\infty(\mathbf{R}^n)$ of bounded functions on $\mathbf{R}^n$.

A locally integrable function f will be said to belong to BMO if the inequality

$$\frac{1}{|B|}\int_B |f(x)-f_B|\,dx \le A \tag{1}$$

holds for all balls B; here $f_B = |B|^{-1}\int_B f\,dx$ denotes the mean value of f over the ball B. The inequality (1) asserts that over any ball B, the average oscillation of f is bounded.

The smallest bound A for which (1) is satisfied is then taken to be the norm of f in this space, and is denoted by $\|f\|_{\mathrm{BMO}}$. Let us begin by making some elementary remarks about functions that are in BMO.

1.1.1 Note first that the null elements in the BMO norm are the constants, so that a function in BMO is, strictly speaking, defined only up to an additive constant. Observe also that f would still be in BMO if the definition (1) were extended to allow arbitrary constants c_B in place of the mean values f_B. Indeed, we would then have $|c_B - f_B| \le A$, and $\|f\|_{\mathrm{BMO}} \le 2A$ follows. Similar reasoning shows that an equivalent definition of BMO arises if we replace the family of all balls appearing in (1) by, say, the family of all cubes.

1.1.2 It is trivial that any bounded function is in BMO. The converse is false. A simple example that already typifies some of the essential properties of BMO is given by the function $f(x) = \log|x|$.

To check that this function is in BMO, note that the scaling transformations $f(x) \to f(\delta x)$, $\delta > 0$, map BMO functions to BMO functions, and in fact do not change their norms. Under these scalings, $\log|x|$ is changed by at most an additive constant. Thus to verify (1) it suffices to check the alternative assertions

$$\int_B |\log|x||\,dx \le A, \qquad \int_B |\log|x| - \log|x_0||\,dx \le A$$

where B is a ball of radius 1 centered at x_0. The first inequality holds when $|x_0| \le 1$; the second holds when $|x_0| \ge 1$.

One generalization of this example that is not as straightforward is the fact that $\log|P(x)| \in$ BMO for any polynomial P on $\mathbf{R}^n$; see §6.1. Another family of interesting examples arises from the theory of A_p weights: If $\omega \in A_p$ then $\log\omega \in$ BMO. An appropriate converse also holds; see Chapter 5, §6.2.

1.1.3 It is a simple but useful fact that the space of real-valued BMO functions forms a lattice. If f and g belong to BMO, then so do $\max(f,g)$ and $\min(f,g)$. This follows from the observation that $|f|$ is in BMO whenever f is, which in turn is a consequence of the fact that $||f| - |f_B|| \le |f - f_B|$.

1.1.4 An element of BMO is a function that is "nearly bounded". We shall see below the precise sense in which this is true. A useful related fact is the assertion that $|f(x)|(1+|x|)^{-n-1}$ is integrable on $\mathbf{R}^n$ whenever $f \in \mathrm{BMO}(\mathbf{R}^n)$.[†] Indeed, one can show that

$$\int_{\mathbf{R}^n} |f(x) - f_{B_1}|\,(1+|x|)^{-n-1}\,dx \le c\|f\|_{\mathrm{BMO}}. \tag{2}$$

Here $B_1 = B(0,1)$ is the unit ball centered at the origin. To prove (2), note that by (1)

$$\int_{B_{2^k}} |f(x) - f_{B_{2^k}}|\,dx \le c2^{nk}\|f\|_{\mathrm{BMO}},$$

where B_{2^k} is the ball of radius 2^k centered at the origin. If we compare this inequality with the corresponding one for $B_{2^{k+1}}$, we easily get that $|f_{B_{2^{k+1}}} - f_{B_{2^k}}| \le c\|f\|_{\mathrm{BMO}}$, from which it follows that $|f_{B_{2^k}} - f_{B_1}| \le ck\|f\|_{\mathrm{BMO}}$ for all $k > 0$. This gives

$$\int_{B_{2^k}} |f - f_{B_1}|\,dx \le ck2^{nk}\|f\|_{\mathrm{BMO}},$$

which leads immediately to (2).

[†] Thus it makes sense to take the Poisson integral of a BMO function.

1.2 Duality of H^1 and BMO. Having set down these preliminary observations about the space BMO, we come to one of our main results, namely the fact that BMO is the dual space to H^1. That is, we shall see that each continuous linear functional ℓ on H^1 can be realized as a mapping

$$\ell(g) = \int_{\mathbf{R}^n} f(x)\, g(x)\, dx, \qquad g \in H^1, \tag{3}$$

when suitably defined, where f is a function in BMO.

It is necessary to clarify one point before we proceed. For general $f \in$ BMO and $g \in H^1$, the integral (3) does not converge absolutely (see §6.2), and thus we need to define (3) initially taking g to be in an appropriate dense linear subspace of H^1. For convenience we take this subspace to be the space of all g that are bounded and have compact support, with $\int_{\mathbf{R}^n} g\, dx = 0$. This is of course the same as the subspace of finite linear combinations of H^1 atoms; we know (by Chapter 3, §2.4) that this subspace is dense in H^1. We denote this subspace by H^1_a. For $g \in H^1_a$, the integral (3) converges, and the ambiguity of the BMO element f (i.e., the additive constant) disappears because $\int g\, dx = 0$.

THEOREM 1. (a) *Suppose $f \in$ BMO. Then the linear functional ℓ given by* (3), *initially defined on the dense subspace H^1_a, has a unique bounded extension to H^1 and satisfies*

$$\|\ell\| \leq c\|f\|_{\mathrm{BMO}}.$$

(b) *Conversely, every continuous linear functional ℓ on H^1 can be realized as above, with $f \in$ BMO, and with*

$$\|f\|_{\mathrm{BMO}} \leq c'\|\ell\|.$$

1.2.1 The proof that, for every f in BMO, the linear functional (3) is defined and bounded on H^1 depends on the inequality

$$\left|\int_{\mathbf{R}^n} fg\, dx\right| \leq c\|f\|_{\mathrm{BMO}}\|g\|_{H^1}, \tag{4}$$

for $f \in$ BMO and g in the dense subspace $H^1_a \subset H^1$.

The above inequality is actually a simple consequence of the atomic decomposition for H^1, once we prove (4) under the extra assumption that f is bounded. With this assumption, we can write

$$\int_{\mathbf{R}^n} fg\, dx = \sum_k \lambda_k \int_{\mathbf{R}^n} f(x) a_k(x)\, dx,$$

where $g = \sum \lambda_k a_k$ is an atomic decomposition for $g \in H^1$, since the sum $\sum \lambda_k a_k$ converges in the $L^1(\mathbf{R}^n)$ norm (for the atomic decomposition, see §2.2 in Chapter 3).

Since the a_k have vanishing mean value, we can write

$$\int f(x) a_k(x)\, dx = \int_{B_k} [f(x) - f_{B_k}] a_k(x)\, dx,$$

where a_k is supported in B_k. Using the fact that $|a_k(x)| \leq |B_k|^{-1}$, we have that

$$\left|\int f(x) g(x)\, dx\right| \leq \sum_k \frac{|\lambda_k|}{|B_k|} \int_{B_k} |f(x) - f_{B_k}|\, dx \leq \sum |\lambda_k| \cdot \|f\|_{\mathrm{BMO}}.$$

The inequality (4) is therefore proved for bounded f (and all $g \in H^1$), where the constant of course does not depend on the L^∞ norm of f.

To conclude the proof of (4), we restrict our attention to $g \in H^1_a$, and assume that $f \in$ BMO is real-valued. We then replace f by $f^{(k)}$, where

$$f^{(k)}(x) = \begin{cases} -k & \text{if } f(x) \leq -k, \\ f(x) & \text{if } -k \leq f(x) \leq k, \\ k & \text{if } k \leq f(x). \end{cases}$$

Then since $\|f^{(k)}\|_{\mathrm{BMO}} \leq c\|f\|_{\mathrm{BMO}}$ (see the remark §1.1.3), we get by the case just proved that $|\int f^{(k)} g\, dx| \leq c\|f\|_{\mathrm{BMO}}\|g\|_{H^1}$. Finally, since $f^{(k)} \to f$ a.e. as $k \to \infty$, the dominated convergence theorem‡ gives (4). We have therefore seen that each $f \in$ BMO gives a bounded linear functional on the dense subspace H^1_a, and thus on all of H^1.

1.2.2 To prove the converse we test the linear functional ℓ on appropriate atoms. For this purpose fix a ball $B \subset \mathbf{R}^n$, and let L^2_B denote the space of all square integrable functions supported on B. This space has the norm $\|\cdot\|_{L^2}$ given by $\|g\|_{L^2} = (\int_B |g(x)|^2\, dx)^{1/2}$. Let $L^2_{B,0}$ denote its closed subspace of functions with integral zero. Note that every element $g \in L^2_{B,0}$ is a multiple of a "generalized atom" for H^1 and that $\|g\|_{H^1} \leq c|B|^{1/2}\|g\|_{L^2}$.†

Thus if ℓ is a given linear functional on H^1 (which we will assume has norm ≤ 1), then ℓ extends to a linear functional on $L^2_{B,0}$ with norm at most $c|B|^{1/2}$. By the Riesz representation theorem for the Hilbert space $L^2_{B,0}$, there exists an element $F^B \in L^2_{B,0}$ so that

$$\ell(g) = \int_B F^B(x)\, g(x)\, dx, \qquad \text{if } g \in L^2_{B,0}, \tag{5}$$

with

$$\left(\int_B |F^B(x)|^2\, dx\right)^{1/2} \leq c|B|^{1/2}. \tag{6}$$

‡ It is here that we use that $g \in H^1_a$.

† See §2.4.3 in Chapter 3.

Hence for each ball B, we get such a function F^B. We want to have a single function f so that, on each ball B, f differs from F^B by a constant. To construct this f, observe that if $B_1 \subset B_2$ are balls, then $F^{B_1} - F^{B_2}$ is constant on B_1. Indeed, both F^{B_1} and F^{B_2} give the same functional on $L^2_{B_1,0}$, so they must differ by a constant on B_1. We can modify F^B, replacing it with $f^B = F^B + c_B$, where c_B is a constant chosen so that f^B has integral zero on the unit ball centered at the origin. It follows that $f^{B_1} = f^{B_2}$ on B_1, if $B_1 \subset B_2$. Therefore we can unambiguously define f on all of $\mathbf{R}^n$ by taking $f(x) = f^B(x)$ for $x \in B$.

Observe that

$$\frac{1}{|B|}\int_B |f(x) - c_B|\,dx \leq \left(\frac{1}{|B|}\int_B |f(x) - c_B|^2\,dx\right)^{1/2}$$
$$= \left(\frac{1}{|B|}\int_B |F^B|^2\,dx\right)^{1/2} \leq c,$$

by (6); therefore $f \in \mathrm{BMO}$ with $\|f\|_{\mathrm{BMO}} \leq c$. Also, by (5),

$$\ell(g) = \int f(x)\,g(x)\,dx,$$

whenever $g \in L^2_{B,0}$ for some B, in particular this representation holds for all $g \in H^1_a$. The converse conclusion (b) of the theorem is therefore proved.

1.3 The John-Nirenberg inequality. We have seen in the proof of Theorem 1 that every $f \in \mathrm{BMO}$ actually satisfies a stronger version of the defining inequality (1), namely

$$\frac{1}{|B|}\int_B |f - f_B|^2\,dx \leq c'.$$

Indeed, already contained in that argument is the fact that a corresponding inequality holds for every $p < \infty$, and moreover that there is a limiting version involving the exponential integrability of f.

COROLLARY. *Suppose that f is in BMO. Then*

(a)*For any $p < \infty$, f is locally in L^p, and*

$$\frac{1}{|B|}\int_B |f - f_B|^p\,dx \leq c_p\|f\|^p_{\mathrm{BMO}}, \tag{7}$$

for all balls B.

(b) *There exist positive constants c_1 and c_2 so that, for every $\alpha > 0$ and every ball B,*

$$|\{x \in B : |f(x) - f_B| > \alpha\}| \leq c_1 e^{-c_2\alpha/\|f\|_{\mathrm{BMO}}} \cdot |B|. \tag{8}$$

1.3.1 Proof. The proof of inequality (7) is a reprise of the argument given in §1.2.2. For any $q > 1$, denote by L^q_B the space of L^q functions with support in B, and let $L^q_{B,0}$ be the subspace of functions whose integral is zero. The argument in §1.2.4 of the previous chapter (see also §5.7 there) shows that each $g \in L^q_{B,0}$ is also in H^1, and

$$\|g\|_{H^1} \leq c_q|B|^{1-1/q}\|g\|_{L^q}, \tag{9}$$

with $c_q = O(1/(q-1))$, as $q \to 1$.‡

Next, if $f \in \mathrm{BMO}$ and $\|f\|_{\mathrm{BMO}} \leq 1$, then by the above theorem the linear functional $g \mapsto \ell(g)$, given by

$$\ell(g) = \int_{\mathbf{R}^n} fg\,dx, \tag{10}$$

is bounded on $L^q_{B,0}$ with $\|\ell\| \leq c \cdot c_q|B|^{1-1/q}$. By the Hahn-Banach theorem, ℓ extends to a linear functional $\tilde{\ell}$ on all of L^q_B with $\|\tilde{\ell}\| = \|\ell\|$. Thus, by (L^p, L^q) duality, there exists an $F^B \in L^p_B$, $1/p + 1/q = 1$, with

$$\left(\frac{1}{|B|}\int_B |F^B|^p\,dx\right)^{1/p} \leq c \cdot c_q, \tag{11}$$

so that

$$\tilde{\ell}(g) = \int_B F^B g\,dx, \quad \text{if } g \in L^q_{B,0}. \tag{12}$$

However, by (10) and (12), we see that $F^B = f - c_B$ in B, where c_B is an appropriate constant. Using (11) and the fact that $c_q \leq c/(q-1)$ as $q \to 1$ (with $q/(q-1) = p$), we get (with a possibly different constant c)

$$\frac{1}{|B|}\int_B |f - f_B|^p\,dx \leq (cp)^p, \quad 1 \leq p < \infty. \tag{13}$$

This is conclusion (7) of the corollary if we take into account the normalization imposed on f. To pass to the second conclusion, we apply Chebysheff's inequality to (13) and obtain

$$|\{x \in B : |f(x) - f_B| > \alpha\}| \leq (cp)^p\alpha^{-p}|B| \tag{14}$$

‡ This is a consequence of the maximal inequality $\|Mg\|_q \leq A_q\|g\|_q$ and the bound for A_q given in Chapter 1, §8.1.3.

for $0 < \alpha < \infty$, $1 \leq p < \infty$.

We exploit (14) by choosing p in terms of α.† If $\alpha \geq 2c$, we take $p = \alpha/2c \geq 1$. Then (14) gives

$$|\{x \in B : |f(x) - f_B| > \alpha\}| \leq (1/2)^p |B| = e^{-c_1\alpha}|B|$$

with $c_1 = (2c)^{-1}\log 2$. However, if $\alpha \leq 2c$, then $e^{-c_1\alpha} \geq e^{-c_1(2c)} = 1/2$, and

$$|\{x \in B : |f(x) - f_B| > \alpha\}| \leq 2e^{-c_1\alpha}|B|$$

in that range of α. Altogether then, if we drop the normalization on f by replacing f by $f/\|f\|_{\mathrm{BMO}}$, we see that (8) is established with $c_1 = 2$ and $c_2 = (2c)^{-1}\log 2$.

Observe that, as a consequence of (8), we obtain

$$\frac{1}{|B|}\int_B e^{\mu|f-f_B|}\,dx \leq c(\mu, f),$$

for all $\mu < c_2$, whenever f is in BMO.

2. The sharp function

The definition and properties of BMO considered above lead us naturally to study the sharp function $f^\sharp$, associated to any locally integrable function f. It is defined by

$$f^\sharp(x) = \sup_{x \in B} \frac{1}{|B|}\int_B |f(y) - f_B|\,dy, \tag{15}$$

where the supremum is taken over all balls B containing x.

Note that a function f is in BMO exactly when $f^\sharp$ is a bounded function. This observation illustrates the fact that sometimes significant aspects of f are most directly expressed in terms of $f^\sharp$. Pursuing this point, we may ask what happens when $f^\sharp \in L^p$, or how the duality of H^1 with BMO might be quantitatively expressed in terms of $f^\sharp$. To begin with, we intend to establish and exploit the following *duality inequality*:

$$\left|\int_{\mathbf{R}^n} f(x)g(x)\,dx\right| \leq c\int_{\mathbf{R}^n} f^\sharp(x)\,\mathcal{M}g(x)\,dx \tag{16}$$

for appropriate f and g.

We use the notation $\mathcal{M}$ to stand for $\mathcal{M}_{\mathcal{F}}$, as defined in §1.2 of the previous chapter; here $\mathcal{F}$ is *any* fixed finite collection of seminorms.

† Note that this part of the argument shows that any F with $\|F\|_{L^p} \leq cp$, $p \to \infty$, belongs to the exponential class.

PROPOSITION. *The duality inequality* (16) *holds whenever* $g \in H^1$ *and* f *is bounded.*

2.1 The inequality (16), like the earlier version (4) which it extends, is a consequence of the ideas used in the atomic decomposition for H^1. Recall that according to §2.3.1 of the previous chapter, any $g \in H^1$ (called f there) can be written as $g = \sum_{j,k} \lambda_{j,k} a_k^j$, where the a_k^j are H^1 atoms supported in balls B_k^j, and $\lambda_{j,k} = c2^j|B_k^j|$. The balls B_k^j contain cubes Q_k^j, with $|B_k^j| \leq c|Q_k^j|$. Finally, for fixed j, the cubes $\{Q_k^j\}_k$ are disjoint, and

$$\bigcup_k Q_k^j = \{x : \mathcal{M}g > 2^j\}.$$

Since $\sum \lambda_{j,k} a_k^j$ converges to g in L^1 norm, we have

$$\int fg\,dx = \sum \lambda_{j,k}\int_{B_k^j} f(x)\,a_k^j(x)\,dx,$$

while

$$\int_{B_k^j} f(x) a_k^j(x)\,dx = \int_{B_k^j} [f(x) - f_{B_k^j}]a_k^j(x)\,dx.$$

However, $|a_k^j| \leq |B_k^j|^{-1}$ and

$$\frac{1}{|B_k^j|}\int_{B_k^j} |f(x) - f_{B_k^j}|\,dx \leq f^\sharp(\bar{x}), \qquad \text{if } \bar{x} \in Q_k^j.$$

Thus

$$\left|\int_{B_k^j} f(x)\,a_k^j(x)\,dx\right| \leq \frac{1}{|Q_k^j|}\int_{Q_k^j} f^\sharp(x)\,dx.$$

Inserting the last inequality and summing gives

$$\left|\int fg\,dx\right| \leq \sum_{j,k} \frac{\lambda_{j,k}}{|Q_k^j|}\int_{Q_k^j} f^\sharp(x)\,dx = c\sum_j 2^j \int_{\mathcal{M}g > 2^j} f^\sharp(x)\,dx,$$

because the disjoint union $\bigcup_k Q_k^j$ equals $\{\mathcal{M}g > 2^j\}$. Interchanging the summation and integration then shows that

$$\left|\int f\,g\,dx\right| \leq c\int f^\sharp(x)\,\mathcal{M}g(x)\,dx,$$

proving the proposition.

2.2 It is obvious that the sharp function is pointwise dominated by the standard maximal function; from this it follows that $\|f^\sharp\|_{L^p} \leq c_p\|f\|_{L^p}$, if $1 < p \leq \infty$. An important consequence of the proposition just proved is the following "converse" inequality.

THEOREM 2. *Suppose* $1 < p < \infty$. *If* $f \in L^p(\mathbf{R}^n)$, *then the following inequality holds:*

$$\|f\|_{L^p} \leq A_p\|f^\sharp\|_{L^p}. \tag{17}$$

Remark. More generally, suppose that $f \in L^{p_0}$, for some $1 < p_0 < \infty$, and that $f^\sharp \in L^p$; then f is also in L^p and (17) holds.

To prove this last statement (and hence Theorem 2), we show that if q_0 is the conjugate exponent to p_0, $1/q_0 + 1/p_0 = 1$, then the basic duality inequality (16) holds for $f \in L^{p_0}$ and $g \in L^{q_0}$.

In fact, given $f \in L^{p_0}$, we can find a sequence $\{f_k\}$, so that $|f_k| \leq |f|$, $f_k \to f$ a.e., and each f_k is bounded. Then clearly $f_k \to f$ in L^{p_0} norm, while the maximal theorem shows that $f_k^\sharp \to f^\sharp$ a.e. and dominatedly; thus $f_k^\sharp \to f^\sharp$ in L^{p_0} norm. Also we can find a sequence $\{g_k\}$ so that each g_k has compact support, $g_k \to g$ a.e., the g_k are dominated by a fixed function in L^{q_0}, and, most importantly, $\int g_k\,dx = 0$ for all k. The last fact insures that each g_k belongs to H^1. To see this, write $\tilde{g}_k = g\chi_{|x|\leq k}$. Then by Hölder's inequality, $|I_k| \leq ck^{n/p_0}$, where $I_k = \int \tilde{g}_k\,dx$. Let

$$g_k = \tilde{g}_k - I_k|B_{2^k}|^{-1}\chi_{B_{2^k}} = \tilde{g}_k - r_k,$$

where B_{2^k} is the ball of radius 2^k centered at the origin. Note that $\int g_k\,dx = 0$ and that $g_k \to g$ almost everywhere; also the $\{g_k\}$ have the common majorant $|g| + r$, where

$$r(x) = \sum_{k=1}^{\infty}|r_k(x)| = \sum_{k=1}^{\infty}|I_k|\cdot|B_{2^k}|^{-1}\chi_{B_{2^k}}(x) \leq c\frac{\log(1+|x|)^{n/p_0}}{(1+|x|)^n},$$

since $|I_k| \leq ck^{n/p_0}$. Thus $|g| + r \in L^{q_0}$, since $q_0 > 1$, and our assertions about the sequence g_k have been verified. By the previous proposition, then

$$\left|\int f_k(x)\,g_k(x)\,dx\right| \leq c\int f_k^\sharp(x)\,\mathcal{M}g_k(x)\,dx.$$

A passage to the limit therefore gives

$$\left|\int fg\,dx\right| \leq c\int f^\sharp\mathcal{M}g\,dx,$$

showing that (16) holds whenever $f \in L^{p_0}$, $g \in L^{q_0}$. Using the fact that $\|\mathcal{M}g\|_{L^q} \leq c\|Mg\|_{L^q} \leq c_q\|g\|_{L^q}$, one obtains

$$\left|\int f\,g\,dx\right| \leq c_q\|f^\sharp\|_{L^p}\cdot\|g\|_{L^q},$$

whenever $g \in L^q \cap L^{q_0}$, where $1/p + 1/q = 1$. Now take the supremum of the left side of this inequality over all such g with $\|g\|_{L^q} \leq 1$. The result is that $f \in L^p$ and (17) holds; Theorem 2 (and its generalization) is therefore proved.

3. An elementary approach and a dyadic version

Our treatment above of the duality of H^1 and BMO, and the related inequalities for BMO, had the virtue of following rather immediately from results of the previous chapter. However, establishing these facts (as well as the re-examination of their proofs needed for the sharp function) required the full panoply of techniques developed for the atomic decomposition, including the grand maximal function. It would therefore be desirable to have a more direct and elementary approach to some of the results of the previous sections of this chapter.

The approach we present now has several advantages of its own and is based on two ideas. The first is to consider "dyadic" analogues of the maximal function and sharp function. The second is to relate them by a "relative distributional inequality". Incidentally, similar ideas will also be very useful in the next chapter when we consider weighted inequalities. We turn first to the dyadic version of the maximal function.

3.1 The *dyadic maximal function* is defined using a family of meshes $\{\mathcal{M}_k\}$. For each integer k, $\mathcal{M}_k$ denotes the family of closed cubes whose sides have length 2^{-k}, and whose vertices are members of the lattice of points of the form $(m_1/2^k, \ldots, m_n/2^k)$ with the m_j being arbitrary integers. Observe that $\mathcal{M}_k = 2^{-k}\mathcal{M}_0$, where $\mathcal{M}_0$ is the mesh of unit cubes whose vertices have integral coordinates. Also, each cube in $\mathcal{M}_k$ gives rise to 2^n cubes in $\mathcal{M}_{k+1}$ by bisecting the sides.

We shall refer to a cube belonging to any of the meshes as a *dyadic cube*. Thus the collection of dyadic cubes is simply $\cup_k\mathcal{M}_k$. We say that two dyadic cubes intersect if their interiors intersect, and that they are disjoint if their interiors are disjoint.

The dyadic maximal operator M^Δ is then defined by

$$(M^\Delta f)(x) = \sup_{x\in Q}\frac{1}{|Q|}\int_Q|f(y)|\,dy, \tag{18}$$

where the supremum is over all dyadic cubes Q that contain x.

What follows is the principal fact about the dyadic maximal function; it is also the basis of the classical Calderón-Zygmund decomposition.‡

‡ Compare with *Singular Integrals*, pp. 17–19.

LEMMA 1. *Let f be a locally integrable function on $\mathbf{R}^n$, and let α be a positive constant so that*

$$\Omega_\alpha = \{x : M^\Delta f(x) > \alpha\}$$

has finite measure. Then Ω_α may be written as a disjoint union of dyadic cubes $\{Q_j\}$ with

(i) $\alpha < |Q_j|^{-1} \int_{Q_j} |f(x)|\, dx$

and

(ii) $|Q_j|^{-1} \int_{Q_j} |f(x)|\, dx \leq 2^n \alpha,$

for each cube Q_j. This has the immediate consequences:

(iii) $|f(x)| \leq \alpha$ *for a.e.* $x \in {}^c\bigcup_j Q_j$

and

(iv) $|\Omega_\alpha| \leq \alpha^{-1} \int_{\mathbf{R}^n} |f(x)|\, dx$.†

Proof. Given any two intersecting dyadic cubes, one of the two must be contained in the other. For each $x \in \Omega_\alpha$, there is a maximal dyadic cube Q so that $x \in Q$ and

$$\frac{1}{|Q|} \int_Q |f(y)|\, dy > \alpha,$$

for otherwise Ω_α would have infinite measure. The above remark shows that any two of these maximal cubes are disjoint. We let $\{Q_j\}$ denote this collection of maximal dyadic cubes. Then it is clear that

$$\Omega_\alpha = \bigcup_j Q_j$$

and that (i) holds automatically. Moreover, since Q_j was maximal, if $\tilde{Q}_j$ denotes the (unique) next larger dyadic cube containing Q_j, then

$$|\tilde{Q}_j|^{-1} \int_{\tilde{Q}_j} |f| \leq \alpha,$$

which implies conclusion (ii). Conclusion (iv) follows by adding the inequalities (i).

† This conclusion is vacuous when f is not integrable on all of $\mathbf{R}^n$. Alternatively, we may replace the right side of (iv) by $\alpha^{-1} \int_{\Omega_\alpha} |f|\, dx$, as follows from (i).

3.2 Since the proof given above depends only on the simple inclusion properties of dyadic cubes, the lemma actually holds in a much wider setting. In fact, conclusions of the kind (i), (iii), and (iv) are valid in the context of general martingales. Only conclusion (ii) is of a more restricted nature; for it is a doubling property (that involves the meshes $\mathcal{M}_k$ and the underlying measure). See §6.22. For a comparison between M^Δ and the usual maximal operator M, see §6.24.

3.3 There are two straightforward observations that we want to make about the above lemma that will be useful in its application below. First, suppose that Q_0 is a fixed dyadic cube and that f is supported in Q_0. If $\alpha_0 = |Q_0|^{-1} \int_{Q_0} |f|\, dx$, then, when we apply Lemma 1 to f with $\alpha \geq \alpha_0$, the cubes $\{Q_j\}$ are all contained in Q_0; in other words, the decomposition guaranteed by the lemma takes place entirely in Q_0.

Second, if $\alpha_1 > \alpha_2$, then

$$\{x : M^\Delta f(x) > \alpha_1\} \subset \{x : M^\Delta f(x) > \alpha_2\}$$

and, by the maximality of the cubes, each dyadic cube in the decomposition at altitude α_1 is contained in a dyadic cube in the decomposition at height α_2.

3.4 In analogy with the dyadic maximal function, we define a similar version of the sharp function, which we denote by $f^\sharp_\Delta$. It is given by

$$f^\sharp_\Delta(x) = \sup_{x \in Q} \frac{1}{|Q|} \int_Q |f - f_Q|\, dx,$$

where the supremum is taken over all dyadic cubes Q containing x.

The relation between $f^\sharp_\Delta$ and M^Δ is given by a relative distributional inequality, of which the following is typical:

$$|\{x : M^\Delta f(x) > \alpha, f^\sharp_\Delta(x) \leq c\alpha\}| \leq 2^{n+1} c |\{x : M^\Delta f(x) > \alpha/2\}|, \tag{19}$$

for all $\alpha, c > 0$.

Some motivation for inequalities of this type, and the L^p estimates they imply, will now be discussed.

3.5 Relative distributional inequalities. The simplest hypothesis that gives the inequality

$$\int [F(x)]^p\, d\mu(x) \leq A \int [G(x)]^p\, d\mu(x), \tag{20}$$

for nonnegative F and G, is that F is pointwise dominated by a constant multiple of G. A trivial generalization of this occurs when F and G are related by the "relative" inequality $F(x) \leq c_1 G(x) + c_2 F(x)$, where $c_2 < 1$. However, in most situations of interest we are not fortunate enough to have pointwise inequalities at our disposal; sometimes then bounds relating distribution functions can be used. Thus an inequality of the form‡

$$\mu\{x : F(x) > \alpha\} \leq c_1 \mu\{x : G(x) > c_2 \alpha\},$$

for all $\alpha > 0$ (with fixed positive constants c_1 and c_2), easily implies (20). These examples are subsumed under a "relative distributional inequality" of the form

$$\mu\{x : F(x) > \alpha;\ G(x) \leq c\alpha\} \leq a\, \mu\{x : F(x) > b\alpha\}, \quad \text{all } \alpha > 0, \tag{21}$$

where a, b, and c are appropriate constants.

LEMMA 2. *Suppose the nonnegative functions F and G satisfy the relative distributional inequality* (21). *Assume also that for some p, $0 < p < \infty$, we have $\|F\|^p_{L^p(d\mu)} = \int_{\mathbf{R}^n} [F(x)]^p \, d\mu(x) < \infty$. Then* (20) *holds with $A = A(a, b, c, p)$, whenever $a < b^p$.*

Recall that

$$\|F\|^p_{L^p(d\mu)} = p \int_0^\infty \alpha^{p-1} \mu\{x : F(x) > \alpha\} \, d\alpha$$

and notice that (21) implies immediately that

$$\mu\{x : F(x) > \alpha\} \leq a\, \mu\{x : F(x) > b\alpha\} + \mu\{x : G(x) > c\alpha\}.$$

Multiply both sides of this inequality by $p\alpha^{p-1}$ and integrate in α. After a change of variables, the result is

$$\|F\|^p_{L^p(d\mu)} \leq ab^{-p} \, \|F\|^p_{L^p(d\mu)} + c^{-p} \, \|G\|^p_{L^p(d\mu)}.$$

When $ab^{-p} < 1$ and $\|F\|^p_{L^p(d\mu)} < \infty$, we can bring the second integral to the left side. We then obtain (20) with $A = \dfrac{c^{-p}}{(1 - ab^{-p})}$, proving the lemma.

Remark. When $b \leq 1$ (which occurs in the applications below), the requirement that $\|F\|_{L^p(d\mu)} < \infty$, can be relaxed to the assumption that $\|F\|_{L^{p_0}(d\mu)} < \infty$ for some p_0, with $p_0 \leq p$.

‡ An example is the inequality (25) in §2.5 of Chapter 2.

To see this, let

$$I_R = p \int_0^R \alpha^{p-1} \mu\{x : F(x) > \alpha\} \, d\alpha.$$

Notice that, by our assumption, I_R is finite for each R. Now by proceeding as above,

$$I_R \leq ab^{-p} I_{Rb} + c^{-p} \|G\|^p_{L^p(d\mu)}.$$

Since $I_{Rb} \leq I_R$ (because $b \leq 1$), we get $I_R \leq A\|G\|^p_{L^p(d\mu)}$; letting $R \to \infty$ gives the desired conclusion.

Without the condition that $\|F\|_{L^{p_0}(d\mu)} < \infty$, the conclusion may fail. A simple example arises on $\mathbf{R}^1$ when $F(x) = |x|^{-1}$, $d\mu = dx$, and $G(x) \equiv 0$.

3.6 The inequality relating $f^\sharp_\Delta$ and $M^\Delta f$. We shall now prove the distributional inequality (19) in a slightly more general form:

PROPOSITION. *For a locally integrable function f, and for b and c positive with $b < 1$, we have the inequality*

$$|\{x : M^\Delta f(x) > \alpha, f^\sharp_\Delta(x) \leq c\alpha\}| \leq a|\{x : M^\Delta f(x) > b\alpha\}|, \tag{22}$$

for all $\alpha > 0$, where $a = 2^n c/(1 - b)$.

In proving (22) we may assume that the set $\{x : M^\Delta f(x) > b\alpha\}$ has finite measure, for otherwise the inequality is vacuous. This implies, by the lemma in §3.1, that this set is the union of disjoint maximal cubes $\{Q_j\}$. We let $Q = Q_j$ denote one of these cubes, and note that it suffices to show that

$$|\{x \in Q : M^\Delta f(x) > \alpha, f^\sharp_\Delta(x) \leq c\alpha\}| \leq a|Q|, \tag{23}$$

because (22) follows from adding these inequalities.

Now let $\tilde{Q} \supset Q$ be the parent of Q; then by the maximality of Q we have $|f|_{\tilde{Q}} \leq b\alpha$. So for all $x \in Q$ for which $M^\Delta f(x) > \alpha$, it follows that $M^\Delta(f\chi_Q)(x) > \alpha$, and also that $M^\Delta[(f - f_{\tilde{Q}})\chi_Q](x) > (1 - b)\alpha$. By the weak-type inequality (conclusion (iv) of the lemma with f replaced by $(f - f_{\tilde{Q}})\chi_Q$, and α replaced by $(1 - b)\alpha$), we have that the measure of the left side of (23) is at most

$$\frac{1}{(1-b)\alpha} \int_Q |f - f_{\tilde{Q}}| \, dx \leq \frac{1}{(1-b)\alpha} \int_{\tilde{Q}} |f - f_{\tilde{Q}}| \, dx \leq \frac{|\tilde{Q}|}{(1-b)\alpha} \inf_{x \in Q} f^\sharp_\Delta(x).$$

The last quantity is majorized by

$$\frac{|\tilde{Q}|}{(1-b)\alpha} \cdot c\alpha = \frac{2^n c}{1-b} |Q|,$$

if the set in question is not empty (i.e., if there is an $x \in Q$ so that $f^\sharp_\Delta(x) \leq c\alpha$). This gives (23), and hence (22), which proves the proposition.

Corollary 1. *Suppose f is a locally integrable function for which $f^\sharp_\Delta$ belongs to L^p, $1 < p < \infty$. If we also assume that f belongs to some L^{p_0}, $p_0 \le p$, then f is in L^p and*

$$\|M^\Delta f\|_{L^p} \le A_p \|f^\sharp_\Delta\|_{L^p}. \tag{24}$$

For the proof of the corollary we merely need to invoke (19) and Lemma 2 (and the remark that follows it), with $a = 2^{n+1}c$, $b = 1/2$, and c chosen to be $2^{-n-2}2^{-p}$. The desired result is therefore obtained with $A_p^p = c^{-p}/(1-ab^{-p}) = 2(2^{n+2}2^p)^p$.

Remark. The result (24) is stronger than the corresponding estimate (17) obtained in §2.2 for $f^\sharp$ on two grounds. First, as is easily seen, $f^\sharp_\Delta \le cf^\sharp$, while $f^\sharp_\Delta$ can be essentially smaller than $f^\sharp$. For instance, in $\mathbf{R}^1$, if $f(x) = \log|x|$ for $x > 0$, and $f(x) = 0$ for $x \le 0$, then $f^\sharp_\Delta$ is bounded, and in fact vanishes for $x < 0$; however $f^\sharp(x) \ge c|\log|x||$. Second, in view of the fact that $M^\Delta f \ge |f|$, we get a further improvement near $p = 1$, since the constant A_p remains bounded as $p \to 1$.

3.7 One can also obtain in this way a version of the John-Nirenberg inequality in §1.3, which is stronger because, among other reasons, it is localized and is restricted to dyadic cubes. We suppose that f is an integrable function given on a dyadic cube Q_0. We shall assume that f is in *dyadic* BMO on this cube (with norm 1) in the sense that

$$\frac{1}{|Q|}\int_Q |f - f_Q|\,dx \le 1, \tag{25}$$

holds for all dyadic subcubes $Q \subset Q_0$.

Corollary 2. *There exist positive constants c_1 and c_2 so that whenever f satisfies* (25), *then for all $\alpha > 0$*

$$|\{x \in Q_0 : |f(x) - f_{Q_0}| > \alpha\}| \le c_1 e^{-c_2\alpha}|Q_0|. \tag{26}$$

Proof. We set $f_1 = (f - f_{Q_0})\chi_{Q_0}$; then $(f_1)^\sharp_\Delta(x) \le 1$ for all x, by our assumption, since any dyadic cube that intersects Q_0 must either be contained in it, or contain it. Let

$$\lambda(\alpha) = |\{x : M^\Delta f_1(x) > \alpha\}|.$$

Then by the more general distribution inequality (22), we have $\lambda(\alpha) \le a\lambda(b\alpha)$, if we choose $c = 1/\alpha$. Next we restrict our attention to $\alpha \ge 2^{n+1}$, and fix b so that $b\alpha = \alpha - 2^{n+1}$, i.e., $b = 1 - 2^{n+1}/\alpha$. Since $a = 2^n c/(1-b)$, we have $a = 1/2$, and therefore

$$\lambda(\alpha) \le \frac{1}{2}\cdot\lambda(\alpha - 2^{n+1}), \quad \text{for } \alpha \ge 2^{n+1}.$$

This easily implies that in this range

$$\lambda(\alpha) \le e^{-c_2\alpha}\lambda(1) \le e^{-c_2\alpha}|Q_0|,$$

where $c_2 = 2^{-n-1}\log 2$; the estimate for $\lambda(1)$ following directly from the maximal inequality (iv) in the lemma in §3.1.

Since

$$|\{x \in Q_0 : |f(x) - f_{Q_0}| > \alpha\}| \le |\{x \in \mathbf{R}^n : M^\Delta f_1 > \alpha\}| = \lambda(\alpha),$$

we have that the quantity on the left is at most $e^{-c_2\alpha}|Q_0|$ when $\alpha \ge 2^{n+1}$. However when $0 < \alpha \le 2^{n+2}$, this quantity is obviously not greater than $|Q_0| = c_1 e^{-c_2\alpha}|Q_0|$, where $c_1 = e^{c_2 2^{n+1}} = 2$. Thus (26) is proved with $c_1 = 2$ and $c_2 = 2^{-n-1}\log 2$.

4. Further properties of BMO

We develop further the theory of BMO. We first show that many singular integrals map L^∞ to BMO and then generalize this to a pointwise estimate involving the sharp function. Later, we study the relationship of square functions and orthogonality to BMO.

As we saw in the previous chapter, a large class of singular integrals map H^1 to L^1; the duality of H^1 and BMO then implies that there should be a corresponding assertion concerning the boundedness from L^∞ to BMO. Moreover it can be reasonably expected that this conclusion could be formulated as an estimate involving sharp functions. To these and related matters we now turn.

4.1 Singular Integrals. We consider the (translation-invariant) singular integrals that were the subject of Chapter 1, §6.2 and Chapter 2, §3. Briefly, we are concerned with a linear operator T, initially defined and bounded from $L^2(\mathbf{R}^n)$ to itself, to which is associated a bounded function $m(\xi)$ and a distribution K, so that m is the Fourier transform of K; also K, away from the origin, is given by a locally integrable function $K(x)$. The connection between m, K and T is as follows: first, $\widehat{Tf} = m\widehat{f}$, whenever $f \in L^2$; second, whenever f has compact support, then

$$(Tf)(x) = \int_{\mathbf{R}^n} K(x-y)\,f(y)\,dy,$$

for almost every x outside the support of f. Finally the crucial quantitative assumptions on T are given by the existence of a bound A so that

$$\|Tf\|_{L^2} \le A\|f\|_{L^2},$$
$$\int_{|x|\ge 2|y|} |K(x-y) - K(x)|\,dx \le A, \quad \text{all } y \ne 0. \tag{27}$$

Recall that (in Chapter 1) we saw that these operators have a natural extension to $L^p(\mathbf{R}^n)$, $1 \le p < \infty$; for $1 < p < \infty$ they are bounded, and for $p = 1$ a weak-type result holds. Also in Chapter 3 we saw that this extended T is also a bounded mapping from $H^1(\mathbf{R}^n)$ to $L^1(\mathbf{R}^n)$. To define the natural extension of T to L^∞, we consider the dual operator $\tilde{T}$, defined analogously to T, but with m replaced by $\tilde{m}$, where $\tilde{m}(\xi) = \bar{m}(-\xi)$. Then the associated distribution $\tilde{K}$ is also a locally integrable function away from the origin, and is given by $\tilde{K}(x) = \bar{K}(-x)$. In fact $\tilde{T}$ is, except for a complex conjugate, the Hilbert space $(L^2(\mathbf{R}^n))$ adjoint of T, for which

$$\int_{\mathbf{R}^n} (Tf)g\,dx = \int_{\mathbf{R}^n} f(\tilde{T}g)\,dx \tag{28}$$

whenever $f, g \in L^2(\mathbf{R}^n)$. Notice that $\tilde{T}$ and $\tilde{K}$ satisfy the same kind of inequalities (27) as do T and K.

The above identity allows us to extend the domain of T and to define Tf for any bounded f; more precisely, Tf is defined, by (28), up to an additive constant and is square-integrable on every compact set. In fact, whenever $g \in L^2_{B,0}$ (in the notation of §1.2.2), we know that $g \in H^1$ with $\|g\|_{H^1} \le c|B|^{1/2}\|g\|_{L^2}$. In particular, we get from the right side of (28) a continuous linear functional on $L^2_{B,0}$, with norm $\le cA\|f\|_{L^\infty}|B|^{1/2}$, since $\tilde{T}$ is bounded from H^1 to L^1 with norm $\le cA$. Thus, on each ball B, Tf is thereby defined as an element of $L^2_{B,0}$, and by the argument in §1.2.2 it can be globally defined, modulo an additive constant, on all of $\mathbf{R}^n$. The same argument of course shows that $\|Tf - (Tf)_B\|_{L^2(B)} \le cA\|f\|_{L^\infty}|B|^{1/2}$, and so

$$\|Tf\|_{\mathrm{BMO}} \le A'\|f\|_{L^\infty}. \tag{29}$$

We summarize our discussion up to this point.

PROPOSITION 1. *Suppose T is extended to L^∞ as described above. Then T maps L^∞ to BMO; and the bound A' in* (29) *depends only on the bound A in* (27).

One can also obtain essentially the same conclusion directly without recourse to duality. We give the simple argument involved because it also leads to an interesting variant related to the sharp function.

Let f be any bounded function with compact support (it is to be noted that the size of the support of f will not enter into any of the estimates that follow). Since by our assumptions f is also in L^2, we need to deal only with the originally defined Tf. For any ball B, let B^* denote its double, and write $f = f_1 + f_2$, with $f_1 = f\chi_{B^*}$, $f_2 = f\chi_{^cB^*}$. Then $Tf = Tf_1 + Tf_2$. For Tf_1 we use the L^2 estimate in (27); so

$$\int_B |Tf_1|\,dx \le |B|^{1/2} \cdot \|Tf_1\|_{L^2} \le A|B|^{1/2} \cdot \|f_1\|_{L^2} \le A2^{n/2}|B| \cdot \|f\|_{L^\infty}.$$

For Tf_2 we use the integral inequality in (27), and define the constant c_B by $c_B = \int_{^cB^*} K(x_0 - y)\,f(y)\,dy$, where x_0 is the center of B. It follows that for $x \in B$

$$Tf_2(x) - c_B = \int_{^cB^*} [K(x-y) - K(x_0 - y)]f(y)\,dy,$$

and therefore $|Tf_2 - c_B| \le A\|f\|_{L^\infty}$. When this is combined with the estimate for Tf_1, we obtain

$$\frac{1}{|B|}\int_B |Tf - c_B|\,dx \le (2^{n/2} + 1)A\|f\|_{L^\infty},$$

which gives us (29), at least for functions of compact support.

Remark. The above argument can be modified to show that the operator T, when properly defined, actually maps BMO to itself (see also §6.3^a).

4.2 The sharp function. The statement below requires, in addition to the sharp function, a more stringent version of the standard maximal function. Let $r \ge 1$ and define $M_r(f) = [M(|f|^r)]^{1/r}$. Then, as Hölder's inequality easily shows, $Mf \le M_r f$; but $M_r f$ can be finite only if f is locally in L^r.

We shall also need to strengthen the assumptions (27) made for the operator T. Instead of the integral inequality occurring in (27), we make the more restrictive hypothesis that for some $\gamma > 0$

$$|K(x-y) - K(x)| \le A\frac{|y|^\gamma}{|x|^{n+\gamma}}, \quad \text{whenever } |x| \ge 2|y|. \tag{30}$$

PROPOSITION 2. *Suppose T satisfies the assumptions above* (*in particular,* (30) *holds*)*. Let $r > 1$, then*

$$(Tf)^\sharp(x) \le A_r(M_r f)(x) \tag{31}$$

whenever $f \in L^p$, with $r \le p \le \infty$.

What is striking about this theorem is that it gives a *pointwise* estimate of (suitable averages of) Tf in terms of f, a result that in essence is stronger that the usual L^p inequalities. In this regard it can be viewed as a culmination of the techniques by which singular integrals are analyzed in terms of maximal functions.

The proof is a simple variant of the argument following Proposition 1 above. Let B be any ball containing the point $\bar{x}$, and let x_0 be the center of B. It suffices to find a constant c_B so that

$$\frac{1}{|B|}\int_B |Tf(x) - c_B|\,dx \le A_r M_r f(\bar{x}), \tag{32}$$

with A_r depending only on A and r.

With B^* denoting the double of the ball B, we write as before $f = f_1 + f_2 = f\chi_{B^*} + f\chi_{{}^cB^*}$. We know by the theory in Chapter 1 that T is bounded on L^r, if $r > 1$; therefore

$$\int_B |Tf_1|\,dx \le |B|^{1-1/r}\left(\int_B |Tf_1|^r\,dx\right)^{1/r} \le |B|^{1-1/r}A_r\|f_1\|_{L^r}$$
$$\le |B|A_r\left(\frac{1}{|B|}\int_{B^*}|f|^r\,dx\right)^{1/r} \le c|B|(M_rf)(\bar{x}).$$

Next observe that if $x \in B$, $y \notin B^*$, then by our assumption (30)

$$|K(x-y) - K(x_0-y)| \le A\frac{t^\gamma}{|\bar{x}-y|^{n+\gamma}},$$

where t is the radius of B. So if we set

$$c_B = \int_{{}^cB^*} K(x_0-y)\,f(y)\,dy,$$

then for $x \in B$,

$$|Tf_2(x) - c_B| \le \int \Phi_t(\bar{x}-y)|f(y)|\,dy,$$

with $\Phi_t(u) = t^{-n}\Phi(u/t)$, and $\Phi(u) = Au^{-n-\gamma}\chi_{|u|\ge 1}$. The last integral can be majorized by the usual maximal function, in view of the remarks in Chapter 2, §2.1.

So we have

$$|Tf_2(x) - c_B| \le AMf(\bar{x}) \le AM_rf(\bar{x}).$$

Combining the estimates for Tf_1 and Tf_2 gives (32) and proves the proposition.

4.3 Relation with Carleson measures and square functions. Recall that, for each fixed $\Phi \in \mathcal{S}$ whose integral vanishes, we have defined (in Chapter 1, §6.3) a closely related pair of "square function" operators $f \mapsto s_\Phi(f)$ and $f \mapsto S_\Phi(f)$, the former given by

$$(s_\Phi f)(x) = \left(\int_0^\infty |f * \Phi_t(x)|^2\frac{dt}{t}\right)^{1/2}, \tag{33}$$

the latter having a similar "nontangential" definition.

The basic property of s_Φ (and S_Φ) is the L^2 boundedness

$$\|s_\Phi f\|_{L^2} \le A\|f\|_{L^2}. \tag{34}$$

An important subclass of these operators are those that arise when Φ is "nondegenerate" in the following sense: There exists a $\Psi \in \mathcal{S}$, also with $\int \Psi\,dx = 0$, so that

$$\int_0^\infty \Phi_t * \Psi_t\frac{dt}{t} = \delta. \tag{35}$$

Here δ is the Dirac delta function, the left side of (35) is defined as

$$\lim_{\substack{\varepsilon\to 0\\ N\to\infty}} \int_\varepsilon^N \Phi_t * \Psi_t\frac{dt}{t},$$

and the limit is taken in the sense of distributions.

The above notion of nondegeneracy is not as mysterious as it seems. The condition is equivalent with the validity of the converse to the L^2 inequality (34); or equivalently that the Fourier transform $\widehat{\Phi}$ does not vanish identically on any ray emanating from the origin (see §6.19 below). In particular, Φ is nondegenerate whenever Φ is radial and does not vanish identically.

We know that the operators s_Φ and S_Φ are bounded from L^p to itself for $1 < p < \infty$.† If Φ is nondegenerate, there are also converse inequalities in these cases (see Chapter 1, §8.23). The corresponding result for BMO is not that $s_\Phi f \in L^\infty$ (or $S_\Phi f \in L^\infty$), but a weaker condition that can be stated as follows.

THEOREM 3. *Let $\Phi \in \mathcal{S}$ with $\int \Phi = 0$.*

(a) *Suppose $f \in$ BMO, and let*

$$d\mu = |f * \Phi_t(x)|^2\frac{dx\,dt}{t}.$$

Then $d\mu$ is a Carleson measure, i.e., in the notation of Chapter 2, §2.2,

$$\sup_B \frac{1}{|B|}\int_{T(B)} |f * \Phi_t(x)|^2\frac{dx\,dt}{t} = \|d\mu\|_C < \infty, \tag{36}$$

with $\|d\mu\|_C \le c\|f\|^2_{\mathrm{BMO}}$; the supremum is over all balls $B \subset \mathbf{R}^n$.

(b) *Conversely, suppose Φ is nondegenerate as in (35),*

$$\int_{\mathbf{R}^n} \frac{|f(x)|\,dx}{1+|x|^{n+1}} < \infty,$$

and $d\mu = |f * \Phi_t(x)|^2\,dx\,dt/t$ satisfies (36); then f is in BMO and $\|f\|^2_{\mathrm{BMO}} \le c\|d\mu\|_C$.

† These operators map H^1 to L^1 as well; this is nothing more than a vector-valued version of Theorem 3 in Chapter 3.

4.3.1 The proof of part (a) of the theorem is easy, and follows the approach used for propositions 1 and 2 above. For any ball B, write $f = f_1 + f_2 + f_3$, where $f_1 = (f - f_{B^*})\chi_{B^*}$, $f_2 = (f - f_{B^*})\chi_{{}^cB^*}$, and $f_3 = f_{B^*}$. Here B^* denotes the double of the ball B and f_{B^*} is the mean value of f in B^*. Then $f * \Phi_t = f_1 * \Phi_t + f_2 * \Phi_t$, since $\int \Phi\, dx = 0$.

Set $d\mu_1 = |f_1 * \Phi_t(x)|^2\, dx\, dt/t$, $d\mu_2 = |f_2 * \Phi_t(x)|^2\, dx\, dt/t$. Thus $d\mu \le 2(d\mu_1 + d\mu_2)$. Treating $d\mu_1$ first, we note that

$$\begin{aligned}\int_B (s_\Phi f_1(x))^2\, dx &\le \|s_\Phi f_1\|_{L^2}^2 \\ &\le A\|f_1\|_{L^2}^2 = A\int_{B^*} |f - f_{B^*}|^2\, dx \le A|B| \cdot \|f\|_{\mathrm{BMO}}^2;\end{aligned}$$

the second inequality is by the L^2 boundedness (34), and the last inequality is by the property (7) in §1.3 for $f \in \mathrm{BMO}$. However

$$\begin{aligned}\int_B (s_\Phi f_1(x))^2\, dx &= \int_{x\in B}\int_0^\infty |(f_1 * \Phi_t)(x)|^2 \frac{dt}{t}\, dx \\ &\ge \int_{T(B)} |(f_1 * \Phi_t)(x)|^2 \frac{dx\, dt}{t}.\end{aligned}$$

So $\int_{T(B)} d\mu_1 \le A|B| \cdot \|f\|_{\mathrm{BMO}}^2$.

Next, as is easily seen, if $(x,t) \in T(B)$

$$|f_2 * \Phi_t(x)| \le c \int_{y\in {}^cB^*} \frac{|f(y) - f_{B^*}| \cdot t}{(t + |x_0 - y|)^{n+1}}\, dy,$$

and when $y \notin B^*$, we have $t/(t + |x_0 - y|)^{n+1} \le ct/(r + |x_0 - y|)^{n+1}$; here x_0 is the center of B^* and r is its radius.

Finally, the elementary inequality (2) in §1.1.4 can be adapted to our needs above, because the BMO space and its norm are translation and dilation invariant. Thus we easily get the following variant of (2):

$$\int_{\mathbf{R}^n} |f(y) - f_{B^*}| \frac{r\, dy}{(r + |x_0 - y|)^{n+1}} \le c\|f\|_{\mathrm{BMO}}.$$

Inserting this in the above gives us that $|f_2 * \Phi_t(x)| \le c(t/r)\|f\|_{\mathrm{BMO}}$, whenever $(x,t) \in T(B)$. Therefore obviously

$$\int_{T(B)} d\mu_2 = \int_{T(B)} |f_2 * \Phi_t(x)|^2 \frac{dx\, dt}{t} \le c|B| \cdot \|f\|_{\mathrm{BMO}}^2,$$

and the proof of conclusion (a) is complete.

4.3.2 At this stage we wish to point out a discrete version of part (a) of the theorem, which will be useful in later applications. With the same assumptions on f and Φ as above, the measure

$$d\mu = \sum_{j=-\infty}^{\infty} |f * \Phi_{2^j}(x)|^2\, dx\, \delta_{2^j}(t), \tag{37}$$

where $\delta_{2^j}(t)$ is the unit Dirac mass at the point $t = 2^j$, is a Carleson measure, and $\|d\mu\|_C \le c\|f\|_{\mathrm{BMO}}^2$.

The proof of this assertion is essentially the same as that just given for (36), once we replace the continuous square function s_Φ by a discrete analogue s_Φ^D, given by

$$s_\Phi^D(f) = \left(\sum_{j=-\infty}^{\infty} |f * \Phi_{2^j}|^2\right)^{1/2}.$$

The fact that s_Φ^D is bounded on L^2 is, like the case for the continuous analogue, a direct consequence of Plancherel's theorem. In this instance, it is equivalent with the majorization

$$\sup_\xi \sum_j |\widehat{\Phi}(2^j\xi)|^2 \le A^2,$$

which itself is an immediate consequence of the fact that $|\widehat{\Phi}(u)| \le A|u|$ and $|\widehat{\Phi}(u)| \le A|u|^{-1}$.

4.3.3 It may also be worthwhile to record here the following observation that is implicit in the proof given in §4.3.1: whenever $f \in \mathrm{BMO}$ and Φ is a fixed test function with $\int \Phi\, dx = 0$, then

$$\sup_{t>0} \|f * \Phi_t\|_{L^\infty} \le A_\Phi \|f\|_{\mathrm{BMO}}$$

4.4 Another duality inequality. The proof of the converse ((b) of Theorem 3) is deeper and requires that we develop some ideas related to the "tent spaces" studied in §2.2 of Chapter 2.

For any measurable function F given on the half-space $\mathbf{R}_+^{n+1}$, we define its *square function*, $\mathfrak{S}(F)$, to be the function on $\mathbf{R}^n$ given by

$$\mathfrak{S}(F)(x) = \left(\int_{\Gamma(x)} |F(y,t)|^2 \frac{dy\, dt}{t^{n+1}}\right)^{1/2}, \quad x \in \mathbf{R}^n;$$

as before, $\Gamma(x) = \{(y,t) : |y - x| < t\}$ is the cone of aperture 1 at x. In analogy with the space $\mathcal{N}$ appearing in Chapter 2, we denote by $\mathcal{N}_2$ the space of all functions F on $\mathbf{R}^{n+1}_+$ so that $\mathfrak{S}(F) \in L^1(\mathbf{R}^n)$; with the norm $\|F\|_{\mathcal{N}_2} = \|\mathfrak{S}(F)\|_{L^1}$, the vector space $\mathcal{N}_2$ becomes a Banach space.

Together with $\mathfrak{S}$, we define a dual operator $\mathfrak{T}$, related to Carleson measures, given by

$$\mathfrak{T}(F)(x) = \sup_{x\in B}\left(\frac{1}{|B|}\int_{T(B)} |F(y,t)|^2 \frac{dy\,dt}{t}\right)^{1/2}, \quad x \in \mathbf{R}^n.$$

Here $T(B)$ is the tent over B, as defined in §2.2 of Chapter 2. The duality inequality that follows is the analogue of Theorem 2 in Chapter 2.

PROPOSITION. (a) *Whenever $G \in \mathcal{N}_2$ and $\mathfrak{T}(F) \in L^\infty(\mathbf{R}^n)$,*

$$\int_{\mathbf{R}^{n+1}_+} |F(x,t)G(x,t)|\frac{dx\,dt}{t} \le c\|\mathfrak{T}(F)\|_{L^\infty}\|G\|_{\mathcal{N}_2}.$$

(b) *More precisely,*

$$\int_{\mathbf{R}^{n+1}_+} |F(x,t)G(x,t)|\frac{dx\,dt}{t} \le c\int_{\mathbf{R}^n} \mathfrak{T}(F)(x)\,\mathfrak{S}(G)(x)\,dx. \tag{38}$$

Proof. We may of course assume that both F and G are positive. For any $\tau > 0$, let $\Gamma^\tau(x) = \{(y,t) : |y-x| < t, t < \tau\}$ be the cone at x truncated at height τ; we also set

$$\mathfrak{S}(F|\tau)(x) = \left(\int_{\Gamma^\tau(x)} |F(y,t)|^2 \frac{dy\,dt}{t^{n+1}}\right)^{1/2}.$$

Note that $\mathfrak{S}(F|\tau)$ is increasing with τ, and that $\mathfrak{S}(F|\infty) = \mathfrak{S}(F)$. For every F, we define the "stopping-time" $\tau(x)$ by

$$\tau(x) = \sup\{\tau > 0 : \mathfrak{S}(F|\tau)(x) \le A\mathfrak{T}(F)(x)\}.$$

Here A is a large constant to be determined later; we shall fix A to depend only on the dimension n. The key observation is that there exists a $c = c_A$ so that whenever B is a ball in $\mathbf{R}^n$ of radius r, then

$$|\{x \in B : \tau(x) \ge r\}| \ge cr^n. \tag{39}$$

Let us temporarily assume (39), and show that it implies our desired conclusion (38). For any nonnegative function H, we have

$$\int_{\mathbf{R}^{n+1}_+} H(y,t)\,t^n\,dy\,dt \le c^{-1}\int_{\mathbf{R}^n}\left\{\int_{\Gamma^{\tau(x)}(x)} H(y,t)\,dy\,dt\right\}dx,$$

which follows directly from (39) by Fubini's theorem.‡ In the above we take $H(y,t) = F(y,t)G(y,t)t^{-n-1}$ and use Schwarz's inequality. The result is

$$\begin{aligned}\int_{\mathbf{R}^{n+1}_+} F(x,t)\,G(x,t)\frac{dx\,dt}{t} &\le c^{-1}\int_{\mathbf{R}^n} \mathfrak{S}(F|\tau(x))(x)\,\mathfrak{S}(G|\tau(x))\,dx \\ &\le c^{-1}A\int_{\mathbf{R}^n} \mathfrak{T}(F)(x)\,\mathfrak{S}(G)(x)\,dx,\end{aligned}$$

by the definition of the stopping time. This gives (38), and we are left with the task of establishing (39), to wit, that on every ball, for x ranging over a positive fraction of its volume, the stopping time is at least as large as the radius of the ball.

Let B be any ball of radius r and $\tilde{B}$ the concentric ball with radius $3r$; note that $\cup\Gamma^r(x)$ is contained in $T(\tilde{B})$, the tent over $\tilde{B}$. Thus

$$\int_B \mathfrak{S}(F|r)^2(x)\,dx \le a_n\int_{T(\tilde{B})} |F(y,t)|^2\frac{dy\,dt}{t}$$

by Fubini's theorem (also compare with (60) in the previous chapter); here a_n denotes the volume of the unit ball in $\mathbf{R}^n$.

Therefore,

$$\frac{1}{|B|}\int_B \mathfrak{S}(F|r)^2(x)\,dx \le a'\frac{1}{|\tilde{B}|}\int_{T(\tilde{B})} |F(y,t)|^2\frac{dy\,dt}{t} \le a'\inf_{x\in B}\mathfrak{T}(F)(x),$$

with $a' = 3^n a_n$. From this it is clear that (39) holds, as long as the constant A^2 exceeds a'; in fact, we get (39) with $c = c_A = a_n(1 - a'/A^2)$. The proposition is therefore proved.

Remarks. The proposition can be used to identify the dual of $\mathcal{N}_2$ with those F for which $\mathfrak{T}(F) \in L^\infty(\mathbf{R}^n)$. Note also that one has the implication $\mathfrak{S}(F) \in L^\infty(\mathbf{R}^n) \Rightarrow \mathfrak{T}(F) \in L^\infty(\mathbf{R}^n)$, but the space of F with $\mathfrak{S}(F) \in L^\infty$ is strictly smaller. See §6.10.

4.4.1 We return to the proof of the converse direction of Theorem 3. We are given a Φ that satisfies the nondegeneracy condition (35), and an f for which

$$\int_{\mathbf{R}^n} |f(x)|(1+|x|)^{-n-1}\,dx < \infty,$$

and where $d\mu = |(f * \Phi_t)(x)|^2 dx\,dx/t$ is a Carleson measure.

We fix a function $g \in H^1_a$, i.e., a bounded function of compact support whose integral vanishes. We require the identity

$$\int_{\mathbf{R}^n} f(x)g(-x)\,dx = \int_{\mathbf{R}^{n+1}_+} F(x,t)G(x,t)\frac{dx\,dt}{t}, \tag{40}$$

‡ Compare also with the variant appearing in (61), §4.4.3 of the previous chapter.

where $F(x,t) = (f * \Phi_t)(x)$ and $G(x,t) = (g * \Psi_t)(x)$.

On the formal level, (40) is an immediate consequence of the fact that $\int_0^\infty \Phi_t * \Psi_t \, dt/t = \delta$. Its actual proof requires a little care; we will return to it momentarily. Assuming (40), we see that the proposition then yields

$$\left| \int f(x) g(-x)\, dx \right| \le c \|\mathfrak{T}(F)\|_{L^\infty} \|\mathfrak{S}(G)\|_{L^1} \le c \|d\mu\|_*^{1/2} \|\mathfrak{S}(G)\|_{L^1}.$$

However, $\mathfrak{S}(G) = S_\Psi(g)$, because

$$(S_\Psi g)(x) = \left(\int_{\Gamma(x)} |g * \Psi_t(y)|^2 \frac{dy\, dt}{t^{n+1}} \right)^{1/2}.$$

Next, the Hilbert-space valued version of the singular integral theorem in Chapter 3, §3.1 shows that $g \mapsto S_\Psi(g)$ is a bounded mapping from H^1 to L^1. Thus $\|\mathfrak{S}(G)\|_{L^1} \le c\|g\|_{H^1}$.

As a result

$$\left| \int_{\mathbf{R}^n} f(x)\, g(-x)\, dx \right| \le c \|d\mu\|_*^{1/2} \|g\|_{H^1}$$

for every g in the dense subspace H^1_a of H^1. This gives a linear functional on H^1, and by part (b) of Theorem 1, f agrees with an element of BMO, and $\|f\|^2_{\mathrm{BMO}} \le c^2 \|d\mu\|_*$.

4.4.2 Finally, there is the matter of establishing the identity (40).

We claim first that there exists an $H \in \mathcal{S}$, with $\int H\, dx = 1$, so that

$$\int_\varepsilon^N \Phi_t * \Psi_t \frac{dt}{t} = H_\varepsilon - H_N, \tag{41}$$

whenever $0 < \varepsilon < N < \infty$; here $H_\varepsilon(x) = \varepsilon^{-n} H(x/\varepsilon)$, with a similar definition for H_N. To see this, let us set

$$h(\xi) = \int_1^\infty \widehat{\Phi}(t\xi) \widehat{\Psi}(t\xi) \frac{dt}{t}, \qquad \text{for } \xi \ne 0. \tag{42}$$

The rapid decrease at infinity of $\widehat{\Phi}$, $\widehat{\Psi}$ and their derivatives insures that h is smooth when $\xi \ne 0$, and that h, together with its derivatives, decays rapidly at infinity.

On the other hand, for $\xi \ne 0$, the integral $\int_0^\infty \widehat{\Phi}(t\xi)\widehat{\Psi}(t\xi) \frac{dt}{t}$ converges absolutely (since $\widehat{\Phi}$ and $\widehat{\Psi}$ vanish at the origin) and its value is 1; this last fact is of course equivalent with the property (35). Therefore, for $\xi \ne 0$,

$$h(\xi) = 1 - \int_0^1 \widehat{\Phi}(t\xi)\widehat{\Psi}(t\xi) \frac{dt}{t}.$$

The smoothness of $\widehat{\Phi}$ and $\widehat{\Psi}$, together with their vanishing at the origin, then shows that $h(\xi)$ is smooth near the origin, and hence that $h \in \mathcal{S}$.

If we make a change of scale in (42), we see that

$$h(\varepsilon\xi) = \int_\varepsilon^\infty \widehat{\Phi}(t\xi)\widehat{\Psi}(t\xi) \frac{dt}{t}.$$

therefore $h(0) = \lim_{\varepsilon \to 0} h(\varepsilon\xi) = 1$; also

$$h(\varepsilon\xi) - h(N\xi) = \int_\varepsilon^N \widehat{\Phi}(t\xi)\widehat{\Psi}(t\xi) \frac{dt}{t}.$$

This establishes (41), when we define H as the inverse Fourier transform of h, and observe that $\int_{\mathbf{R}^n} H\, dx = h(0) = 1$.

From (41), we see that

$$\begin{aligned} \int_{\mathbf{R}^n} f(x)\,(g * H_\varepsilon)(-x)\, dx &- \int_{\mathbf{R}^n} f(x)\,(g * H_N)(-x)\, dx \\ &= \int_{\mathbf{R}^n} \int_\varepsilon^N F(x,t)\, G(x,t) \frac{dx\, dt}{t}. \end{aligned} \tag{43}$$

We now let $\varepsilon \to 0$, $N \to \infty$. The right side of (43) converges to the right side of (40), since we have seen that the integral in question converges absolutely. Also, because $g \in H^1_a$, the argument in Chapter 3, §2.2 shows that

$$\sup_t |(g * H_t)(x)| \le \frac{c}{(1+|x|)^{n+1}}.$$

Moreover, $g * H_\varepsilon(-x) \to g(-x)$ a.e. as $\varepsilon \to 0$, while $g * H_N(-x) \le cN^{-n} \to 0$ as $N \to \infty$. Since $\int |f(x)|(1+|x|)^{-n-1}\, dx < \infty$, the dominated convergence theorem gives $\int f(x) g(-x)\, dx$ as the limit of the left side of (43), when $\varepsilon \to 0$ and $N \to \infty$. The identity (40) is therefore established, and the proof of Theorem 3 is complete.

4.4.3 The assumptions that Φ (and Ψ) are in $\mathcal{S}$ can be substantially relaxed without changing the proof. An interesting example arises when

$$\Phi(x) = \left. \frac{\partial P_t(x)}{\partial t} \right|_{t=1}$$

and $P_t(x) = t^{-n} P(x/t)$ is the Poisson kernel $P(x) = c_n(1+|x|^2)^{-(n+1)/2}$. Then if $\Psi = 4\Phi$, one has the identity (35) for the representation of the Dirac delta function. A consequence of this is that, writing $u(x,t) = (f * P_t)(x)$ for the Poisson integral of f, the necessary and sufficient condition that f is in BMO becomes the condition that $t|\nabla u(x,t)|^2\, dx\, dt$ is a Carleson measure.

Incidentally, these remarks allow us to give another proof of the duality between H^1 and BMO. The key here is Theorem 3 (and the proposition on which it depends), together with the H^1 inequality given by the case $p = 1$ of Proposition 4 in §4 of the previous chapter. This approach does not depend on the atomic decomposition of H^1. See §6.12.

4.5 Quasi-orthogonal expansions for functions in BMO. The study of square functions and their variants involve, by their very nature, properties closely related to orthogonality. Because of this it is natural to expect that the previous characterization of functions in BMO will have an equivalent formulation in terms of orthogonal decompositions. It is this we shall now present, albeit in a rather simple and primitive form. Various refinements and extensions of the same ideas are possible, giving also characterizations of many other function spaces besides BMO, leading in addition to what are now known as "wavelet" decompositions.

Our orthogonal decompositions (more precisely, "quasi-orthogonal" decompositions) will be given in terms of a family of "bump" functions; each such function will be associated to a dyadic cube. We fix our notation as follows: the letter Q will be reserved for a dyadic cube, and $B = B_Q$ will be the ball with the same center and twice the diameter (thus $B_Q \supset Q$); similarly the ball B_j will be associated to Q_j, etc.

For each dyadic cube Q, we will be given a function ϕ_Q, supported in B_Q, that satisfies certain natural size, regularity, and moment conditions. We shall assume that

$$\begin{cases} |D^\alpha \phi_Q| \le \dfrac{\ell(Q)^{-|\alpha|}}{|Q|^{1/2}}, & 0 \le |\alpha| \le n, \text{ and} \\ \int x^\alpha \phi_Q(x)\, dx = 0, & 0 \le |\alpha| \le n. \end{cases} \tag{44}$$

with $\ell(Q)$ denoting the length of a side of the cube $Q \subset \mathbf{R}^n$.†

We shall be dealing with functions f that can be represented in the form

$$f = \sum_Q a_Q \phi_Q, \tag{45}$$

where $\{a_Q\}$ is a suitable collection of constants, and the summation in (45) is carried over all dyadic cubes.

PROPOSITION. (a) *Suppose the coefficients $\{a_Q\}$ satisfy the inequalities*

$$\sum_{Q \subset Q_0} |a_Q|^2 \le A|Q_0| \tag{46}$$

for all dyadic cubes Q_0, where the summation in (46) is taken over all dyadic subcubes of Q_0. Then the series (45) gives an $f \in$ BMO in the sense that

$$\lim_{\substack{\rho_1 \to 0 \\ \rho_2 \to \infty}} \sum_{\rho_1 \le \ell(Q) \le \rho_2} a_Q \phi_Q = f$$

exists in the weak topology of BMO *(considered as the dual space of H^1).*

† While somewhat less regularity and fewer moment conditions could be required, this possible generalization is not pursued here.

(b) *Conversely, suppose $f \in$ BMO. Then there is a collection of functions $\{\phi_Q\}$ and a collection of coefficients $\{a_Q\}$ that satisfy* (44) *and* (46) *respectively, so that f is representable as the sum* (45), *in the sense asserted in part* (a).

The smallest A for which (46) *holds is comparable with $\|f\|^2_{\text{BMO}}$.*

4.5.1 The (quasi) orthogonal nature of the expansions (45) is subsumed in the following lemma:

LEMMA. *Suppose $f = \sum a_Q \phi_Q$ is a finite sum. Then we have*

$$\int |f|^2\, dx \le c \sum |a_Q|^2. \tag{47}$$

Proof. We begin by calculating $\int \phi_{Q_1} \bar{\phi}_{Q_2}\, dx$. Writing $B_j = B_{Q_j}$, we may assume that $B_1 \cap B_2 \ne \emptyset$ and $\ell(Q_1) \ge \ell(Q_2)$. Suppose p is the $(n-1)$-degree Taylor polynomial of ϕ_{Q_1}, taken about $\bar{x}$, the center of Q_2. Then $\int \phi_{Q_1}\bar{\phi}_{Q_2}\, dx = \int (\phi_{Q_1} - p)\bar{\phi}_{Q_2}\, dx$, because of the vanishing moments of ϕ_{Q_2}. The smoothness and size conditions therefore imply

$$|\phi_{Q_1}(x) - p(x)| \le c \left(\frac{\ell(Q_2)}{\ell(Q_1)}\right)^n |Q_1|^{-1/2}, \qquad \text{for } x \in B_2,$$

while of course $|\phi_{Q_2}| \le |Q_2|^{-1/2}$. Since the support of ϕ_{Q_2} has measure $|B_2| \le c|Q_2|$, we get

$$\left| \int \phi_{Q_1} \bar{\phi}_{Q_2}\, dx \right| \le c \left(\frac{\ell(Q_2)}{\ell(Q_1)}\right)^n (|Q_2|/|Q_1|)^{1/2};$$

i.e.,

$$\begin{cases} \left| \int \phi_{Q_1} \bar{\phi}_{Q_2}\, dx \right| \le c\gamma(Q_1, Q_2) \\ \text{where} \quad \gamma(Q_1, Q_2) = 2^{-(3/2)nj}, \end{cases} \tag{48}$$

if $\ell(Q_2) = 2^{-j}\ell(Q_1)$, $j = 0, 1, \ldots$. Also, $\gamma(Q_1, Q_2) = 0$ if $B_1 \cap B_2 = \emptyset$. The inequality (48) expresses the quasi-orthogonality of the $\{\phi_Q\}$. The only real significance of the exponent $3/2$ in the above is that it is strictly greater than 1.

Now

$$\begin{aligned} \int |f|^2\, dx &= \sum_{Q_1, Q_2} a_{Q_1} \bar{a}_{Q_2} \int \phi_{Q_1} \bar{\phi}_{Q_2}\, dx \\ &\le 2 \sum_{\ell(Q_2) \le \ell(Q_1)} |a_{Q_1}|\, |a_{Q_2}| \int \phi_{Q_1} \bar{\phi}_{Q_2}\, dx \\ &\le c \sum_{\ell(Q_2) \le \ell(Q_1)} (|a_{Q_1}|^2 + |a_{Q_2}|^2) \gamma(Q_1, Q_2). \end{aligned}$$

Next, with Q_1 fixed,

$$\sum_{\ell(Q_2)\le\ell(Q_1)} \gamma(Q_1,Q_2) \le c\sum_{j=0}^{\infty} 2^{nj}2^{-(3/2)nj} \le c,$$

because for each Q_1, there are at most $c2^{nj}$ cubes Q_2 with $\ell(Q_2) = 2^{-j}\ell(Q_1)$, and with $B_2 \cap B_1 \ne \emptyset$. Similarly, because for each fixed Q_2, there are at most c cubes Q_1 with $\ell(Q_2) = 2^{-j}\ell(Q_1)$, so that $B_2 \cap B_1 \ne \emptyset$, we have that $\sum_{Q_1} \gamma(Q_1,Q_2) \le c$. Inserting these estimates in the sum

$$\sum_{\ell(Q_2)\le\ell(Q_1)} (|a_{Q_1}|^2 + |a_{Q_2}|^2)\gamma(Q_1,Q_2)$$

gives the desired estimate (47), and the lemma is proved.

4.5.2 Using the lemma, we now prove part (a) of the proposition. We assume that only finitely many of the a_Q are nonzero, but we make estimates that are independent of this number.

Fix a ball $\bar{B}$ of radius r. Then for $x \in \bar{B}$, write $\sum a_Q\phi_Q = f_1(x) + f_2(x)$, where f_1 is the sum taken over those cubes Q so that $\ell(Q) \le r$ and $B_Q \cap \bar{B} \ne \emptyset$. Similarly, f_2 is the sum taken over cubes for which $\ell(Q) > r$, and again $B_Q \cap \bar{B} \ne \emptyset$.

It is important to observe that given $\bar{B}$, there are a bounded number of dyadic cubes Q_0, so that $\bar{B}$ is contained in the union of these cubes, while each $\ell(Q_0) \le r$. Thus, by (47) of the lemma and the assumption (46),

$$\int_{\bar{B}} |f_1(x)|^2\,dx \le Ac|\bar{B}|. \tag{49}$$

Next, let $c_{\bar{B}} = \sum_{\ell(Q)>r} a_Q\phi_Q(\bar{x})$, where $\bar{x}$ is the center of the ball $\bar{B}$.

Then for $x \in \bar{B}$,

$$|f_2(x) - c_{\bar{B}}| \le \sum_{\ell(Q)>r} |a_Q|\cdot|\phi_Q(x) - \phi_Q(\bar{x})|.$$

To estimate the right side of the above inequality we use the following simple observations: first, $|a_Q| \le A|Q|^{1/2}$, because of (46); second, $|\phi_Q(x)-\phi_Q(\bar{x})| \le cr|Q|^{-1/2}\ell(Q)^{-1}$, because of (44); and third, for each x and each length, there are at most a bounded number of cubes Q whose sides have that length for which $x \in B_Q$. Thus the sum is majorized by $Acr\sum_{2^j\ge r} 2^{-j} \le Ac'$. Therefore $|f_2(x) - c_{\bar{B}}| \le Ac'$, for $x \in \bar{B}$. It follows that $\int_{\bar{B}} |f(x) - c_{\bar{B}}|^2\,dx \le Ac|\bar{B}|$, for each ball $\bar{B}$, and so $f \in \mathrm{BMO}$, with $\|f\|^2_{\mathrm{BMO}} \le Ac$.

We now drop the assumption that there are only finitely many a_Q different from zero and write

$$S_{\rho_1}^{\rho_2} = \sum_{\rho_1\le\ell(Q)\le\rho_2} a_Q\phi_Q = S_{\rho_1}^1 + S_1^{\rho_2}$$

if $\rho_1 \le 1 \le \rho_2$. What we have just proved shows that $S_{\rho_1}^1$ and $S_1^{\rho_2}$ are uniformly in BMO, and so to prove the asserted weak convergence, we need only show the existence of the limits

$$\lim_{\rho_1\to 0}\int S_{\rho_1}^1(x)g(x)\,dx \qquad \text{and} \qquad \lim_{\rho_2\to\infty}\int S_1^{\rho_2}(x)g(x)\,dx$$

whenever g is an element of a dense subspace of H^1 (the predual of BMO). We take H_a^1, the bounded functions of compact support with vanishing integral, to be that subspace. Fix $g \in H_a^1$ and let $K = \operatorname{supp} g$.

We first consider $\lim_{\rho_1\to 0}\langle S_{\rho_1}^1, g\rangle$. Since $g \in L^2$, it suffices to show that $S_{\rho_1}^1$ converges in L^2 as $\rho_1 \to 0$. This follows from (46) and (47) once we observe that, since K is a fixed compact set, there is a finite collection of dyadic cubes Q_j so that, if Q is any dyadic cube with $\ell(Q) \le 1$ and $B_Q \cap K \ne \emptyset$, we have $Q \subset \bigcup Q_j$.

It remains to prove that

$$\lim_{\rho_1,\rho_2\to\infty}\int S_1^{\rho_1}(x)\,g(x)\,dx - \int S_1^{\rho_2}(x)\,g(x)\,dx = 0.$$

Now

$$\begin{aligned}\left|\int S_1^{\rho_1}g - \int S_1^{\rho_2}g\right| &\le \sum_{\ell(Q)\ge\min(\rho_1,\rho_2)} |a_Q|\cdot\left|\int \phi_Q(x)\,g(x)\,dx\right| \\ &\le c\sum |a_Q|\cdot\int|\phi_Q(x) - \phi_Q(\bar{x})|\,dx,\end{aligned}$$

where $\bar{x}$ is some point in K. However, by (44),

$$\int_K |\phi_Q(x) - \phi_Q(\bar{x})|\,dx \le c_K|Q|^{-1/2}\ell(Q)^{-1}.$$

Again using the compactness of K, we see that there exists a constant $N = N_K$ so that, for each fixed side length $2^j \ge 1$, there are at most N dyadic cubes Q with $\ell(Q) = 2^j$ and $B_Q \cap K \ne \emptyset$. If we recall that $|a_Q| \le c|Q|^{1/2}$, we see that the sum is dominated by $c_K\sum_{2^j\ge\min(\rho_1,\rho_2)} 2^{-j}$, which tends to zero as $\rho_1,\rho_2 \to \infty$.

4.5.3 To prove the converse, we will proceed as follows. First we shall use that there exists a suitable pair Φ, Ψ, with Ψ having support in the unit ball, so that if $F(x,t) = f * \Phi_t(x)$, then

$$f(x) = \int_0^\infty F(x,t) * \Psi_t \frac{dt}{t} = \int\int_{\mathbf{R}^{n+1}_+} F(y,t)\Psi_t(x-y)\,dy\frac{dt}{t}, \qquad (50)$$

in the sense of weak convergence (i.e., when paired with H^1 functions). We then rewrite this as

$$f(x) = \sum_Q \int_{\ell(Q)/2}^{\ell(Q)} \left\{ \int_Q F(y,t)\Psi_t(x-y)\,dy \right\} \frac{dt}{t} = \sum_Q f_Q(x), \qquad (51)$$

where the sum is taken over *all* dyadic cubes Q. Finally we write $f_Q = a_Q \phi_Q$, with appropriate constants a_Q, and verify (46).

In detail, let $\Phi \in C^\infty$ be a real-valued, radial function, supported in the unit ball, which is not identically zero, but for which‡

$$\int x^\alpha \Phi(x)\,dx = 0, \qquad \text{whenever } 0 \le |\alpha| \le n.$$

If $\widehat{\Phi}$ is its Fourier transform then, for $\xi \neq 0$, the integral

$$\int_0^\infty |\widehat{\Phi}(t\xi)|^2 \frac{dt}{t}$$

converges, since $\widehat{\Phi}$ vanishes at the origin and is rapidly decreasing at infinity. Because $\widehat{\Phi}$ is radial and the measure dt/t is scale-invariant, the value of this integral is a function of ξ that is invariant under rotations and dilations, and hence is a constant c. Now we set $\Psi = c^{-1}\Phi$, and notice that $\widehat{\Phi}(\xi)$ is real, because Φ is radial. The result is

$$\int_0^\infty \widehat{\Phi}(t\xi)\,\widehat{\Psi}(t\xi)\frac{dt}{t} \equiv 1,$$

which means that

$$\lim_{\substack{\varepsilon\to 0\\ N\to\infty}} \int_\varepsilon^N \Phi_t * \Psi_t \frac{dt}{t} = \delta$$

in the sense of distributions.

To see that

$$\int_\varepsilon^N f * \Phi_t * \Psi_t \frac{dt}{t} = \int_\varepsilon^N \int_{\mathbf{R}^n} F(y,t)\Psi_t(x-y)\,dy\frac{dt}{t} = f_\varepsilon^N$$

‡ Such a Φ may be easily obtained by taking a nonzero radial function $\Phi^{(0)}$ supported in the unit ball and setting $\Phi = (\Delta)^n\Phi^{(0)}$, where Δ is the Laplacian.

converges weakly to f as $\varepsilon \to 0$, $N \to \infty$, we note that the argument in §4.4.1 shows that

$$\left| \int f_\varepsilon^N g\,dx \right| \le c\|f\|_{\mathrm{BMO}}\|g\|_{H^1},$$

whenever $g \in H^1$; Theorem 1 then gives $\|f_\varepsilon^N\|_{\mathrm{BMO}} \le c$. Also, by §4.4.2,

$$\int f_\varepsilon^N(x)\,g(-x)\,dx = \int f(x)(H_\varepsilon g)(-x)\,dx - \int f(x)(H_N g)(-x)\,dx,$$

and therefore whenever $g \in H^1_a$, we have that

$$\int f_\varepsilon^N(x)\,g(-x)\,dx \to \int f(x)\,g(-x)\,dx,$$

as $\varepsilon \to 0$ and $N \to \infty$. Since H^1_a is dense in H^1, this establishes the weak convergence of f_ε^N to f, and we have (50) in the sense asserted above. Next, if we choose $\varepsilon = 2^{-j}$, $N = 2^k$, then we have (51) in the sense that

$$f_\varepsilon^N = {\sum_Q}' \int_{\ell(Q)/2}^{\ell(Q)} \left\{ \int_Q F(y,t)\,\Psi_t(x-y)\,dy \right\} \frac{dt}{t} = {\sum}' f_Q, \qquad (52)$$

where the symbol $\sum_Q'$ denotes the summation over all dyadic cubes with side length between 2^{-j-1} and 2^k. In fact, the identity (52) is obvious if we sum first over all the dyadic cubes at a fixed side length, and then over the finitely many lengths.

Now define the coefficients a_Q by

$$\gamma a_Q = \left(\int_{\ell/2}^{\ell} \int_Q |F(y,t)|^2\,dy\frac{dt}{t} \right)^{1/2}; \qquad (53)$$

here $\ell = \ell(Q)$ and γ is a constant to be chosen momentarily. Recall that

$$f_Q = \int_{\ell/2}^{\ell} \int_Q F(y,t)\,\Psi_t(x-y)\frac{dy\,dt}{t}. \qquad (54)$$

From this it follows that

$$|f_Q| \le \left(\int_{\ell/2}^{\ell} \int_Q |F(y,t)|^2 \frac{dy\,dt}{t} \right)^{1/2} \left(\int_{\ell/2}^{\ell} \int_Q |\Psi_t(y,t)|^2 \frac{dy\,dt}{t} \right)^{1/2}.$$

However $|\Psi_t| \le ct^{-n}$, therefore

$$\left(\int_{\ell/2}^{\ell} \int_Q |\Psi_t(y,t)|^2 \frac{dy\,dt}{t} \right)^{1/2} \le c|Q|^{-1/2}$$

and we have that $|f_Q| \le \gamma a_Q c|Q|^{-1/2}$. Similarly,

$$\left|\left(\frac{\partial}{\partial x}\right)^\alpha f_Q\right| \le \gamma a_Q c_\alpha |Q|^{-1/2}\ell(Q)^{-|\alpha|}.$$

Thus if we take γ so small that $\gamma c_\alpha \le 1$ for $0 \le |\alpha| \le n$, and let $\phi_Q = f_Q/a_Q$, we have that $f_Q = a_Q\phi_Q$ and that

$$\left|\left(\frac{\partial}{\partial x}\right)^\alpha \phi_Q\right| \le |Q|^{-1}\ell(Q)^{-|\alpha|}, \qquad 0 \le |\alpha| \le n.$$

Next, the $\{\phi_Q\}$ satisfy the moment conditions in (44), because the $\{f_Q\}$ do, and this in turn is a consequence of the corresponding property for Ψ (and Φ). Also observe from (54) that f_Q (and hence ϕ_Q) is supported in B, where B is the ball with the same center as Q but with twice the diameter. Therefore all the requisite conditions for the $\{\phi_Q\}$ have been verified.

To check the size condition (46) for the $\{a_Q\}$, we note that by (53),

$$\sum_{Q\subset Q_0} |a_Q|^2 \le \gamma^{-2}\int_0^{\ell(Q_0)}\int_{Q_0} |F(y,t)|^2\frac{dy\,dt}{t}$$
$$\le \gamma^{-2}\int_{T(\tilde B)} |F(y,t)|^2\frac{dy\,dt}{t} = \gamma^{-2}\int_{T(\tilde B)} |f * \Phi_t(y)|^2\frac{dy\,dt}{t}.$$

Here $\tilde B$ is the ball with the same center as Q_0 but with three times the diameter; the above inequality is an immediate consequence of the geometric observation that the cylinder $Q_0 \times (0, \ell(Q_0))$ is contained in the tent $T(\tilde B)$. From Theorem 3 (a) (in §4.3), we then see that

$$\sum_{Q\subset Q_0} |a_Q|^2 \le c\|f\|^2_{\mathrm{BMO}}|Q_0|,$$

and our proposition is proved.

4.5.4 Concluding remarks.

1. One can also obtain similar characterizations for the spaces L^p, $1 < p < \infty$, or more generally H^p, $0 < p < \infty$. If the system $\{\phi_Q\}$ is chosen appropriately, then the condition that an f represented by the sum (45) belongs to H^p is that the square function

$$\tilde S(x) = \left(\sum_{x\in Q} |a_Q|^2|Q|^{-1}\right)^{1/2}$$

belongs to L^p. See §6.26.

2. A simplified version of the system $\{\phi_Q\}$ occurs in the dyadic context, and is given by the Haar basis. We describe the situation in one dimension. Suppose h is the function supported in the unit interval $[0,1]$ that equals 1 in the left half and -1 in the right half. For any dyadic interval Q, set

$$h_Q = 2^{j/2}h(2^jx - k), \qquad \text{if } Q = [k2^{-j}, (k+1)2^{-j}].$$

While the $\{h_Q\}$ satisfy only the size condition $|h_Q| \le |Q|^{-1/2}$ and the moment condition $\int h_Q\,dx = 0$ (and not the full conditions (44)), they have the compensating merit of forming a complete orthonormal basis for $L^2(\mathbf{R}^1)$. For $f = \sum a_Q h_Q$, the property

$$\sum_{Q\subset Q_0} |a_Q|^2 \le c|Q_0|$$

is then equivalent with f being in BMO in the dyadic sense, i.e., that $f^\sharp_\Delta \in L^\infty$, which we know is weaker than $f \in$ BMO. See §6.23.

3. Returning to the stronger regularity and moment conditions (44), one can construct such systems $\{\phi_Q\}$ so that they form a complete orthonormal basis of $L^2(\mathbf{R}^n)$. In one dimension this can also be done so that, in addition, the $\{\phi_Q\}$ arise as the dyadic dilates and integer translates of a *single* function ϕ, just as the $\{h_Q\}$ were formed from h above. The systems so obtained, and their generalizations to $\mathbf{R}^n$, are the *wavelet expansions.* See §6.26.

5. An interpolation theorem

An important fact about the theory of BMO is that it can be applied to obtain sharp estimates for large classes of interesting operators. One method that occurs in practice relies on an interpolation theorem that is applicable to an analytic family of operators; this allows one to pass from hypotheses for certain p (e.g., $p = 2$ and $p = \infty$) to conclusions involving a range of p (e.g., $2 < p < \infty$). Here BMO plays the key role as the necessary substitute for L^∞.

5.1 We shall now briefly motivate the use of such an interpolation theorem by citing a specific example of its applicability. Here we shall only state, without proof, the requisite facts of this example. Later, in Chapter 9, we return to it in detail but in a much more general setting, when we study it in the context of Fourier integral operators.

We consider the uniform (Lebesgue) measure on the unit sphere in $\mathbf{R}^n$, normalized so that its total mass is one, and we denote it by $d\sigma$. Our focus is the convolution operator

$$f \mapsto f * d\sigma.$$

Of course, $(f * d\sigma)(x)$ is the mean-value of f taken over the unit sphere centered at x. Besides its intrinsic interest (from its simple symmetric structure), this operator and its variants arise in many other problems, such as the theory of the wave equation.†

Now what are the L^p estimates for this operator? To begin with, it is obvious that $\|f * d\sigma\|_{L^p} \le \|f\|_{L^p}$, $1 \le p \le \infty$, because $d\sigma$ has total mass one. However this is far from the best that can be said (if $p \ne 1, \infty$). For example, when $p = 2$, we can use the important fact (whose extension is the subject of Chapter 8, §3) that $\widehat{d\sigma}$, the Fourier transform of $d\sigma$, has the decay estimate

$$|\widehat{d\sigma}(\xi)| \le A|\xi|^{-(n-1)/2}.$$

From this it follows that the operator

$$f \mapsto \left(\frac{\partial}{\partial x}\right)^{\alpha} f * d\sigma$$

is bounded on L^2, whenever $|\alpha| \le (n-1)/2$. A better way of stating this (because it is more precise when n is even) is that the operator

$$U_\gamma(f) = (-\Delta)^{\gamma/2}(f * d\sigma)$$

is bounded on L^2, whenever $0 \le \gamma \le (n-1)/2$.

Interpolating between L^2 and L^∞, one can reasonably expect that U_γ is bounded from L^p to itself if $0 \le \gamma \le (n-1)/p$ and $2 \le p \le \infty$. Such a result indeed holds, but its proof depends (in an essential way) on estimates for the operator $(-\Delta)^{\gamma/2}$, with γ purely imaginary. These operators, which are given by Marcinkiewicz multipliers‡ (as treated in Chapter 1, §6.2, and in §4.1 above), are not bounded from L^∞ to itself, but do map L^∞ to BMO; this accounts for the role of BMO in the interpolation theorem that follows.

To get our result from the general theorem below, we must first multiply U_γ by a holomorphic function $h(\gamma)$ that mitigates the (polynomial) growth of

$$\|(-\Delta)^{\gamma/2}\|_{L^\infty \to \mathrm{BMO}} \quad \text{when } \gamma = it,\ t \in \mathbf{R},\ t \to \pm\infty,$$

and then set $1 - s = 2\gamma/(n-1)$. Here we may take $h(\gamma) = e^{\gamma^2}$; a similar argument occurs in Chapter 9, §1.2.3.

† When $n = 3$, $f * d\sigma$ is the value at $t = 1$ of the solution $u(x,t)$ of the equation

$$\sum_{j=1}^{3} \frac{\partial^2 u}{\partial x_j^2} = \frac{\partial^2 u}{\partial t^2}, \qquad u(x,0) = 0, \qquad \frac{\partial u}{\partial t}(x,0) = f(x).$$

‡ It is often the case when applying complex interpolation that the imaginary direction corresponds to a Marcinkiewicz multiplier (or a related operator).

5.2 The interpolation theorem. We now formulate the interpolation theorem alluded to above. Our assumptions are as follows.

(i) Let S denote the closed strip $0 \le \mathrm{Re}(s) \le 1$ in the complex s-plane. For each $s \in S$, we have a bounded linear operator

$$T_s : L^2(\mathbf{R}^n) \to L^2(\mathbf{R}^n).$$

(ii) T_s is a holomorphic function of s, in the sense that

$$s \mapsto \int_{\mathbf{R}^n} T_s(f)\, g\, dx$$

is continuous in S and analytic in the interior of S, whenever $f, g \in L^2$.

(iii) The operator norms of the T_s are uniformly bounded, i.e., there is an M with

$$\|T_s\|_{L^2 \to L^2} \le M \qquad \text{for all } s \in S.$$

This bound, however, will *not* enter into the quantitative conclusion below.

(iv) The key hypothesis is the existence of a bound A so that

$$\|T_{it}(f)\|_{L^2} \le A\|f\|_{L^2}, \quad f \in L^2,\ -\infty < t < \infty, \quad \text{and} \tag{55}$$

$$\|T_{1+it}(f)\|_{\mathrm{BMO}} \le A\|f\|_{L^\infty}, \quad f \in L^2 \cap L^\infty,\ -\infty < t < \infty. \tag{56}$$

THEOREM 4. *Under the above assumptions we can conclude that*

$$\|T_\theta(f)\|_{L^p} \le A_\theta\|f\|_{L^p}, \qquad f \in L^2 \cap L^p, \tag{57}$$

whenever $0 \le \theta = 1 - \dfrac{2}{p} < 1$. *Here A_θ depends only on A and θ, but not on M.*

5.3 The proof of the theorem follows the same lines as that of the more standard version for L^p spaces,† except that the intervention of the sharp function requires an auxiliary linearizing device.

We fix a measurable function $x \mapsto B_x$, from points in $\mathbf{R}^n$ to balls in $\mathbf{R}^n$, with $x \in B_x$ for all $x \in \mathbf{R}^n$, so that the volumes $|B_x|$ are bounded above and below. We also fix a measurable function $\eta_x(y)$, with $|\eta_x(y)| \le 1$ for $(x,y) \in \mathbf{R}^n \times \mathbf{R}^n$. It is important that the estimates made immediately below are independent of these two functions. Starting with an $f \in L^2$, we set $F^s = T_s(f)$, for s in the strip S, and write‡

$$U^s(f)(x) = \frac{1}{|B_x|}\int_{B_x} [F^s(y) - F^s_{B_x}]\eta_x(y)\, dy.$$

† See, for instance, Chapter 9, §1.2.5 or *Fourier Analysis*, Chapter 5, §4.

‡ Recall that F_B denotes the average of F over B.

The key observations are, first,

$$|U^s(f)(x)| \le (T_s f)^\sharp(x); \tag{58}$$

and, second,

$$\sup |U^s(f)(x)| = (T_s f)^\sharp(x), \tag{59}$$

where the supremum is taken over all possible functions B_x, η_x described above.

Our first main goal will be the estimate

$$\|U^\theta(f)\|_{L^p} \le A'\|f\|_{L^p} \tag{60}$$

where $\theta = 1 - 2/p$, whenever f is a simple function (i.e., a finite linear combination of characteristic functions of sets of finite measure). In proving (60), we may also assume that f has been normalized so that $\|f\|_{L^p} \le 1$. Thus (60) is equivalent to

$$\left|\int U^\theta(f)\, g\, dx\right| \le A' \tag{61}$$

where g is an arbitrary simple function, with $\|g\|_{L^{p'}} \le 1$; here $1/p' + 1/p = 1$.

Now write $f = \sum a_j \chi_{E_j}$ (with E_j disjoint), and set

$$f_s = \sum |a_j|^{(1-s)p/2} \frac{a_j}{|a_j|} \chi_{E_j}.$$

Note that $f_\theta = f$, because $(1-\theta)p/2 = 1$. Also

$$\|f_{it}\|_{L^2}^2 = \sum |a_j|^p \cdot |E_j| = \|f\|_{L^p}^p \le 1.$$

Moreover, $\|f_{1+it}\|_{L^\infty} \le 1$.

Similarly, write $g = \sum b_k \chi_{E'_k}$, and set

$$g_s = \sum |B_k|^{(s+1)p'/2} \frac{b_k}{|b_k|} \chi_{E'_k}.$$

Then, as is easily verified, $g_\theta = g$, while $\|g_{it}\|_{L^2} \le 1$ and $\|g_{1+it}\|_{L^1} \le 1$.

We now consider the function $I(s)$ defined in the strip S, given by

$$\begin{aligned} I(s) &= \int U^s(f_s) g_s\, dx \\ &= \sum_{j,k} |a_j|^{(1-s)p/2} \frac{a_j}{|a_j|} \cdot |b_k|^{(1+s)p'/2} \frac{b_k}{|b_k|} \int U^s(\chi_{E_j})\chi_{E'_k}\, dx. \end{aligned}$$

From this and our assumptions, it is clear that $I(s)$ is continuous in S, analytic in the interior, and bounded throughout the strip. Moreover

$$\begin{aligned} |I(t)| &\le \|U^{it}(f_{it})\|_{L^2} \cdot \|g_{it}\|_{L^2} \\ &\le \|(T_{it} f_{it})^\sharp\|_{L^2} \le c\|M(T_{it} f_{it})\|_{L^2} \le c\|T_{it} f_{it}\|_{L^2} \le \bar{A}. \end{aligned}$$

These inequalities are (in order) consequences of (58), the L^2 boundedness of the maximal operator M, and the hypothesis (55) of the theorem. Similarly

$$|I(1+it)| \le A,$$

if we use the hypothesis (56). Applying the three lines theorem,† we get

$$|I(\theta)| \le \bar{A},$$

which is (61), and therefore gives (60). Notice that our estimates up to this point have been independent of the functions B_x and η_x. If we now invoke (59), we obtain

$$\|(T_\theta f)^\sharp\|_{L^p} \le \bar{A}\|f\|_{L^p}.$$

We are now in a position to use the inverse inequality for the sharp function (Theorem 2 in §2.2, and the remark that follows with $p_0 = 2$). The result is our conclusion (57), for all f that are simple. However if $f \in L^2 \cap L^p$, we can find a sequence of simple functions $\{f_k\}$ so that $f_k \to f$ in L^p norm, and this establishes our desired conclusion in general.

6. Further results

A. Properties of BMO

6.1 (a) Let P be any polynomial on $\mathbf{R}^n$. Then the function $\log|P|$ belongs to $\mathrm{BMO}(\mathbf{R}^n)$. Moreover, we can choose a bound for its BMO norm that is independent of the coefficients and depends only on the degree of P and on the dimension n. Stein [1986]; also see §6.5 in Chapter 5, where the assertion is deduced from the fact that $|P|$ satisfies a "reverse Hölder" inequality.

(b) Another proof can be given using the following idea. Suppose we decompose $\mathbf{R}^n = \mathbf{R}^{n_1} \times \mathbf{R}^{n_2}$, $x = (x^1, x^2)$, $x^j \in \mathbf{R}^{n_j}$, with $n = n_1 + n_2$. Whenever $x^1 \mapsto f(x^1, x^2)$ is in $\mathrm{BMO}(\mathbf{R}^{n_1})$ uniformly in x_2, and $x^2 \mapsto f(x^1, x^2)$ is in $\mathrm{BMO}(\mathbf{R}^{n_2})$ uniformly in x_1, then $f \in \mathrm{BMO}(\mathbf{R}^n)$. Cwickel [1987].

† In effect, the maximum modulus theorem, which holds in this context.

6.2 That for $f \in$ BMO, $g \in H^1$, the pairing (3) may not be given by an absolutely convergent integral can be seen as follows. Suppose $g_0(r)$, $0 < r < \infty$, is a positive decreasing function of r, with $\int_0^\infty g_0(r)\,dr < \infty$. Then the function $g(x) = \text{sign}(x)g_0(|x|)$ is in $H^1(\mathbf{R}^1)$. Thus we need only choose $g_0(r) = r^{-1}(\log r^{-1})^{-1-\varepsilon}$ for $0 < r < 1/2$, $g_0(r) = 0$ for $r \geq 1/2$, and $f(x) = \log|x|$ to see that (3) does not converge absolutely.

To prove that $g \in H^1$, set

$$a_k(x) = \frac{2^{-k-2}}{g_0(2^k)} g_0(|x|)\,\text{sign}(x), \qquad 2^k \leq |x| < 2^{k+1},$$

$a_k(x) = 0$ elsewhere; then a_k is an atom and $g = \sum \lambda_k a_k$, with $\lambda_k = g_0(2^k)2^{k+2}$ and $\sum \lambda_k \leq 8\int_0^\infty g_0(r)\,dr$. A similar result holds in $\mathbf{R}^n$: Whenever Ω is a bounded function, homogeneous of degree 0, and $\int_{|x|=1} \Omega(x)\,d\sigma(x) = 0$, then

$$g(x) = \Omega(x)\frac{g_0(|x|)}{|x|^{n-1}} \in H^1(\mathbf{R}^n).$$

6.3 The following are examples of elements of BMO.

(i) $(I_{-n}f)(x) = c_n \int_{\mathbf{R}^n} \log|x-y|\, f(y)\,dy$ for $f \in L^1(\mathbf{R}^n)$; see also Chapter 3, §5.21. A variant is $J_{-n}f = f * G_n$, where $\widehat{G}_n(\xi) = (1 + 4\pi^2|\xi|^2)^{-n/2}$. An interpretation of these results, when n is even, is the following assertion: Suppose f is locally integrable and $\int_{\mathbf{R}^n} |f(x)|(1+|x|)^{-n-1}\,dx < \infty$. Assume that $\Delta^{n/2} f \in L^1(\mathbf{R}^n)$ in the sense of distributions. Then $f \in \text{BMO}(\mathbf{R}^n)$.

(ii) The next four examples are periodic functions on $\mathbf{R}^1$. Suppose $f(x) = \sum_{-\infty}^{\infty} a_k e^{ikx}$, where $|a_k| \leq k^{-1}$. Then $f \in \text{BMO}(\mathbf{R}^1)$.

(iii) A similar conclusion if $f = \sum_{k=0}^{\infty} a_k e^{i2^k x}$, with $\sum |a_k|^2 < \infty$.

(iv) Suppose that $\sum b_k e^{ikx} \in$ BMO and $b_k \geq 0$; then $\sum a_k e^{ikx} \in$ BMO whenever $|a_k| \leq b_k$.

(v) If P is a real-valued polynomial on $\mathbf{R}^1$, then $\sum_{k=1}^{\infty} \dfrac{e^{iP(k)x}}{k}$ is in BMO.

Assertions (ii) and (iii) are due to C. Fefferman; (iv) was proved by Journé using the criterion §6.4 below. For (v), see Stein and Wainger [1990].

6.3^a When $Tf = f * K$ is a singular integral satisfying (27), we have seen that T extends to $L^\infty(\mathbf{R}^n)$, mapping that space to BMO. Here are three amplifications of this result.

(a) The same conclusion holds for a more general class of T that are not necessarily translation-invariant. Indeed, suppose T is bounded on $L^2(\mathbf{R}^n)$, and is representable by $(Tf)(x) = \int_{\mathbf{R}^n} K(x,y)\,f(y)\,dy$ for $x \notin \text{supp} f$, with

$$\int_{|y-x_0| \geq 2|x-x_0|} |K(x,y) - K(x_0,y)|\,dy \leq A$$

(for all x, x_0). Then the dual T^* maps H^1 to L^1, and $T : L^\infty \to$ BMO.

(b) Returning to the translation-invariant case, one can show that T actually extends to map BMO to itself; see §5.25 in the previous chapter; also Peetre [1966].

(c) There is an analogue of the above for the maximal operator. Namely, if $f \in$ BMO then $Mf \in$ BMO, with

$$\|Mf\|_{\text{BMO}} \leq c(\|f\|_{\text{BMO}} + |f_{B_1}|);$$

here B_1 is the unit ball. Bennet, DeVore, and Sharpley [1981].

6.4 That "all" BMO functions arise in the form $T(f)$, with T as in §4.1 and f bounded, is contained in the following characterization of BMO: for every $f \in \text{BMO}(\mathbf{R}^n)$, there exist $f_0, f_1, \ldots, f_n \in L^\infty(\mathbf{R}^n)$ so that

$$f = f_0 + \sum_{j=1}^{n} R_j(f_j);$$

here R_j are the Riesz transforms.

In fact, this assertion is essentially the dual to the characterization of H^1 in terms of Riesz transforms, given in §4.3 of the previous chapter. Further details are in C. Fefferman and Stein [1972]; see also §6.16 below.

6.5 For any f defined on $\mathbf{R}^n$, let M_f be the multiplication operator $\phi \mapsto \phi f$. Then $f \in \text{BMO}(\mathbf{R}^n)$ if and only if the commutators $M_f R_j - R_j M_f$ are bounded from $L^2(\mathbf{R}^n)$ to itself, for $1 \leq j \leq n$. Further details are in Coifman, Rochberg, and G. Weiss [1976]; see also §5.23(a) of the previous chapter.

6.6 The defining condition for BMO can be weakened substantially. In particular, for any $p > 0$, the condition

$$\sup_B |B|^{-1} \int_B |f - c_B|^p\,dx < \infty,$$

for appropriate constants c_B, implies that $f \in$ BMO. More generally, suppose that there is a fixed γ, $0 \leq \gamma < 1/2$, and a $\lambda > 0$, so that for all balls B

$$|\{x \in B : |f(x) - c_B| > \lambda\}| \leq \gamma|B|.$$

Then $f \in$ BMO, and the BMO norm of f is comparable to the smallest λ for which the above inequality holds. See John [1964], Strömberg [1979a].

6.7 (a) Suppose $\Phi \in \mathcal{S}$ and $a \in L^\infty$. Let $x \mapsto t(x)$ be an arbitrary measurable function from $\mathbf{R}^n$ to $\mathbf{R}^+$ (a "stopping time"). Then

$$f(x) = \int_{\mathbf{R}^n} a(y)\,\Phi_{t(y)}(x-y)\,dy \in \text{BMO}.$$

(b) A more elaborate version of this is to replace the integral by the sum $\sum \lambda_k \Phi_{t_k}(x - y_k)$. Here $|\lambda_k| \leq M$, and if B_k denotes the ball of radius t_k about y_k, then we assume that $\sum_{B_k \subset B^*} |B_k| \leq c|B|$ for all balls B; here B^* is the double of B.

(c) There is a converse to (a). It states that if $\int \Phi\, dx = 1$, then any BMO function can be so represented, modulo an additive L^∞ function. There is also a corresponding converse to (b).

The proof that the integral in (a) represents a BMO function is a direct consequence of the fact that it can be paired with an H^1 function. The proof of (b) is similar, if one uses the fact that $d\mu = \sum d\mu_k$ is a Carleson measure, where $d\mu_k$ is the unit mass situated at (y_k, t_k). For the converses see Carleson [1976], Garnett and P. Jones [1978]. The converse to (a) gives another proof that BMO is the dual of H^1. The converse to (b) arises in measuring the distance of a BMO function to L^∞; see also §6.20 below.

6.8 The definition $f \in$ BMO requires that the averages $|B|^{-1}\int_B |f - f_B|\, dx$ be uniformly bounded as B ranges over all balls in $\mathbf{R}^n$. If, in addition, we require that these averages tend to zero uniformly as the radius of B tends either to zero or to infinity, we say that f belongs to VMO ("vanishing mean oscillation"). VMO is a closed subspace of BMO. Some further properties of VMO are:

(a) VMO is the closure of C_0 (the continuous functions that vanish at infinity) in BMO.

(b) Let $\Phi \in \mathcal{S}$, $\int \Phi\, dx = 1$. Suppose $f \in$ BMO. Then $f \in$ VMO if and only if $f * \Phi_t \in C_0$ for all $t > 0$, and $\|f - f * \Phi_t\|_{\mathrm{BMO}} \to 0$ as $t \to 0$.

(c) If $f \in C_0$ then $Tf \in$ VMO, where T is as in §4.1.

(d) The pairing (3) (in §1.2) allows one to realize $H^1(\mathbf{R}^n)$ as the dual of VMO$(\mathbf{R}^n)$.

Further details are in Sarason [1975], Coifman and G. Weiss [1977a].

6.9 Invertible linear transformations of $\mathbf{R}^n$ are easily seen to preserve the space BMO$(\mathbf{R}^n)$. When $n > 1$, the most general homeomorphism of $\mathbf{R}^n$ that has this property is given by a *quasi-conformal* mapping. Namely, suppose $F : \mathbf{R}^n \to \mathbf{R}^n$, and let ∇F be its Jacobian matrix. If F satisfies certain regularity conditions and

$$\|\nabla F(x)\|^n \leq K|\det \nabla F(x)| \qquad \text{for each } x,$$

then, whenever $f \in$ BMO, it follows that $f \circ F \in$ BMO with $\|f \circ F\|_{\mathrm{BMO}} \leq c_k \|F\|_{\mathrm{BMO}}$. Conversely, if F is a homeomorphism that enjoys the latter property, then F is quasi-conformal. See Reimann [1974]; for the case $n = 1$, see Chapter 5, §6.8.

B. Tent spaces and square functions

6.10 We shall now describe some further aspects of the theory of "tent spaces". Here our concern is with those spaces that correspond to square functions, in the same way as the space $\mathcal{N}$ (which was the subject of Chapter 2, §2) corresponded to the nontangential maximal function F^*. Some of the basic

definitions involved have already been set down in §4.4 of this chapter: for a given function F on $\mathbf{R}^{n+1}_+$, we define the square function

$$(\mathfrak{S}F)(x) = \left(\int_{\Gamma(x)} |F(y,t)|^2 \frac{dy\, dt}{t^{n+1}}\right)^{1/2},$$

and another dual operator

$$(\mathfrak{T}F)(x) = \sup_{x \in B}\left(\frac{1}{|B|}\int_{T(B)} |F(y,t)|^2 \frac{dy\, dt}{t}\right)^{1/2}.$$

When $0 < p < \infty$, we define $\mathcal{N}^p = \mathcal{N}^p_2$ to consist of all F for which $\mathfrak{S}(F) \in L^p(\mathbf{R}^n)$ with "norm" $\|F\|_{\mathcal{N}^p} = \|\mathfrak{S}(F)\|_{L^p(\mathbf{R}^n)}$. The space $\mathcal{N}^p_r$ can be defined similarly; then $\mathcal{N}^1_\infty$ is the space $\mathcal{N}$ treated in Chapter 2. Note that $\mathcal{N}^1 = \mathcal{N}^1_2$ is the space $\mathcal{N}_2$ that occurred in §4. The analogy between the $\mathcal{N}^p$ spaces and the H^p spaces, $0 < p < \infty$, is brought out by the following properties.

(a) *Duality for* $1 < p < \infty$. The pairing

$$\langle F, G\rangle = \int_{\mathbf{R}^{n+1}_+} F(x,t)\, G(x,t) \frac{dx\, dt}{t}$$

exhibits the duality between $\mathcal{N}^p$ and $\mathcal{N}^q$, if $1 < p, q < \infty$ and $p^{-1} + q^{-1} = 1$.

(b) *Duality for* $p = 1$. The above pairing also gives a duality between $\mathcal{N}^1 = \mathcal{N}_2$ and the space of G for which $\mathfrak{T}(G) \in L^\infty(\mathbf{R}^n)$.

(c) *Alternate definition.* When $2 \leq p < \infty$, we have $F \in \mathcal{N}^p$ if and only if $\mathfrak{T}(F) \in L^p(\mathbf{R}^n)$ and $\|F\|_{\mathcal{N}^p} \approx \|\mathfrak{T}(F)\|_{L^p(\mathbf{R}^n)}$. However, when $p = \infty$, this equivalence breaks down.

(d) *Atomic decomposition.* We define an $\mathcal{N}^p$ atom to be a function a supported in a tent $T(B)$, B a ball in $\mathbf{R}^n$, for which

$$\left(\int_{T(B)} |a(x,t)|^2 \frac{dx\, dt}{t}\right)^{1/2} \leq |B|^{1/2 - 1/p}.$$

If $p \leq 2$ then $a \in \mathcal{N}^p$ with $\|a\|_{\mathcal{N}^p} \leq c$; moreover, each $F \in \mathcal{N}^p$, $p \leq 1$, can be written as $F = \sum \lambda_k a_k$, where the a_k are atoms and $\sum |\lambda_k|^p \leq c\|F\|^p_{\mathcal{N}^p}$.

The proof of (a) depends in part on §4.4.3 of the previous chapter. Duality for $p = 1$ is essentially contained in inequality (38) of the present chapter. The atomic decomposition is a more elaborate version of the argument for the space $\mathcal{N}$ given in Chapter 2. Further details are in Coifman, Y. Meyer, and Stein [1985].

6.11 Beyond these analogies with H^p, there is a genuine identification that comes about via the square functions S_Φ considered in §4.3. Suppose $\Phi \in \mathcal{S}$ and $\int_{\mathbf{R}^n} \Phi\, dx = 0$. Define the mapping π_Φ, initially acting on bounded functions F on $\mathbf{R}^{n+1}_+$ having compact support, by

$$(\pi_\Phi F)(x) = \int_{\mathbf{R}^{n+1}_+} F(y,t)\, \Phi_t(x-y)\, \frac{dy\, dt}{t}.$$

(a) π_Φ extends to a bounded mapping from $\mathcal{N}^p$ to $L^p(\mathbf{R}^n)$, $1 < p < \infty$; and more generally from $\mathcal{N}^p$ to $H^p(\mathbf{R}^n)$ when Φ satisfies the moment conditions $\int x^\alpha\, \Phi(x)\, dx = 0$, for $|\alpha| \leq n(p^{-1} - 1)$, if $p \leq 1$.

(b) If we redefine $\mathcal{N}^\infty$ (in keeping with §6.10(b) above) to consist of those F for which $\mathfrak{T}(F) \in L^\infty(\mathbf{R}^n)$ with $\|F\|_{\mathcal{N}^\infty} = \|\mathfrak{T}(F)\|_{L^\infty}$, then π_Φ extends to a bounded mapping of $\mathcal{N}^\infty$ to $\mathrm{BMO}(\mathbf{R}^n)$.

(c) The mappings π_Φ above are actually onto. More precisely, suppose that Φ is nondegenerate in the sense of §4.3 (see also §6.19 below). Then whenever $f \in H^p$, $0 < p < \infty$, we can choose $F \in \mathcal{N}^p$ with $f = \pi_\Phi(F)$ and $\|F\|_{\mathcal{N}^p} \le c\|f\|_{H^p}$. A similar result holds for $\mathcal{N}^\infty$ and BMO.

To prove (a) for $p \le 1$, it suffices to check the assertion for $\mathcal{N}^p$ atoms. For the converse, suppose $f \in H^p$, $p \le 1$. Since Φ is nondegenerate, there exists a $\Psi \in \mathcal{S}$, $\int \Psi = 0$, with $\int_0^\infty \Phi_t * \Psi_t \, dt/t = \delta$. Thus $f = \pi_\Phi(F)$, where $F(x,t) = (f * \Psi_t)(x)$. Now $\mathfrak{S}(F) = S_\Psi(F)$, while $S_\Psi(f) \in L^p(\mathbf{R}^n)$; the latter is a consequence of the vector-valued version of Theorem 4 in the previous chapter. Thus $\mathfrak{S}(F) \in L^p$ and $F \in \mathcal{N}^p$. The statement (b) and its converse are variants of the quasi-orthogonal expansion of BMO functions treated in §4.5.

For further details see Coifman, Y. Meyer, and Stein [1985]; some relevant earlier ideas are in Calderón [1977b].

6.12 The properties of the tent spaces can be used to clarify various points in the theory of H^p spaces. We give two examples of this.

(a) *Another proof of the atomic decomposition of H^p.* The argument that we gave in §2 of the previous chapter was based on the grand maximal function. Alternatively, one can take as a starting point the area integral S that was discussed in §4.4 of that chapter.

First, if $\Psi = (\partial P/\partial t)|_{t=1}$, with P the Poisson kernel, one can always find a C^∞ function Φ, supported in the unit ball and having a prescribed number of vanishing moments, so that

$$\int_0^\infty \Psi_t * \Phi_t \frac{dt}{t} = \delta;$$

see §6.19 below.

Next, if $f \in H^p$, $p \le 1$, and u is the Poisson integral of f, it follows that

$$(S_\Psi f)(x) = \left(\int_{\Gamma(x)} \left|\frac{\partial u}{\partial t}(y,t)\right|^2 t^{1-n}\, dy\, dt\right)^{1/2} \le (Su)(x) \in L^p(\mathbf{R}^n).$$

Thus if $F(x,t) = t \cdot \partial u/\partial t(x,t)$, we have $F \in \mathcal{N}^p$ and can decompose F into $\mathcal{N}^p$ atoms. That is, by §6.10(d), $F = \sum \lambda_k a_k$, where a_k is supported in $T(B_k)$.

Now let $\tilde a_k = \pi_\Phi(a_k)$. Since Φ is supported in the unit ball, $\tilde a_k$ is supported in $\tilde B_k$, the double of B_k. Moreover $\int_{\tilde B_k} |\tilde a_k|^2\, dx \le c|\tilde B_k|^{1-2/p}$; also $\tilde a_k$ has the required number of moment conditions because Φ does. As a result, $f = \sum \lambda_k \tilde a_k$ is an atomic decomposition of f. Note however that the atoms here have their size controlled by an L^2 norm and not an L^∞ norm; for this point see §5.7 of the previous chapter. A variant of this argument is given by Wilson [1985].

(b) *Another proof of the duality of H^1 and BMO.* The approach used in §1 of this chapter depended on the atomic decomposition for H^1. An alternative argument is as follows. If $\Phi = (\partial P/\partial t)|_{t=1}$ and $\Psi = 4\Phi$, we have $\int_0^\infty \Phi_t * \Psi_t \, dt/t = \delta$; see §6.19 below. Thus, for appropriate f and g,

$$\int_{\mathbf{R}^n} fg\, dx = 4\int_0^\infty (f * \Phi_t)(g * \Phi_t)\frac{dt}{t},$$

and as a result of §4.4,

$$\left|\int fg\, dx\right| \le c\|\mathfrak{S}(g)\|_{L^1(\mathbf{R}^n)}\|\mathfrak{T}(F)\|_{L^\infty(\mathbf{R}^n)}$$

with $F(x,t) = (f * \Phi_t)(x)$, $G(x,t) = (g * \Phi_t)(x)$. By the argument of Theorem 3(a), we have $\mathfrak{T}(F) \in L^\infty(\mathbf{R}^n)$ if $f \in \mathrm{BMO}$. Moreover, if u is the Poisson integral of g, then $\mathfrak{S}(G) = S_\Phi(g) \le S(u)$. Hence, Chapter 3 §4.4 implies $\|S(u)\|_{L^1} \le c\|g\|_{H^1}$, and we have that

$$\left|\int_{\mathbf{R}^n} fg\, dx\right| \le c'\|f\|_{\mathrm{BMO}}\|g\|_{H^1}.$$

For this, and still another proof of duality, see C. Fefferman and Stein [1972].

6.13 (a) The space $\mathcal{N}^p$ is independent of the aperture of the cone used in its definition. Indeed, for $a > 0$, define

$$(\mathfrak{S}_a F)(x) = \left(\int_{\Gamma_a(x)} |F(y,t)|^2 \frac{dy\, dt}{t^{n+1}}\right)^{1/2};$$

here $\Gamma_a(x) = \{(y,t) \in \mathbf{R}^{n+1}_+ : |x-y| < at\}$. Then

$$\|\mathfrak{S}_a(F)\|_{L^p} \approx \|\mathfrak{S}_b F\|_{L^p}$$

whenever $0 < p < \infty$ and $0 < b < a < \infty$. The proof uses §4.4.3 of the previous chapter when $p \le 2$, and the fact that

$$\int_{\mathbf{R}^n} [(\mathfrak{S}_a F)(x)]^2 \phi(x)\, dx \le c\int_{\mathbf{R}^n} [(\mathfrak{S}_b F)(x)]^2 (M\phi)(x)\, dx$$

when $2 \le p < \infty$. These results break down for $p = \infty$. See Coifman, Y. Meyer, and Stein [1985].

(b) In particular, if u is harmonic and we take $F(x,t) = t|\nabla u(x,t)|$, we have the equivalence between area integrals $\|S_a(u)\|_{L^p} \approx \|S_b(u)\|_{L^p}$, $0 < p < \infty$. Here, too, the inequalities break down at $p = \infty$. A counterexample is due to P. Jones [1984].

6.14 Relative distributional inequalities (as described in §3.5) hold for the area integral and nontangential maximal function treated in §4.4 of the previous chapter. Indeed, with appropriate a and b, we can assert that for $\gamma \leq 1$ and all $\alpha > 0$:

$$|\{x : (S_b u)(x) > \alpha; u_a^*(x) \leq \gamma\alpha\}| \leq c\gamma^2 |\{x : (S_b u)(x) > \gamma\alpha\}|$$

and

$$|\{x : u_b^*(x) > \alpha; (S_a u)(x) \leq \gamma\alpha\}| \leq c\gamma^2 |\{x : u_b^*(x) > \gamma\alpha\}|.$$

The bounds $c\gamma^2$ that occur above can be improved to $c\gamma^k$ for any $k \geq 0$, and in fact to exponential decay; see Burkholder and Gundy [1972], R. Fefferman, Gundy, Silverstein, and Stein [1982], Murai and Uchiyama [1986].

6.15 (a) It follows as a result of the above considerations that

$$\|S(u)\|_{L^p} \leq O(p^{1/2}) \|u^*\|_{L^p}, \qquad \text{as } p \to \infty.$$

As a consequence, if $f \in L^\infty$ with compact support, and u is the Poisson integral of f, then for some $c > 0$, $\exp(c[Su(x)]^2)$ is locally integrable.

(b) In the converse direction, whenever f has compact support and $S(u) \in L^\infty$, then $\exp(c|f|^2)$ is locally integrable for some $c > 0$.

For (a), see Burkholder [1979]; conclusion (b) is in S.-Y. A. Chang, Wilson, and Wolff [1985].

C. Miscellaneous

6.16 We may ask which collections of singular integral operators (besides $\{I, R_1, \ldots, R_n\}$) characterize BMO($\mathbf{R}^n$) and $H^1(\mathbf{R}^n)$, in the sense of §6.4 (and §4.3 of the previous chapter).

(a) Some examples are: in $\mathbf{R}^2$, let $R = R_1 + iR_2$, then $\{I, R\}$ has this property, as do analogous systems in $\mathbf{R}^n$, $n \geq 3$, built out of "spinors". However, the pair $\{I, R^2\}$ does not enjoy this property. See Stein and G. Weiss [1968], and Gandulfo, García-Cuerva, and Taibleson [1976] for the negative example.

(b) A general result is the following. Suppose T_j, $1 \leq j \leq N$, are singular integral operators of the kind described in §8.18 of Chapter 1; i.e., $T_j f = f * K_j$, where $\widehat{K}_j = m_j$ is homogeneous of degree 0 and C^∞ away from the origin. Then a necessary condition that $\{T_1, T_2, \ldots, T_N\}$ characterize BMO and H^1 as above is that the vectors $(m_1(\xi), \ldots, m_N(\xi))$ and $(m_1(-\xi), \ldots, m_N(-\xi))$ be linearly independent for every $\xi \neq 0$. Janson [1977].

(c) The general condition above is also sufficient. This is a deep result which, unlike the case involving Riesz transforms, cannot be obtained by harmonic majorization. See Uchiyama [1982]. A significant step in the proof is the utilization of the quasi-orthogonal decomposition of BMO functions given in §4.5.

6.17 It is useful to be able to interpolate H^p spaces by the complex method. We will not formulate the most general result of this kind here, but limit ourselves to a version that already contains the substance of the matter.

Suppose $0 < p_0 < p_1 < \infty$ are given, and for each s, $0 \leq \mathrm{Re}(s) \leq 1$, the operator T_s is defined as a mapping from $H^{p_0} \cap H^{p_1}$ to locally integrable functions on $\mathbf{R}^n$. Assume that $\{T_s\}$ is analytic, in the sense that whenever $f \in H^{p_0} \cap H^{p_1}$ and g is bounded with compact support, the function $s \mapsto \int_{\mathbf{R}^n} T_s(f)\, g\, dx$ is analytic in $0 < \mathrm{Re}(s) < 1$, and is continuous and bounded in the closed strip. We assume also that the following boundary inequalities are satisfied:

$$\|T_s(f)\|_{L^{q_0}} \leq A \|f\|_{H^{p_0}}, \qquad \mathrm{Re}(s) = 0,$$

and

$$\|T_s(f)\|_{L^{q_1}} \leq A \|f\|_{H^{p_1}}, \qquad \mathrm{Re}(s) = 1.$$

Then we can conclude that $\|T_\theta(f)\|_{L^q} \leq A_\theta \|f\|_{H^p}$ whenever $0 < \theta < 1$, $1/q = (1-\theta)/q_0 + \theta/q_1$, and $1/p = (1-\theta)/p_0 + \theta/p_1$.

In the case $n = 1$, this was proved by "complex methods" in Stein and G. Weiss [1957]; see also Calderón and Zygmund [1950]. For $n > 1$, with $p_0 = 1$ and $p_1 > 1$, it follows from its dual, which can be proved by the argument in §5.2; see C. Fefferman and Stein [1972]. For the general case, consult Calderón and Torchinsky [1977]. One approach is via tent spaces and §6.11, since a corresponding complex interpolation theory holds for the space $\mathcal{N}^p$; see Coifman, Y. Meyer, and Stein [1985]. The results above can be completed to include the limiting case corresponding to $p_1 = \infty$ (which space can be taken to be either BMO or L^∞). This can be based in part on the reiteration theorem of Wolff [1982], or the proofs given in Janson and P. Jones [1982], P. Jones [1981].

6.18 While the atomic decomposition for H^p spaces is limited to $p \leq 1$, there is a substitute version that is valid for $1 < p < \infty$:

Suppose $f \in L^p(\mathbf{R}^n)$. Then there exist functions f_k, balls B_k, and constants λ_k so that $f = \sum f_k$, while $|f_k| \leq \lambda_k \chi_{B_k}$, $\int f_k\, dx = 0$, and

$$\Big\| \sum_k \lambda_k \chi_{B_k} \Big\|_{L^p} \leq c_p \|f\|_{L^p}.$$

This result is due to Uchiyama. Extensions may be found in Janson and P. Jones [1982]. The result can be used to give an alternate proof of Theorem 2.

6.19 We shall now deal with the question of the existence of a pair of functions Φ and Ψ so that the reproducing identity

$$\int_0^\infty \Phi_t * \Psi_t \frac{dt}{t} = \delta$$

holds; here δ is the Dirac delta function.

(a) Suppose $\Phi \in \mathcal{S}$ and $\int \Phi\, dx = 0$. Then Φ is nondegenerate (i.e., there is a $\Psi \in \mathcal{S}$ with $\int \Psi\, dx = 0$ so that the above holds) if and only if $\widehat{\Phi}$ does not vanish identically on any of the rays emanating from the origin (i.e., for every $\xi \neq 0$, there exists a $t > 0$ with $\widehat{\Phi}(t\xi) \neq 0$).

The necessity is obvious because $1 = \int_0^\infty \widehat{\Phi}(t\xi)\widehat{\Psi}(t\xi)\, dt/t$. For the sufficiency, note that there exists a nonnegative, compactly supported, C^∞ function η that vanishes near the origin, so that $|\widehat{\Psi}(\xi)|^2\eta(\xi)$ does not vanish identically on any ray through the origin. Then take

$$\widehat{\Psi}(\xi) = \frac{\overline{\widehat{\Phi}(\xi)}\eta(\xi)}{h(\xi)} \qquad \text{where} \quad h(\xi) = \int_0^\infty |\Phi(t\xi)|^2\eta(t\xi)\,\frac{dt}{t}.$$

(b) In particular, if Φ is radial (and does not vanish identically), we can take $\Psi = c\Phi$ for an appropriate constant c.

(c) If $\Phi = (\partial P_t/\partial t)|_{t=1}$, then the above identity holds with $\Psi = 4\Phi$. This is because $\widehat{P}_t(\xi) = e^{-2\pi t|\xi|}$.

(d) If $\Phi \in L^1$ is radial (and does not vanish identically), $\widehat{\Phi}(\xi) \geq 0$ for all ξ, and N is given, then there is a $\Psi \in C^\infty$, supported in the unit ball, that enjoys the moment conditions $\int_{\mathbf{R}^n} x^\alpha\, \Psi(x)\, dx = 0$ for $0 \leq |\alpha| \leq N$, so that $\int_0^\infty \Phi_t * \Psi_t\, dt/t = \delta$.

Indeed, let Ψ' be *any* radial C^∞ function, supported in $|x| \leq 1/2$, satisfying the moment conditions, that does not vanish identically. Then we can take $\Psi = c\Psi' * \Psi'$, for an appropriate c.

6.20 Suppose $f \in$ BMO, $\Phi \in \mathcal{S}$, and $\int \Phi\, dx = 1$. Notice that while $f * \Phi_t \to f$ as $t \to 0$ and $f * \Phi_t \to 0$ as $t \to \infty$ in the *weak* sense (i.e., when tested against H^1), it is not true that such convergence is in the norm of BMO; to see this, consider $f(x) = \log|x|$. These remarks raise the following question: Given a BMO function, what is its distance (in the BMO norm) from the subspace L^∞? In this regard one has:

Suppose $f \in$ BMO. We know by §1.3 that there are $\mu > 0$ so that

$$\sup_B |\{x \in B : |f - f_B| > \alpha\}| \leq e^{-\alpha/\mu} \cdot |B|$$

holds for all sufficiently large α. The infimum of such μ is comparable with

$$\inf_{g \in L^\infty} \|f - g\|_{\mathrm{BMO}}.$$

See Garnett and P. Jones [1978].

6.21 A part of the theory of H^p spaces extends to the general setting described in §1 and §2 of Chapter 1. This generalization can be carried out if one takes as one's point of departure the "atomic" definition of H^p. This is carried out in Coifman and G. Weiss [1977a], Macias and Segovia [1979], Uchiyama [1980], Strömberg and Torchinsky [1989].

If the context is not as general, a richer theory can be developed. See Calderón and Torchinsky [1975] and [1977], which deal with the nonisotropic and translation-invariant setting of Chapter 1, §2.3. For the theory in the framework of homogeneous groups (as in Chapter 1, §2.5 and Chapter 13, §5), see Folland and Stein [1982].

D. Martingales and wavelets

6.22 Ideas from probability theory, in particular martingales and Brownian motion, have had a strong connection with the subject treated in the last four chapters. In fact, general concepts of real-variable theory have been borrowed by probability theory, and this debt has been repaid several times. While it is beyond the scope of our work to explain in detail the related notions of probability theory, we will try to indicate briefly the main concepts that are relevant to our subject.

For the sake of simplicity, we take $\mathbf{R}^n$, equipped with its usual Lebesgue measure, as our "probability space". One considers on it an increasing sequence $\mathcal{B}_k$, $-\infty < k < \infty$, of collections of measurable sets (i.e., $\mathcal{B}_k \subset \mathcal{B}_{k+1}$). We suppose that each $\mathcal{B}_k$ is closed under complements and countable unions, and that it is generated by its elements of finite measure.

For each k, we then define the "conditional expectation" operator E_k as follows. For each locally integrable f, we take $E_k(f) = E(f|\mathcal{B}_k)$ to be the (essentially unique) $\mathcal{B}_k$-measurable function for which $\int_{\mathbf{R}^n} E_k(f)\, g\, dx = \int_{\mathbf{R}^n} f\, g\, dx$ for all $\mathcal{B}_k$-measurable functions g that are bounded and are supported on a set of finite measure.

The E_k then form an increasing sequence of orthogonal projections on $L^2(\mathbf{R}^n)$. Associated to any f that is integrable on sets of finite measure is the sequence $\{f_k = E_k(f)\}$. More generally, we may consider sequences f_k for which $f_k = E_k(f_j)$ whenever $j \geq k$; such sequences are called *martingales*. The structure given by the collection $\{\mathcal{B}_k\}$ and the associated projections $\{E_k\}$ is also called a martingale.

(a) The maximal theorem, comprising a weak-type (1,1) inequality and L^p boundedness (for $1 < p \leq \infty$), holds in this setting. In fact,

$$|\{x \in \mathbf{R}^n : \sup_k |(E_k f)(x)| > \alpha\}| \leq \frac{\|f\|_{L^1}}{\alpha},$$

and

$$\|\sup_k |(E_k f)(x)|\,\|_{L^p(dx)} \leq A_p \|f\|_{L^p}, \quad 1 < p \leq \infty.$$

(b) There is also an analogue of the square function, namely

$$S(f) = \Big(\sum_k |E_k f - E_{k-1} f|^2\Big)^{1/2}.$$

The basic L^p equivalence holds:

$$\|S(f)\|_{L^p} \approx \|f\|_{L^p}, \quad 1 < p < \infty.$$

(c) The prototypical example of the above is the *dyadic* martingale on $\mathbf{R}^n$. Here the collection $\mathcal{B}_k$ is the algebra of sets generated by the mesh $\mathcal{M}_k$ of dyadic cubes having side length 2^{-k} (as in §3.1). In this case $\sup_k E_k|f| = M^\Delta(f)$.

The formulation and proof of the maximal theorem is in Doob [1953], and the result for the square function is in Burkholder [1966] (see also the exposition in Stein [1970a], as well as the literature cited below). Example (c) (for $n = 1$) was already treated in Paley [1932]. Paley's motivation arose from the fact that, if f is supported on $[0,1]$, then $E_k(f)$ is the partial sum of order 2^k of the orthogonal expansion of f in the Walsh basis. The Walsh basis is the set of all characters on $[0,1]$, when $[0,1]$ is given the usual identification with the countably infinite product of two-point groups $\prod_0^\infty \mathbf{Z}/2\mathbf{Z}$.

6.23 Besides the L^p theory of martingales, there are also variants of some of the results of H^p theory. While a part of this can be done in the framework discussed in §6.22, we shall state the results in a simpler setting, in which we assume that the martingales are "regular" in the following sense: There exists a $c > 0$ so that whenever $A \in \mathcal{B}_k$, there is a $B \in \mathcal{B}_{k-1}$ with $B \supset A$ and $|B| \leq c|A|$. Note that the dyadic martingales (and the examples given in §6.24 below) are regular. We remark that Lemma 1 of §3.1 (and in particular, the critical condition (ii)) holds for regular martingales. Assuming then that our martingales are regular, we can assert:

(a) One has the *a priori* equivalence

$$\|S(f)\|_{L^p} \approx \|\sup_k |E_k f|\,\|_{L^p}, \quad 0 < p < \infty.$$

This is in analogy with §4.4 of the previous chapter.

(b) The relative distributional inequalities in §6.14 have analogues that link $S(f)$ and $\sup_k |E_k(f)|$ in a corresponding way.

(c) If, in this context, we consider H^1 to be the set of functions $f \in L^1$ for which $\sup_k |E_k f| \in L^1$, then we can identify its dual with the set of f for which

$$\sup_k E_k(|f - E_k f|^2) \in L^\infty.$$

For (a) see Burkholder and Gundy [1970], Davis [1970]; (b) is in Burkholder [1979]; for (c) consult C. Fefferman and Stein [1972], Garsia [1973], and Herz [1974b]. Note that in the case of dyadic martingales on $\mathbf{R}^1$, if Q^0 is a cube of size 2^{-k}, then the value of $E_k(|f - E_k f|^2)$ on Q^0 equals $|Q^0|^{-1} \sum_{Q \subset Q_0} |a_Q|^2$, when $f \sim \sum a_Q h_Q$; here the h_Q are the Haar functions (see §4.5.4).

6.24 There are two ways to link martingale theory with the analysis described in the last four chapters. The first is straightforward and direct.

(a) Suppose we consider the dyadic martingale in $\mathbf{R}^n$. We have already treated the dyadic version of BMO in §3. Here we ask: What is the relation between M^Δ (see §3.1 and §6.22(c)) and the standard maximal operator M? It is obvious that $(M^\Delta f)(x) \leq c(Mf)(x)$ for all x, but the reverse inequality is easily seen to fail. Control of M by M^Δ is possible, however, in terms of distributional inequalities. In fact,

$$|\{x : (Mf)(x) > c\alpha\}| \leq c_1 |\{x : (M^\Delta f)(x) > \alpha\}|, \quad \text{for all } \alpha > 0.$$

To prove this let Q_j be the disjoint cubes so that $\cup_j Q_j = \{x : (M^\Delta f)(x) > \alpha\}$ (as in §3.1). If B_j is the ball with the same center as Q_j but twice its diameter, it can be shown that $\{x : (Mf)(x) > c\alpha\} \subset \cup_j B_j$.

(b) Some of these considerations are applicable to the general structure described in §1 of Chapter 1. Indeed, for some $\delta > 0$ and $C > 0$, there exists a collection of sets $\{Q_j^k\}$ so that:

(i) $B(x_j^k, \delta^k) \subset Q_j^k \subset B(x_j^k, c\delta^k)$ for some x_j^k.

(ii) If $Q_j^k \cap Q_{j'}^\ell \neq \emptyset$ and $\ell \geq k$, then either $\ell > k$ and $Q_{j'}^\ell \subset Q_j^k$ or $\ell = k$ and $j = j'$.

(iii) $\mathbf{R}^n = \bigcup_{k,j} Q_j^k$.

If we take $\mathcal{B}_k$ to be the collection of sets generated by the Q_j^k, then we have a martingale structure, and an analogue of (a) holds. For the construction of the $\{Q_j^k\}$, see David [1988], Christ [1990].

(c) The examples above are regular (in the sense of §6.23), as are those that arise in p-adic theory; for this see, e.g., Taibleson [1975], Chao and Janson [1981].

6.25 There is a deeper connection between martingales and our subject; it involves Brownian motion. This fundamental notion can be formulated as follows. There is an (abstract) probability space Ω so that for each t, $-\infty < t < \infty$, we are given an $\mathbf{R}^N$-valued function B_t on Ω with the following properties:

(i) $B_0 \equiv 0$.

(ii) The increments $B_{t_1} - B_{t_2}, B_{t_2} - B_{t_3}, \ldots, B_{t_{k-1}} - B_{t_k}$, $t_1 < t_2 < \cdots < t_k$, are independent.

(iii) $B_{t_1} - B_{t_2}$ has a Gaussian distribution with variance $t_2 - t_1$.

The basic premise linking Brownian motion, martingale theory, and the analysis of the last four chapters states that whenever u is a harmonic function on $\mathbf{R}^N$, the functions $F_t = u(B_t)$ form a (continuous version of a) martingale. Note that here our underlying probability space is Ω (whose points label the Brownian paths), and not $\mathbf{R}^n$ as in §6.22 above.

This principle (and is variants) can be applied to analysis in $\mathbf{R}^n$ by taking $N = n+1$, and letting u be the Poisson integral of a function (or distribution) f on $\mathbf{R}^n$. Indeed, the first proof of the equivalence of the maximal characterization of H^p in terms of Poisson integrals (as well as that given by square functions) was done in this way (in the case $n = 1$); see Burkholder, Gundy, and Silverstein [1971], also Burkholder and Gundy [1973].

An inherent difficulty in this approach (and part of the initial obstacle to its extension to $n > 1$) comes from the fact that the probabilistic maximal and square operators give functions defined on Ω, while the original objects are functions on $\mathbf{R}^n$. The interested reader may consult the expositions of these and related topics in the paper of Stroock and Varadhan [1974], and in the monographs of Petersen [1977] and Durrett [1984].

6.26 The dyadic martingale, for $n = 1$, leads us naturally to the archetypical example of a "wavelet" expansion. Indeed, if we let $h(x) = 1$ when

$0 < x < 1/2$, $h(x) = -1$ when $1/2 < x < 1$, and $h(x) = 0$ elsewhere, then the double-indexed system $\{h_{k,j}\}$ given by

$$h_{k,j}(x) = 2^{k/2}\, h(2^k x - j)$$

is orthonormal and complete in $L^2(\mathbf{R}^1)$; it is the Haar system. Here the index k indicates the scale of $h_{k,j}$ and the index j determines the position of its support.

The connection with the dyadic martingale comes about as follows. If for $f \in L^2(\mathbf{R}^n)$, we write $f \sim \sum_{k,j} c_{k,j} h_{k,j}$ with $c_{j,k} = \langle f, h_{k,j}\rangle$ to indicate its expansion in the Haar system, we have that

$$(E_k - E_{k-1})f = \sum_j c_{k,j} h_{k,j};$$

hence the conditional expectation E_k is the orthogonal projection onto the subspace generated by the elements of the Haar system of scale $\le k$.

To describe the general theory, we adopt the following n-dimensional notation. For each dyadic interval Q, which we can write as $Q = 2^k Q_0 + j$ (where Q_0 is the "unit interval"), we set $w_Q = w_{k,j}$, $w_{k,j}(x) = 2^{nk/2} w(2^k x - j)$, where $j, k \in \mathbf{Z}$ and w is a given (fixed) function. When the $\{w_Q\}$ form a complete orthonormal system for $L^2(\mathbf{R}^1)$, we speak of a *wavelet* system. There are a number of additional properties one might wish to impose on our system, via its generating function w. These include:

(1) Vanishing of a prescribed number of moments of w.

(2) That w have some amount of smoothness.

(3) That w has compact support, or that it vanishes rapidly at infinity.

While these conditions are to an extent in competition, there are wavelet systems that satisfy (1), (2), and (3) to a substantial degree. Among the principal features of such wavelet expansions are the following, which we state in imprecise terms:

(a) If we write $f \sim \sum c_Q w_Q$ with $c_Q = \langle f, w_Q\rangle$, then the *size* of the c_Q, for the Q near x^0, determines the size of f (or its smoothness) near x^0.

(b) One expression of this principle is in terms of the square function $\tilde S$ attached to the system $\{w_Q\}$. Here

$$(\tilde S f)(x) = \left(\sum_{x\in Q} |c_Q|^2 |Q|^{-1}\right)^{1/2}.$$

Then $\|\tilde S(f)\|_{L^p} \approx \|f\|_{H^p}$, $0 < p < \infty$. Also $f \in \mathrm{BMO}(\mathbf{R}^n)$ exactly when

$$\sum_{Q\subset Q_0} |c_Q|^2 \le c|Q_0|$$

for all dyadic cubes Q_0.

(c) Related to this is the fact that the system $\{w_Q\}$ forms an unconditional basis of L^p, $1 < p < \infty$, and of H^p, $p \le 1$; similarly for the space Λ_γ occurring in Chapter 6, §5.3.

(d) The operators that are "approximately diagonal" in an appropriate basis of wavelets are the general Calderón-Zygmund operators (which are bounded on L^2); these are treated in §3 of Chapter 7.

When $n > 1$, the situation is analogous but more complicated: the system $\{w_Q\}$ must be replaced by $2^n - 1$ such systems $\{w_Q^{(\epsilon)}\}$.

Examples of wavelets were constructed by Strömberg [1981], Lemarié, and Y. Meyer [1986], and Daubechies [1988]. An important motivation for the theory was the proof that H^1 has an unconditional basis, and that different versions of H^1 are, in fact, isomorphic as Banach spaces. For this see Maurey [1980], Carleson [1980]. The general theory of wavelets was put in systematic form by Y. Meyer [1990]; this excellent exposition should be consulted for precise versions of statements (a)–(d) above, as well as for further ramifications of wavelet theory.

The construction of the generating function w (or functions $w^{(\epsilon)}$) for such a system can be a tricky business. In practice, matters can be simplified by proceeding instead in the same way as was done for the quasi-orthogonal decomposition of BMO given in §4.5. One can, with substantial flexibility, construct pairs (ϕ, ψ) so that $\{\phi_Q\}$ and $\{\psi_Q\}$ are "bi-orthogonal", and in particular so that $f \sim \sum s_Q \psi_Q$ with $s_Q = \langle f, \phi_Q\rangle$. For such systems, properties akin to (a)–(d) above also hold. For this, see Frazier and Jawerth [1985], [1988].

Notes

The development of BMO can be briefly schematized as follows. First, the paper of John and Nirenberg [1961], where the notion of BMO originated, together with the proof of the corollary in §3.7. Next, the realization the singular integrals map L^∞ to BMO in Spanne [1966], Peetre [1966], and Stein [1967]. Then, the duality of H^1 and BMO in C. Fefferman [1971]; together with Stein [1971] and C. Fefferman and Stein [1972] (which encompasses the first two papers), this also included the L^p inequality for $f^\#$ (Theorem 2) by the proof given in §3.6, the characterization of BMO by Carleson measures (Theorem 3), and the complex interpolation theorem of §5.

§2. The formulation of the duality inequality (16) is an unpublished result of R. Fefferman, who proved a martingale analogue. The proof here is taken from Folland and Stein [1982]; for some related ideas, see Strömberg [1979a].

§3. The idea of relative distributional inequalities, in differing forms, was exploited in Burkholder and Gundy [1970], [1972], and Burkholder [1979].

§4. For an extension of the proposition in §4.2 for $f^\#$, see Muckenhoupt and Wheeden [1976].

The quasi-orthogonal decomposition described in §4.5 can be found in S.-Y. A. Chang and R. Fefferman [1980].

§*5*. The first interpolation theorem involving analytic families of operators for H^p was valid in the one-dimensional complex setting; see Stein and G. Weiss [1957]. Earlier ideas are in Calderón and Zygmund [1950].

CHAPTER V

Weighted Inequalities

The theory of weighted inequalities for the maximal function and singular integrals is a natural outgrowth of the point of view and methods of the previous four chapters. One of the principal problems it addresses can be stated as follows.

We let M denote the standard maximal operator on $\mathbf{R}^n$,

$$(Mf)(x) = \sup_{r>0} \frac{c_n}{r^n} \int_{|y|\leq r} |f(x-y)|\, dy, \tag{1}$$

and wish to characterize the nonnegative measures $d\mu$ on $\mathbf{R}^n$ so that

$$\int_{\mathbf{R}^n} [Mf(x)]^p \, d\mu(x) \leq A \int_{\mathbf{R}^n} |f(x)|^p \, d\mu(x), \tag{2}$$

for some p, $1 < p < \infty$.

While a variety of special results related to this question (and its singular integral analogue) began to appear more than 50 years ago, the full solution to these problems is more recent: it turns out that there is a class of functions A_p so that (2), and other results of this kind, hold exactly when $d\mu(x) = \omega(x)\, dx$ with $\omega \in A_p$. Thus the identification of A_p and the study of its properties become the main goals of this chapter.

In pursuing these objectives we shall proceed in several stages, following roughly the order in which the understanding of the subject developed. First, we elucidate the definition of A_p, describing several alternate formulations, in part suggested by the requirements of the problem. Next we come to the proof of the main result for the maximal function, and realize that it can be achieved by an unexpected "openness" property of the classes A_p, namely that if $\omega \in A_p$ then automatically $\omega \in A_{p_1}$ for some $p_1 < p$. To prove this last fact, we need to invoke the deepest and most significant part of the whole theory: the idea of "reverse" Hölder inequalities and the fact that a weight belongs to one of the A_p classes exactly when it satisfies such an inequality. Connected to this is a "fairness" principle governing A_p weights: For any cube $Q \subset \mathbf{R}^n$ and any subset $E \subset Q$, the ratio $\omega(E)/\omega(Q)$ is bounded by quantities that depend *only* on the ratio $|E|/|Q|$. It should be mentioned that reverse Hölder inequalities also serve an important role in such diverse areas as quasi-conformal mappings and certain refined estimates for elliptic partial differential equations.

As a consequence of the above results, we show that the A_p classes also answer the question analogous to (2) in which M is replaced by a singular integral operator. In addition, we study further the key A_1 class (already encountered in Chapter 2, §1.3) and show how A_p weights, $p > 1$, can be fashioned from A_1 weights. When $p = 2$ the statement is particularly elegant: a weight is in A_2 exactly when it is a quotient of two A_1 weights.

1. The class A_p

As we have said above, it turns out that (2) holds, for a given measure $d\mu$, exactly when $d\mu$ is absolutely continuous (with respect to Lebesgue measure), $d\mu(x) = \omega(x)\,dx$, and the locally integrable function ω satisfies the A_p inequality,

$$\frac{1}{|B|}\int_B \omega(x)\,dx \cdot \left[\frac{1}{|B|}\int_B \omega(x)^{-p'/p}\,dx\right]^{p/p'} \leq A < \infty, \tag{3}$$

for all balls B in $\mathbf{R}^n$.† Here p' is the dual to p, i.e., $1/p' + 1/p = 1$. To avoid trivialities we assume always that ω is nonzero, which by (3) implies that it can vanish only on a set of measure zero. The smallest constant A for which (3) holds, (denoted by $A_p(\omega)$), is the A_p bound of ω. We next make a series of easy observations about the class A_p.

1.1 If ω belongs to A_p, then so do its dilates ω_δ for $\delta > 0$, where $\omega_\delta(x) = \omega(\delta x)$ and $\delta x = (\delta x_1, \ldots, \delta x_n)$ denotes an isotropic dilation; the A_p bound of ω_δ is the same as that of ω. A similar statement holds if we replace $\omega(x)$ with a translate $\omega(x - h)$, $h \in \mathbf{R}^n$, or if we multiply ω by a positive scalar.

1.2 A useful observation—equivalent to "duality" as we shall see below—is the fact that if $\omega \in A_p$ then the function $\sigma = \omega^{-p'/p}$ belongs to $A_{p'}$, where $1/p + 1/p' = 1$. Also, $A_{p'}(\sigma)^{1/p'} = A_p(\omega)^{1/p}$. In fact, the passage from ω to σ has essentially the effect of reversing the order of the two factors on the left side of (3). We remark also that in the self-dual case $p = 2$, the condition A_2 has a simpler and more symmetric appearance, namely

$$\left[\frac{1}{|B|}\int_B \omega\,dx\right] \cdot \left[\frac{1}{|B|}\int_B \omega^{-1}\,dx\right] \leq A.$$

† Throughout this chapter, we continue to write $|E|$ for the Lebesgue measure of the set E.

1.3 Another noteworthy feature of the A_p classes is that they increase with p: if $\omega \in A_{p_1}$ then also $\omega \in A_{p_2}$ whenever $p_1 < p_2$; moreover $A_{p_2}(\omega) \leq A_{p_1}(\omega)$. This is a direct consequence of the definition (3), Hölder's inequality, and the fact that if $q_j = p_j'/p_j = p_j' - 1$ then $q_2 < q_1$ if $p_1 < p_2$.

1.4 There is an alternate way of defining A_p that is more obviously related to the boundedness (2) of the maximal operator (1). For any locally integrable function f and any ball B, we write f_B for the mean value of f on B, that is $f_B = |B|^{-1}\int_B f(x)\,dx$. Using the abbreviation $\omega(B)$ for $\int_B \omega(x)\,dx$, we get the following equivalent characterization of A_p: the weight ω belongs to A_p exactly when the p^{th} power of the mean value f_B is bounded by the mean value of f^p taken with respect to the measure $\omega(x)\,dx$. More precisely, $\omega \in A_p$ if and only if

$$(f_B)^p \leq \frac{c}{\omega(B)}\int_B f^p\omega\,dx \tag{4}$$

holds for all nonnegative f and all balls B. Moreover, the smallest c for which (4) is valid equals the A_p bound of ω.

Notice first that (4) follows from the A_p maximal inequality (2) (when $d\mu = \omega\,dx$) if we use the fact that $Mf(x) \geq c' f_B$ whenever $x \in B$. To prove that (4) is equivalent to (3), assume first that $\omega \in A_p$. Now

$$f_B = \frac{1}{|B|}\int_B f(x)\,dx = \frac{1}{|B|}\int_B f\,\omega^{1/p}\omega^{-1/p}\,dx.$$

Applying Hölder's inequality with exponents p and p' yields

$$(f_B)^p \leq |B|^{-p}\left[\int_B f^p\omega\,dx\right]\cdot\left[\int_B \omega^{-p'/p}\,dx\right]^{p/p'}.$$

Then (3) gives us (4), with $c \leq A_p(\omega)$. Conversely, if (4) holds, we are tempted to set $f = \omega^{-p'/p}$; then $f^p\omega = \omega^{-p'+1} = \omega^{-p'/p}$ (since $p' - 1 = p'/p$). If we knew that $\int_B \omega^{-p'/p}\,dx$ were finite, we could then establish (3) with $A \leq c$. Instead, we replace f by $(\omega + \varepsilon)^{-p'/p}$ where $\varepsilon > 0$. Since $\int_B (\omega+\varepsilon)^{-p'/p}\,dx \leq \int_B (\omega+\varepsilon)^{-p'}\omega\,dx < \infty$, we get

$$\left[\frac{1}{|B|}\int_B \omega\,dx\right]\cdot\left[\frac{1}{|B|}\int_B (\omega+\varepsilon)^{-p'}\omega\,dx\right]^{p/p'} \leq c.$$

Passing to the limit as $\varepsilon \to 0$ completes the proof.

1.5 It is an immediate consequence of the alternative characterization (4) that, whenever $\omega \in A_p$, then $\omega(x)\,dx$ is a doubling measure; that is

$$\omega(B_2) \le c'\,\omega(B_1), \tag{5}$$

if $B_1 = \{x : |x-y| < \delta\}$ and $B_2 = \{x : |x-y| < 2\delta\}$. To see this, apply (4) with $B = B_2$ and $f = \chi_{B_1}$. This gives (5) with $c' = c\,2^{np}$.

The fact that an element of A_p gives a doubling measure is already an interesting insight, but it does not capture the essence of the matter. For example, measures of the form $\omega(x)\,dx$ with $\omega(x) = |x|^\alpha$ are doubling measures when $-n < \alpha$; but such an ω belongs to A_p only if, in addition, $\alpha < n(p-1)$. See also §6.4.

1.6 The doubling property of ω (and the fact that by §1.2, $\omega^{-p'/p}$ is also a doubling measure), together with the characterization (4), shows that in the definition (3) of A_p, we could have replaced the family of balls by the family of cubes or other such equivalent families.

1.7 Related to the doubling property is the important observation that any A_p weight assigns to a subset F of a ball B a fair share of B's weight; "fair" when compared to the ratio of the Lebesgue measures of F and B. We call this the A_∞ property because, as we shall see in §3.1 and §5.1, it characterizes the union of all the A_p classes. Thus we shall say that $\omega \in A_\infty$ if for any α, $0 < \alpha < 1$, there exists a β, $0 < \beta < 1$, so that, for all balls B and all subsets $F \subset B$,

$$|F| \ge \alpha|B| \quad \Rightarrow \quad \omega(F) \ge \beta\omega(B). \tag{6}$$

Since we have no lower bound for β as $\alpha \to 1$, this condition is primarily of interest for small α. However, as we shall see in §3.1 below, the A_∞ property is actually equivalent to the inequality (6) for a *fixed* pair of constants α, β.

To see that (6) holds for any A_p weight, use the characterization (4) in §1.4, taking $f = \chi_F$. This shows that $(|F|/|B|)^p \le c(\omega(F)/\omega(B))$, which gives (6) with $\beta = \alpha^p/c$.‡

Taking complements in (6) yields an equivalent definition, namely that $\omega \in A_\infty$ if, for all γ with $0 < \gamma < 1$, there exists a δ with $0 < \delta < 1$ so that, for all balls B and all subsets $E \subset B$,

$$|E| \le \gamma|B| \quad \Rightarrow \quad \omega(E) \le \delta\omega(B). \tag{6'}$$

Also observe that (6) and (6′) are equivalent to the corresponding statements with cubes Q replacing the balls B, since any one of these properties implies that ω is a doubling measure. In the sequel, we shall use the definition of A_∞ that is most convenient in each context.

‡ Observe that the constant c in (4) is automatically ≥ 1.

1.8 Another essential feature of the A_p weights ω is that the average fluctuation of the magnitude of ω on every ball is uniformly controlled. A precise version of this assertion is the observation that whenever $\omega \in A_p$, the function $\log\omega$ is in BMO.

We give the proof for $p = 2$. Suppose f is a real function. Then by Jensen's inequality, we have that

$$\exp\left(\frac{1}{|B|}\int_B f\,dx\right) \le \frac{1}{|B|}\int_B e^f\,dx. \tag{7}$$

Set $\lambda(x) = \log\omega(x)$, and apply this inequality to the positive part of $[\lambda(x) - \lambda_B]$ (which we denote by $[\lambda(x)-\lambda_B]^+$). The result is

$$\frac{1}{|B|}\int_B [\lambda(x)-\lambda_B]^+\,dx \le \log\left\{\frac{1}{|B|}\int_B e^{\lambda(x)-\lambda_B}\,dx\right\}. \tag{8}$$

A similar inequality holds for $[\lambda(x)-\lambda_B]^-$, the negative part of $[\lambda(x)-\lambda_B]$. Adding these two inequalities gives

$$\frac{1}{|B|}\int_B |\lambda(x)-\lambda_B|\,dx \le \log[\omega_B\cdot(\omega^{-1})_B] \le A,$$

with $A = \log A_2(\omega)$, showing that $\lambda \in$ BMO. A parallel argument works for any p. For a converse, see §6.2.

1.9 Let us consider the limiting case of the A_p spaces, when p tends to 1. If we pass to the limit in definition (3), noting that

$$\left[\frac{1}{|B|}\int_B \omega^{-p'/p}\,dx\right]^{p/p'} \to \|\omega^{-1}\|_{L^\infty(B)} \qquad \text{as } p \to 1,$$

we are led to consider the space A_1 consisting of nonnegative functions ω so that for all balls B

$$\frac{1}{|B|}\int_B \omega\,dx \le A\,\omega(x), \quad \text{for a.e. } x \in B. \tag{9}$$

This condition is equivalent to

$$M\omega(x) \le A'\omega(x), \tag{10}$$

in the sense that (9) implies (10) with $A' \le A$, while conversely (10) implies (9) with $A \le 2^n A'$.

We have already seen the condition (10) in Chapter 2, §1.3 when we considered our first weighted inequalities for the maximal function. The reader will have no difficulty in verifying that the class A_1 satisfies many of the same properties as do the A_p with $p > 1$. In particular, $A_1 \subset A_p$ for all $p > 1$ and A_1 can be characterized by the analogue of (4): namely, $f \in A_1$ if and only if

$$f_B \le c\omega(B)^{-1}\int_B f\,\omega\,dx \tag{11}$$

for all nonnegative f and all balls B.

2. Two further characterizations of A_p

We present two additional characterizations of the class A_p. Each can be viewed as a more elementary version of the main theorem (involving the maximal function) that is proved below in §3. While not going as far as that theorem, these characterizations do have the virtue of extending directly to the situation of unequal weights (for this point, see §6.11).

2.1 The first result represents historically the initial appearance of the classes A_p. We are concerned with convolution operators

$$f \mapsto T_\varepsilon f = f * \Phi_\varepsilon.$$

For simplicity, we assume that Φ is nonnegative, radial, and (radially) decreasing, with $\int \Phi(x)\,dx = 1$, and we define $\Phi_\varepsilon(x) = \varepsilon^{-n}\Phi(x/\varepsilon)$. We designate the class of such Φ by $\mathcal{R}$. Notice that if Φ is in $\mathcal{R}$, then so is Φ_ε. It is useful to recall that

$$(Mf)(x) = \sup_{\Phi\in\mathcal{R}} |f| * \Phi(x).$$

In fact, if $B_\varepsilon = \{x : |x| < \varepsilon\}$ then $|B_\varepsilon|^{-1}\chi_{B_\varepsilon} \in \mathcal{R}$, and the supremum over these elements of $\mathcal{R}$ is, by definition, equal to $(Mf)(x)$. For the other direction, we need only recall that any element of $\mathcal{R}$ is a limit of weighted averages of the $|B_\varepsilon|^{-1}\chi_{B_\varepsilon}$ (see Chapter 2, §2.1).

PROPOSITION 1. *Suppose $d\mu$ is a nonnegative Borel measure, $1 \le p < \infty$, and Φ belongs to $\mathcal{R}$.*

(a) *If the inequality*

$$\int_{\mathbf{R}^n} |T_\varepsilon f(x)|^p\,d\mu(x) \le A \int_{\mathbf{R}^n} |f(x)|^p\,d\mu(x) \tag{12}$$

holds with A independent of ε, then $d\mu$ is absolutely continuous, $d\mu(x) = \omega(x)\,dx$, and $\omega \in A_p$.

(b) *Conversely, if $d\mu(x) = \omega(x)\,dx$ with $\omega \in A_p$, then* (12) *holds with a bound independent of ε and Φ.*

Proof. Let us first assume that (12) holds with $\Phi_\varepsilon = |B_\varepsilon|^{-1}\chi_{B_\varepsilon}$. We write $d\mu = \omega(x)\,dx + d\nu$ where $d\nu$ is totally singular. If $d\nu \ne 0$, then there is a compact set K with $|K| = 0$ and $\nu(K) > 0$. Let U_n be the open set $\{x : \operatorname{dist}(x, K) < 1/n\}$, and let f_n denote the characteristic function of $U_n \backslash K$. Since these sets decrease and have void intersection, $f_n \to 0$ pointwise everywhere. Noting that (12) implies that μ is finite on compact sets, we apply the dominated convergence theorem and conclude

that $\int |f_n|^p\,d\mu \to 0$. However, $T_{1/n}(f_n)(x) = 1$ whenever $x \in K$. To see this, write

$$T_{1/n}(f_n)(x) = |B_{1/n}|^{-1} \int f_n(y)\,\chi_n(x-y)\,dy,$$

where $\chi_n = \chi_{B_{1/n}}$. Now

$$\int f_n(y)\,\chi_n(x-y)\,dy = \int_{U_n\backslash K} \chi_n(x-y)\,dy = \int_{{}^cK} \chi_n(x-y)\,dy,$$

since $x \in K$ and $|x-y| < 1/n$ imply that $y \in U_n$. Since K has Lebesgue measure zero, the last integral equals $|B_{1/n}|$. This shows that $T_{1/n}(f_n)(x) = 1$ for $x \in K$, and since $\mu(K) > 0$, we get a contradiction with (12). Thus $d\mu(x)$ is absolutely continuous and we may write $d\mu(x) = \omega(x)\,dx$.

Now take $f \ge 0$, and let B be any ball. Let us remark that if $x \in B$ and the radius of B is δ, then the ball centered at x of radius 2δ contains B. Thus if $x \in B$ and $\varepsilon = 2\delta$, then $T_\varepsilon f(x) \ge 2^{-n} f_B$, where f_B denotes the average of f over B. The inequality (12) therefore gives us (4), with $c = 2^{np}A$, and if we invoke the mean-value characterization described in §1.4, we can conclude that $\omega \in A_p$.

One passes from the particular family $\Phi_\varepsilon = |B_\varepsilon|^{-1}\chi_{B_\varepsilon}$ to one arising from a general $\Phi \in \mathcal{R}$ by noting that, if $\Phi \in \mathcal{R}$, then $\Phi(x) \ge c_1\chi_{B_{c_2}}(x)$, for some positive constants c_1 and c_2. Thus, the inequality (12) implies the one used in the previous case (perhaps with a different constant A), and the same conclusion follows.

In proving part (b), we consider first the family $\Phi_\varepsilon = |B_\varepsilon|^{-1}\chi_{B_\varepsilon}$. It then suffices to prove that

$$\int_{\mathbf{R}^n} |T_1 f(x)|^p \omega(x)\,dx \le A \int_{\mathbf{R}^n} |f(x)|^p \omega(x)\,dx, \tag{13}$$

with the constant A depending only on the A_p norm of ω. In fact, once (13) is proved for all $\omega \in A_p$, the same result holds with T_1 replaced by T_ε, in view of the dilation invariance of A_p pointed out in §1.1. We therefore turn to (13), and observe that it suffices to consider nonnegative f. Let $B_1 = \{x : |x| < 1\}$, $B_2 = \{x : |x| < 2\}$. Then if $x \in B_1$, the ball centered at x of radius 1 is contained in B_2. This implies that $T_1 f(x) \le 2^n f_{B_2}$ whenever $x \in B_1$. Hence,

$$\begin{aligned}\int_{B_1} |T_1 f(x)|^p\,\omega(x)\,dx &\le 2^{np}(f_{B_2})^p \cdot \omega(B_1)\\ &\le 2^{np}(f_{B_2})^p \cdot \omega(B_2) \le c' \int_{B_2} (f(x))^p\,\omega(x)\,dx,\end{aligned}$$

because an A_p weight always satisfies (4). We can rewrite what we have just obtained as

$$\int (T_1 f(x))^p \chi_{B_1}(x)\, \omega(x)\, dx \leq c' \int (f(x))^p \chi_{B_2}(x)\, \omega(x)\, dx.$$

However, the class A_p is not only dilation-invariant but is translation-invariant as well. Using this, and the fact that T_1 commutes with translations, we get that

$$\int_{\mathbf{R}^n} (T_1 f(x))^p \chi_{B_1}(x-y)\, \omega(x)\, dx \leq c' \int_{\mathbf{R}^n} (f(x))^p \chi_{B_2}(x-y)\, \omega(x)\, dx,$$

for all $y \in \mathbf{R}^n$.

Finally, we integrate the above in y and interchange the order of integration. The result is (13) with $A = c' \cdot 2^n$, which proves (12) for the family $\Phi_\varepsilon = |B_\varepsilon|^{-1}\chi_{B_\varepsilon}$. The case of general Φ then follows by writing it as the limit of weighted averages of the $|B_\varepsilon|^{-1}\chi_{B_\varepsilon}$.

The last argument leads to the following corollary.

COROLLARY. *Suppose $d\mu$ is a nonnegative Borel measure that satisfies the analogue of* (4), *namely*

$$(f_B)^p \leq \frac{c}{\mu(B)} \int_B f^p\, d\mu, \tag{14}$$

for all nonnegative f and all balls B. Then $d\mu$ is absolutely continuous, $d\mu(x) = \omega(x)\, dx$, and, of course, $\omega \in A_p$.

Indeed, the proof we just gave also shows that (14) implies that the hypothesis (12) of Proposition 1 is satisfied.

2.2 PROPOSITION 2. *Let $d\mu$ be a nonnegative Borel measure and let $1 \leq p < \infty$. Then the maximal operator $f \mapsto M(f)$ is of weak-type $(L^p(d\mu), L^p(d\mu))$, i.e.,*

$$\mu\{x : Mf(x) > \alpha\} \leq \frac{A}{\alpha^p} \int |f(x)|^p\, d\mu(x), \quad \textit{all } \alpha > 0, \tag{15}$$

if and only if $d\mu$ is absolutely continuous, $d\mu(x) = \omega(x)\, dx$, with $\omega \in A_p$.

Proof. Suppose (15) holds. Since M is clearly bounded on L^∞, the Marcinkiewicz interpolation theorem implies that M is bounded from $L^{p_1}(d\mu)$ to itself whenever $p_1 > p$, so (12) holds, with p_1 replacing p. Therefore by Proposition 1, $d\mu$ is absolutely continuous, and we can write $d\mu = \omega\, dx$. Now let $f \geq 0$ be supported in a ball B. If $x \in B$ then the ball centered at x, whose radius is twice that of B, contains B; therefore $Mf(x) \geq 2^{-n} f_B$ for $x \in B$. If we take $\alpha = 2^{-n-1} f_B$ in (15), we obtain

$$(f_B)^p \cdot \omega(B) \leq A \cdot 2^{(n+1)p} \int_B |f(x)|^p\, d\mu(x),$$

which by §1.4 proves that $\omega \in A_p$.

To prove the converse, assume $\omega \in A_p$, and define M_ω by

$$M_\omega f(x) = \sup_{\delta > 0} \frac{1}{\omega(B(x,\delta))} \int_{B(x,\delta)} |f(y)|\, \omega(y)\, dy.$$

Since we know that ω is a doubling measure (see §1.5), all the metric properties required in Chapter 1, §1 are satisfied for the system of standard balls together with the measure $d\mu(x) = \omega(x)\, dx$. Conclusion (b) of the theorem in Chapter 1, §3.1 then shows that

$$\omega\{x : M_\omega f(x) > \alpha\} \leq \frac{A}{\alpha} \int |f(x)|\, \omega(x)\, dx, \quad \text{all } \alpha > 0. \tag{16}$$

However, by (4), if one takes the supremum over all balls B centered at x, one sees that $\omega \in A_p$ implies

$$(Mf(x))^p \leq c M_\omega(|f|^p)(x). \tag{17}$$

The weak-type inequality (15) is then a consequence of (16), if in (16) we replace f by $|f|^p$ and α by α^p/c.

3. The main theorem about A_p

Each of the two propositions above shows us that if a measure $d\mu$ leads to a weighted maximal inequality (2), then necessarily $d\mu(x) = \omega(x)\, dx$ and ω belongs to A_p. We now turn to the converse.

THEOREM 1. *Suppose $1 < p < \infty$ and $\omega \in A_p$. Then*

$$\int_{\mathbf{R}^n} (Mf(x))^p\, \omega(x)\, dx \leq A \int_{\mathbf{R}^n} |f(x)|^p\, \omega(x)\, dx, \tag{18}$$

for all $f \in L^p(\omega(x)\, dx)$.

Theorem 1 will be deduced from the following fundamental property of A_p weights.

PROPOSITION 3 *(Reverse Hölder inequality). If $\omega \in A_\infty$ then there exists an $r > 1$ and a $c > 0$ (both depending on ω) so that*

$$\left[\frac{1}{|B|} \int_B \omega^r\, dx\right]^{1/r} \leq \frac{c}{|B|} \int_B \omega\, dx, \tag{19}$$

for all balls B.

Observe that (except for the constant c) this is the reverse of Hölder's inequality, which holds automatically for all nonnegative functions. This proposition can be viewed as another quantitative statement of the fact that, on the average, the values of an A_p weight do not fluctuate too much (for instance, (19) is trivial for ω that are bounded above and below). In §5.1, we will see that the reverse Hölder inequality has a converse.

Proposition 3 has a surprising consequence.

COROLLARY. *Suppose that $\omega \in A_p$ for some $1 < p < \infty$. Then there is a $p_1 < p$ so that $\omega \in A_{p_1}$.*

To prove the corollary, let $1/p' = 1 - 1/p$, and recall (§1.2) that $\sigma = \omega^{-p'/p} \in A_{p'} \subset A_\infty$. Then σ satisfies a reverse Hölder inequality

$$\left(\frac{1}{|B|}\int_B \sigma^r\,dx\right)^{1/r} \le \frac{c}{|B|}\int_B \sigma\,dx$$

for some $r > 1$. Now $rp'/p = r/(p-1) = 1/(p_1 - 1) = p_1'/p_1$ for some $1 < p_1 < p$, and therefore $\omega \in A_{p_1}$ by the definition (3).

Theorem 1 is easily deduced from the corollary. Let $\omega \in A_p$ be given. Then $\omega \in A_{p_1}$ for some $p_1 < p$ and, by the proposition in §2.2, the maximal operator M is of weak type $(L^{p_1}(\omega\,dx), L^{p_1}(\omega\,dx))$. By the Marcinkiewicz interpolation theorem (together with the obvious boundedness of M on $L^\infty(\omega\,dx)$), M is then bounded on $L^p(\omega\,dx)$.

3.1 The proof of the reverse Hölder inequality is simplified by using the dyadic maximal function, which already appeared in §3 of the previous chapter. In order to exploit this tool, we note that since ω is a doubling measure, it suffices to prove the analogue of (19) with cubes Q replacing the balls B.

We will prove a slightly stronger statement than Proposition 3 and assume only that ω satisfies a weaker version of the A_∞ property, namely that there exists *one* pair of constants γ, δ with $0 < \gamma, \delta < 1$ so that, for all cubes Q and all subsets $E \subset Q$,

$$|E| \le \gamma|Q| \Rightarrow \omega(E) \le \delta\omega(Q). \tag{20}$$

Later (in §5.1) we shall see that this weaker version is actually equivalent to the full A_∞ property (as defined in §1.7).

PROPOSITION 4. *If ω has the weak A_∞ property (20), then there exists an $r > 1$ and a $c > 0$ so that*

$$\left(\frac{1}{|Q|}\int_Q \omega^r\,dx\right)^{1/r} \le \frac{c}{|Q|}\int_Q \omega\,dx \tag{21}$$

for all cubes Q.

Observe that the class of ω that satisfy (21) is invariant under multiplication by positive scalars, translation, and dilation, as is the class of ω that have the weak A_∞ property. Thus we may assume that $Q = Q_0$ is a unit dyadic cube with $\omega(Q_0) = |Q_0| = 1$, and we must show that

$$\int_{Q_0} \omega^r\,dx \le c. \tag{22}$$

To do this, we let $f = \omega\chi_{Q_0}$ and apply the dyadic maximal operator M^Δ defined in Chapter 4, §3.1.

Set $E^k = \{x \in Q_0 : M^\Delta f(x) > 2^{Nk}\}$, where N is a large integer to be chosen momentarily. The key observation is that

$$|E^k \cap Q| \le 2^{n-N}|Q| \tag{23}$$

for all cubes Q comprising E^{k-1}. To see this, recall the remarks in Chapter 4, §3.3 and let Q_j be one of the maximal dyadic cubes in E^k that is also contained in Q. By conclusion (i) of the lemma in Chapter 4, §3.1 (with $\alpha = 2^{Nk}$),

$$|Q_j| \le 2^{-Nk}\int_{Q_j} f\,dx.$$

Summing over all such $Q_j \subset Q$ gives

$$|E^k \cap Q| = \sum |Q_j| \le 2^{-Nk}\int_Q f\,dx,$$

since the Q_j are disjoint and their union is $E^k \cap Q$. By conclusion (ii) of the same lemma (with $\alpha = 2^{N(k-1)}$),

$$\int_Q f\,dx \le 2^n \cdot 2^{N(k-1)}.$$

Putting these facts together gives (23).

Now we choose N so that $2^{n-N} \le \gamma$, where γ is as in the assumed weak A_∞ property (20). It follows that

$$\omega(E^k \cap Q) \le \delta\omega(Q).$$

Taking the union over all Q comprising E^{k-1} gives $\omega(E^k) \le \delta\omega(E^{k-1})$, and therefore

$$\omega(E^k) \le \delta^k\omega(E^0) \le \delta^k,$$

since $\omega(E^0) \le \omega(Q_0) = 1$.

Now

$$\int_{Q_0} \omega^r\,dx \le \int_{Q_0} (M^\Delta f)^{r-1}\,\omega\,dx = \int_{Q_0\cap\{M^\Delta f \le 1\}} + \sum_{k=0}^{\infty}\int_{E^k\setminus E^{k+1}}.$$

The first integral is bounded by 1 and the k^{th} integral in the sum is bounded by

$$2^{N(k+1)(r-1)} \cdot \omega(E^k) \le 2^{N(k+1)(r-1)}\delta^k.$$

Since $\delta < 1$, the sum $\sum_{k=0}^{\infty} 2^{N(k+1)(r-1)}\delta^k$ converges if r is sufficiently close to 1. This proves (21) and, with it, Proposition 3.

4. Weighted inequalities for singular integrals

We saw in Chapter 1, §5 how estimates for singular integral operators could be obtained as consequences of a closely related analysis of maximal functions. In deducing weighted inequalities for these operators we shall follow the same procedure (in principle but not in detail) and reduce our estimates to those we have just obtained for maximal functions.

4.1 The class of singular integrals we shall deal with will be a subclass of the operators considered in §7 of Chapter 1. This particular subclass is chosen primarily for clarity of exposition.† We shall be concerned with operators $f \mapsto T(f)$, initially defined for $f \in C_0^\infty(\mathbf{R}^n)$, given as $Tf = f * K$, where K is an appropriate tempered distribution. We shall make the following assumptions about K. First, that the convolution operator it defines will be bounded on $L^2(\mathbf{R}^n)$; i.e., there is a constant A so that

$$\|Tf\|_2 \leq A\,\|f\|_2, \quad \text{for all } f \in C_0^\infty(\mathbf{R}^n). \tag{24}$$

The boundedness in (24) is of course equivalent with the fact that Fourier transform of K is an L^∞ function whose norm is at most A.‡

Next we suppose that, away from the origin, the distribution K agrees with a $C^1(\mathbf{R}^n \backslash \{0\})$ function. That is, there exists a function (which we call $K(x)$), defined for $x \neq 0$, that is C^1 there, so that the distribution K agrees with the function $K(x)$ when tested against C_0^∞ functions that vanish near the origin. About the function $K(x)$, we make the further assumption that it satisfies the inequalities

$$\left|\left(\frac{\partial}{\partial x}\right)^\alpha K(x)\right| \leq A\,|x|^{-n-\alpha}, \quad \text{for all } x \neq 0 \text{ and } |\alpha| \leq 1. \tag{25}$$

It is an interesting point that, in treating the operator $Tf = f * K$, it will be necessary to consider its truncated "approximations" T_ε, $\varepsilon > 0$, defined by

$$T_\varepsilon f(x) = (F * K_\varepsilon)(x) = \int_{\mathbf{R}^n} K_\varepsilon(x-y)\, f(y)\, dy,$$

where $K_\varepsilon(x) = K(x)\chi_{|x|\geq\varepsilon}$.

Moreover, associated with the family $\{T_\varepsilon\}$ is the maximal operator T_*, defined by $T_* f(x) = \sup_{\varepsilon>0} |T_\varepsilon f(x)|$ (for $f \in L^1$ the functions $T_\varepsilon f(x)$ are continuous and hence $T_* f$ is semicontinuous). The relation between T_* (and therefore the T_ε) and T we need is given by the following simple inequality: There exists a constant $c > 0$ so that

$$|Tf(x)| \leq T_* f(x) + c|f(x)|. \tag{26}$$

Observe that our operator T satisfies all the hypotheses of Chapter 1, §7. Here $q = 2$, $\rho(x,y) = |x-y|$, and (10), (22), and (27) of that chapter hold because

$$|K(x-y)| \leq \frac{A}{|x-y|^n}$$

and

$$|K(x-y) - K(x-\bar{y})| \leq \frac{A\,|y-\bar{y}|}{|x-\bar{y}|^{n+1}}, \quad \text{if } |y-\bar{y}| \leq \tfrac{1}{2}|x-\bar{y}|,$$

by our assumptions. So (26) is a consequence (it is (27) of Chapter 1).

4.2 After these preliminaries, we can state the main results about weighted inequalities for singular integrals.

THEOREM 2. *Suppose that the operator $Tf = f * K$ satisfies (24) and (25), and T_* is the associated maximal operator as above. Let $\omega \in A_\infty$, and suppose that $0 < p < \infty$. Then the inequality*

$$\int_{\mathbf{R}^n} [(T_* f)(x)]^p\, \omega(x)\, dx \leq A_{p,\omega} \int_{\mathbf{R}^n} [(Mf)(x)]^p\, \omega(x)\, dx \tag{27}$$

holds for every $f \in C_0^\infty(\mathbf{R}^n)$ for which the right-hand side is finite.†

Recalling that $A_p \subset A_\infty$ for $1 \leq p < \infty$ and using (26) gives the following consequence:

COROLLARY. *Suppose T is as above and $\omega \in A_p$, with $1 < p < \infty$. Then for $f \in C_0^\infty$,*

$$\int_{\mathbf{R}^n} |Tf(x)|^p\, \omega(x)\, dx \leq A \int_{\mathbf{R}^n} |f(x)|^p\, \omega(x)\, dx. \tag{28}$$

There are two further points that should be added to these results. The first is a clarification of how one passes from $f \in C_0^\infty(\mathbf{R}^n)$ to "general" f in the inequalities (27) and (28).‡ The second point to be made is a converse to the corollary, namely if (28) holds for appropriate T, then necessarily $\omega \in A_p$. These matters will be taken up in §4.6, after the proof of Theorem 2.

† More general results proved by the same arguments are sketched in §6.13 below.

‡ See, e.g., *Fourier Analysis*, Chapter 1.

† In stating inequalities of the form (27), (28), and others that follow, it will always be understood that the bounds A, $A_{p,\omega}$, etc., are independent of f.

‡ We mention that our choice here of C_0^∞ as the subspace on which to prove these inequalities is dictated by convenience and our desire to avoid deeper properties of ω at this stage. For an alternative formulation see §6.13 below.

4.3 The relative distributional inequality. The proof of Theorem 2 will be based on a relative distributional inequality that gives control of $T_*(f)$ in terms of $M(f)$. This inequality is of the kind already discussed in Chapter 4, §3.5. Here it takes the following form. For any weight $\omega \in A_\infty$, and for appropriate positive constants a, b, and c, we have that

$$\omega\{x : T_*f(x) > \alpha; Mf(x) \le c\alpha\} \le a\omega\{x : T_*f(x) > b\alpha\} \tag{29}$$

holds for all $\alpha > 0$. The lemma in Chapter 4, §3.5 then allows us to obtain the desired L^p inequality as a consequence.

4.4 Since the statement of inequality (29) is somewhat simpler when $\omega\, dx$ is Lebesgue measure (and since its proof easily implies the general case), we first consider that special situation.

In the rest of this section we shall hold in reserve the letters a, b, and c for their roles in the distribution inequality (29), and we shall use the letter A to denote a generic constant (chosen not to depend on a, b, or c).

PROPOSITION 5. *Suppose T_* is as described above. Then there exists an $A > 0$ so that the following holds: For every b with $0 < b < 1$, and $c > 0$,*

$$|\{x : T_*f(x) > \alpha;\ Mf(x) \le c\alpha\}| \le \frac{Ac}{1-b}\,|\{x : T_*f(x) > b\alpha\}|, \tag{30}$$

for all $\alpha > 0$.

PROPOSITION 6. *Let $\omega \in A_\infty$. Then there exists an a, $a = a_\omega$, with $a < 1$, so that the following holds: For any $b < 1$ we can find $c > 0$ so that*

$$\omega\{x : T_*f(x) > \alpha;\ Mf(x) \le c\alpha\} \le a\,\omega\{x : T_*f(x) > b\alpha\}, \tag{31}$$

for all $\alpha > 0$.†

We prove Proposition 5 first.

Since, for each $\varepsilon > 0$, the function $T_\varepsilon f(x)$ is continuous, the set O where $T_*f(x) = \sup_{\varepsilon>0} |T_\varepsilon f(x)| > b\alpha$ is open. Therefore we can decompose it as a disjoint union $O = \bigcup Q_j$ of "Whitney cubes".‡ Let $Q = Q_j$ be one of these cubes, and suppose we can show that

$$|\{x \in Q : T_*f(x) > \alpha;\ Mf(x) \le c\alpha\}| \le \frac{Ac}{1-b}\,|Q|, \tag{32}$$

with A independent of j. Summing over all the Q_j would then give us (30).

† While Theorem 3 is stated for $f \in C_0^\infty$, we note that M and T_* are well defined for $f \in L^1$. It is for this larger class that we prove (30) and (31).

‡ See Chapter 1, §3.2, or *Singular Integrals*, Chapter 6, §1.

The idea for proving (32) is to localize one of the usual estimates for singular integrals to an appropriate neighborhood of Q. The inequality that is needed here is the weak-type estimate

$$|\{x \in \mathbf{R}^n : T_*f(x) > \alpha\}| \le \frac{A}{\alpha}\int_{\mathbf{R}^n} |f(x)|\,dx, \tag{33}$$

which holds for every $f \in L^1$.

We recall that (33) is contained in Corollary 2 of Chapter 1, §7.3, and now turn to an examination of some simple geometric facts about the cube Q. We let d denote its diameter.

Since Q is one of the cubes of a Whitney decomposition of O, we can find a point $\bar{x}$ in the complement of O (i.e., an $\bar{x}$ with $T_*f(\bar{x}) \le b\alpha$) so that $\operatorname{dist}(\bar{x}, Q) \le 4d$. We let $B = B(\bar{x}, 6d)$ be the ball centered at $\bar{x}$ of radius $6d$. Then $Q \subset B$, and each point in Q is at least a distance d from ${}^cB$.

Next we may choose a point $\bar{y}$ in Q, so that $Mf(\bar{y}) \le c\alpha$; for if there were no such points then the set appearing on the left side of (32) would be empty, and the required inequality would hold automatically. Let B' be the smallest ball centered at $\bar{y}$ that contains B; we have that $|B'| \le A|Q|$ (see Figure 1).

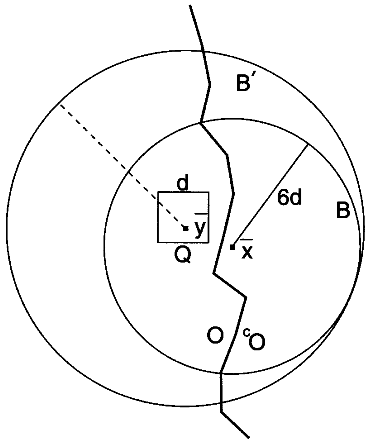

Figure 1. The proof of Proposition 5.

Next, we split $f = f_1 + f_2$ where $f_1 = \chi_B f$, $f_2 = \chi_{^cB} f$. Since $T_*f \le T_*f_1 + T_*f_2$, we have

$$\{T_*f > \alpha\} \subset \{T_*f_1 > b_1\alpha\} \cup \{T_*f_2 > b_2\alpha\},$$

whenever $b_1 + b_2 = 1$.

Using (33), we see that $|\{x \in Q : T_*f_1(x) > b_1\alpha\}| \le \dfrac{A}{b_1\alpha}\displaystyle\int |f_1|\,dx$, while

$$\int |f_1|\,dx = \int_B |f|\,dx \le \int_{B'} |f|\,dx \le c\alpha|B'| \le Ac\alpha|Q|,$$

because $Mf(\bar{y}) \le c\alpha$. Therefore

$$|\{x \in Q : T_*f_1(x) > b_1\alpha\}| \le A\,\frac{c}{b_1}\,|Q|. \tag{34}$$

We still have to deal with f_2, that part of f supported outside the ball B. We claim that the following estimate holds:

$$|T_\varepsilon f_2(\bar{x}) - T_\varepsilon f_2(x)| \le A\cdot Mf(\bar{y}), \quad \text{whenever } x \in Q. \tag{35}$$

To see this we begin by remarking that the left side of (35) is clearly majorized by

$$\int_{^cB} |K_\varepsilon(\bar{x} - y) - K_\varepsilon(x - y)|\cdot|f(y)|\,dy.$$

A straightforward calculation using (25) shows that

$$|K_\varepsilon(\bar{x} - y) - K_\varepsilon(x - y)| \le \frac{Ad}{|y - \bar{y}|^{n+1}}, \quad \text{whenever } x \in Q \text{ and } y \in {}^cB,$$

and moreover ${}^cB \subset \{y : |y - \bar{y}| \ge d\}$. Therefore, the left side of (35) can be dominated by

$$\begin{aligned} Ad\int_{|y-\bar{y}|\ge d} \frac{|f(y)|}{|y-\bar{y}|^{n+1}}\,dy &= Ad\int_{|y|\ge d} \frac{|f(\bar{y}-y)|}{|y|^{n+1}}\,dy \\ &= Ad\sum_{k=0}^{\infty}\int_{2^k d\le|y|<2^{k+1}d} \frac{|f(\bar{y}-y)|}{|y|^{n+1}}\,dy \\ &\le A'\left(\sum_{k=0}^{\infty} 2^{-k}\right) Mf(\bar{y}), \end{aligned}$$

which proves (35). Since the ball B is centered at $\bar{x}$, $T_\varepsilon f_2(\bar{x}) = T_{\varepsilon'}f(\bar{x})$, where $\varepsilon' = \max(\varepsilon, 6d)$ (recall that B has radius $6d$ and that $f_2 = f$ in ${}^cB$). We can now take the supremum over ε in (35) to get

$$T_*f_2(x) \le T_*f(\bar{x}) + A\cdot Mf(\bar{y}) \le b\alpha + Ac\alpha, \quad \text{for } x \in Q.$$

The key point is that we can make this last quantity less than $b_2\alpha$ by taking $b_2 \ge b + Ac$, and it has the effect of making the set $\{x \in Q : T_*f_2(x) > b_2\alpha\}$ empty.

We combine this with (34) to get

$$|\{x \in Q : T_*f(x) > \alpha; Mf(x) \le c\alpha\}| \le A\,\frac{c}{b_1}\,|Q|, \tag{36}$$

when $b_1 + b_2 = 1$ and if b_2 can be chosen so that $b_2 \ge b + Ac$. Now set $b_1 = (1-b)/2$, $b_2 = (1+b)/2$ (recall that $0 < b < 1$). If we have $(1+b)/2 > b + Ac$ (i.e., $Ac/(1-b) < 1/2$), then (36) gives us the desired conclusion. If however $Ac/(1-b) \ge 1/2$, then we automatically get (32) by choosing A twice as large. This concludes the proof of Proposition 5.

4.5 Proof of Proposition 6. We have only to reexamine the argument we just gave to deduce from it also Proposition 6. Recall that if ω belongs to A_∞ and if we fix a $\gamma < 1$, then there is a $\delta < 1$ so that, for all cubes Q and all subsets $E \subset Q$,

$$|E| \le \gamma|Q| \Rightarrow \omega(E) \le \delta\omega(Q).$$

Looking back at (32), we see that (with b given, $0 < b < 1$) we need merely choose c small enough so that $Ac/(1-b) \le \gamma$. Consequently

$$\omega\{x \in Q : T_*f(x) > \alpha;\ Mf(x) \le c\alpha\} \le \delta\omega(Q). \tag{37}$$

Summing again over all $Q = Q_j$ then gives the conclusion (31) with $a = \delta$.

To prove Theorem 2 we may use the lemma cited in §4.3, since if $a < 1$ we can always choose $b < 1$ so that $a < b^p$. The theorem is immediate once we verify that

$$\int (Mf(x))^p\,\omega(x)\,dx < \infty \Rightarrow \int (T_*f(x))^p\,\omega(x)\,dx < \infty, \tag{38}$$

whenever $f \in C_0^\infty$.

4.5.1 Proof of (38). If we assume that $\int_{\mathbf{R}^n}(Mf)^p\,\omega\,dx < \infty$, even for one $f \in C_0^\infty$, then necessarily

$$\int_{\mathbf{R}^n} \frac{\omega(x)}{(1+|x|)^{np}}\,dx < \infty, \tag{$*$}$$

because whenever $f \not\equiv 0$ then $Mf(x) \ge C(1+|x|)^{-n}$. However, and this is the main point, $|T_*f(x)| \le A(1+|x|)^{-n}$, whenever $f \in C_0^\infty$.†

† The smoothness assumption on f can be dispensed with if we use more refined properties of the weight ω; see §6.13 below.

Indeed, assume that f is supported in $|x| \le R$. Then $|T_\varepsilon f(x)| \le A|x|^{-n}$, for $|x| \ge 2R$ (with A independent of ε). Since f is smooth, and the convolution of a smooth function with a tempered distribution is smooth, $Tf(x)$ is bounded when $|x| \le 2R$. By (30) in Chapter 1, §6.3, $T_* f$ is also bounded when $|x| \le 2R$, proving the claim. Thus we have shown that if $f \in C_0^\infty$ and $\|Mf\|_{L^p(\omega\,dx)}$ is finite, then $\|T_* f\|_{L^p(\omega\,dx)}$ is finite, which is (38).

4.5.2 Passage from C_0^∞ to L^p. As far as the corollary (28) is concerned, the passage to general f is straightforward: Every element of $L^p(\omega\,dx)$ can be approximated (in the norm) by C_0^∞ functions. This leads to a natural extension of T to $L^p(\omega\,dx)$ that satisfies (28). However, the corresponding situation for the theorem (27) is a little trickier. For $0 < p < \infty$, let us write $ML^p(\omega)$ for the space of locally integrable functions f for which

$$\|f\|^p \equiv \int_{\mathbf{R}^n} (Mf(x))^p\,\omega(x)\,dx$$

is finite. When $1 \le p$, we take $\|\cdot\|$ to be a norm on $ML^p(\omega)$, and when $p < 1$, $\|\cdot\|^p$ allows us to define a metric on $ML^p(\omega)$.

We may assume that $ML^p(\omega)$ is not empty, i.e., that (*) holds. The theorem (for general f) follows from the fact that $C_0^\infty(\mathbf{R}^n)$ is dense in $ML^p(\omega)$ whenever $\omega \in A_\infty$. A proof of this fact is sketched in §6.14 below.

4.6 The converse. We prove a converse of the corollary in §4.2: If an inequality of the form (28) holds for a measure $d\mu$, then $d\mu = \omega(x)\,dx$, with $\omega \in A_p$.

We consider an operator T, with $Tf = f * K$, that satisfies (24) and (25) and an additional nondegeneracy condition that we formulate as follows: There exists a constant $a > 0$, and a unit vector u_0, so that

$$|K(x)| \ge a|x|^{-n}, \quad \text{whenever } x = t \cdot u_0 \text{ with } -\infty < t < \infty. \tag{39}$$

Note that this condition holds whenever T is, e.g., any *one* of the Riesz transforms.

PROPOSITION 7. *Suppose T satisfies the conditions above. Let $d\mu$ be a nonnegative Borel measure, and let $1 < p < \infty$. Assume that*

$$\int_{\mathbf{R}^n} |Tf|^p\,d\mu \le A \int_{\mathbf{R}^n} |f|^p\,d\mu \tag{40}$$

holds for all $f \in C_0^\infty(\mathbf{R}^n)$. Then $d\mu$ is absolutely continuous $d\mu(x) = \omega(x)\,dx$, with $\omega \in A_p$.

The proof is based on the following observation. By choosing $u = tu_0$, with t fixed but sufficiently large, we can guarantee that

$$|K(r(u+v)) - K(ru)| \le \frac{1}{2}|K(ru)|, \tag{41}$$

whenever $r \in \mathbf{R}\backslash\{0\}$ and $|v| \le 2$.

To see this, let $x \ne 0$ lie on the line $\{tu_0\}$, and take $y \in \mathbf{R}^n$ so that $|y| \le c|x|$. First, if $c \le 1/2$, then $|K(x+y) - K(x)| \le cA'|x|^{-n}$, because of the condition (25) assumed about the kernel K. Now (41) is merely a restatement of this if we write $x = ru$, $y = rv$; the requirement $|y| \le c|x|$ becomes $|v| \le c|u| = c|t|$, and so (41) can always be satisfied if $t \ge 4A'/a\,(\ge |v|/c)$.

Next, for each ball $B = B(\bar{x}, r)$, we shall assign a translate B' in the direction of u_0, $B' = B(\bar{x} + ru, r)$. Similarly, we define a translate B'' in the opposite direction by $B'' = B(\bar{x} - ru, r)$.

Now let f be any nonnegative C_0^∞ function supported in B. Let us consider $Tf(x)$ for $x \in B'$. We have $Tf(x) = \int K(x-y)f(y)\,dy$, with $x = \bar{x} + ru + rx'$, and since $y \in B$, we have $y = \bar{x} + ry'$; here $|x'| \le 1$ and $|y'| \le 1$. Thus $x - y = r(u+v)$ with $|v| \le 2$, and (41) gives us

$$|Tf(x)| \ge \frac{1}{2} f_B |K(ru)|, \quad \text{for all } x \in B'.$$

Our assumption (40) then implies that

$$\mu(B')(f_B)^p \le A \int_B f^p\,d\mu. \tag{42}$$

By a simple passage to the limit, this inequality extends to any nonnegative function f supported in B. Similarly, by reversing the roles of B and B' (using the translate $(B')'' = B$), one also has

$$\mu(B)(f_{B'})^p \le A \int_{B'} f^p\,d\mu. \tag{43}$$

However (43), with $f = \chi_{B'}$, shows that $\mu(B) \le A\mu(B')$. Substituting this in (42) finally yields

$$\mu(B)(f_B)^p \le A \int_B f^p\,d\mu.$$

We need only invoke the corollary in §2.1 to complete our proof.

5. *Further properties of A_p weights*

We now intend to look more closely at the class of A_p functions. It will turn out that a particularly fruitful question is that of characterizing those functions that belong to some A_p, without specifying p ahead of time. This is the role of the class A_∞.

5.1 The class A_∞.

THEOREM 3. *A function $\omega \geq 0$ belongs to some A_p, $1 \leq p < \infty$, if and only if ω satisfies either of the following equivalent properties:*

(i) *ω belongs to A_∞.*

(ii) *ω satisfies a reverse Hölder inequality* (19).

We have already seen that if $\omega \in A_p$, then $\omega \in A_\infty$, and as a result it satisfies a reverse Hölder inequality (Proposition 3). Therefore what remains to be done is to show that whenever ω satisfies a reverse Hölder inequality, then it belongs to some A_p.

In order to do this, we shall have to reverse the roles of the measures $\omega(x)\,dx$ and dx, standing their previous relation on its head. To explain this, let us for a moment consider the more general situation of two non-negative measures $d\mu_1$ and $d\mu_2$ that are mutually absolutely continuous. We say that μ_1 satisfies a reverse Hölder inequality with respect to μ_2 if, for some $r > 1$ and $c > 0$,

$$\left[\frac{1}{\mu_2(B)}\int_B \left(\frac{d\mu_1}{d\mu_2}\right)^r d\mu_2\right]^{1/r} \leq c\frac{\mu_1(B)}{\mu_2(B)} \tag{44}$$

for all balls B. Then this is the inequality (19) if $d\mu_1(x) = \omega(x)\,dx$ and $d\mu_2(x) = dx$. However, if $d\mu_1(x) = dx$ and $d\mu_2(x) = \omega(x)\,dx$ (so $d\mu_1/d\mu_2 = \omega^{-1}$), then (44) becomes[‡]

$$\left[\frac{1}{\omega(B)}\int_B \omega^{1-r}\,dx\right]^{1/r} \leq c\frac{|B|}{\omega(B)}. \tag{45}$$

Now (45) is merely a rewriting of the definition (3) of A_p, with $r - 1 = p'/p$, i.e., $r = p'$. Thus it suffices to prove (45) for some $r > 1$.

We keep with the choice $d\mu_1 = dx$, $d\mu_2 = \omega\,dx$. Since ω is a doubling measure, we prove (44) with cubes Q replacing the balls B. Now the class of ω that satisfy (19), with given fixed r and c, is invariant under dilations, translations, and multiplication by positive scalars, as is the class of ω that satisfy (45). Thus we may normalize the situation, taking

‡ Recall the notation $\omega(B) = \int_B \omega(x)\,dx$.

Q_0 to be a unit dyadic cube, requiring that $\omega(Q_0) = |Q_0| = 1$, and set ourselves the task of proving

$$\int_{Q_0} \omega^{1-\bar{r}} \leq \bar{c} \tag{46}$$

where $\bar{r} > 1$, $\bar{c} > 0$, and both depend only on r and c.

It follows immediately from the assumed reverse Hölder property that there exist constants $0 < \gamma, \delta < 1$ so that, for all cubes Q and all subsets $E \subset Q$

$$\omega(E) \leq \gamma\omega(Q) \Rightarrow |E| \leq \delta|Q|. \tag{47}$$

We shall now follow closely the argument in §3.1 (the proof of the reverse Hölder property), but now using the dyadic maximal function with weight ω, M_ω^Δ, defined by

$$M_\omega^\Delta f(x) = \sup_{x\in Q} \omega(Q)^{-1}\int_Q |f(y)|\,\omega(y)\,dy,$$

where the supremum is taken over all dyadic cubes containing x. As we have already remarked in §3.2 of Chapter 4, this is again a martingale maximal function, so conclusions (i), (iii), and (iv) of the lemma in Chapter 4, §3.1 hold for it as before. In addition, a version of (ii) holds (with only a change of constant), since (19) guarantees that ω is a doubling measure. Let us summarize these facts in an inequality. There exists an integer m, so that for any $\alpha > 0$, the set $\{x : M_\omega^\Delta f(x) > \alpha\}$ can be written as a disjoint union of dyadic cubes $\bigcup Q_j$, with

$$\alpha < \frac{1}{\omega(Q_j)}\int_{Q_j} |f(x)|\,\omega(x)\,dx \leq 2^m\alpha. \tag{48}$$

(The integer m is chosen so that $\omega(\tilde{Q}) \leq 2^m\omega(Q)$, where $\tilde{Q}$ is the dyadic "parent" of Q; the existence of such an m follows from the doubling property of ω).

Let us now consider $M_\omega^\Delta(f)$, with $f = \omega^{-1}\chi_{Q_0}$; note that $f \in L^1(\omega\,dx)$. Then, in parallel with the argument in §3.1, we set

$$E^k = \{x \in Q_0 : M_\omega^\Delta f(x) > 2^{Nk}\},$$

and, as before (using relations like (48)), we can conclude that

$$\omega(E^k \cap Q) \leq 2^{m-N}\omega(E^{k-1} \cap Q)$$

for each cube Q making up E^{k-1}. We now take γ as in (47) and fix N so large that $2^{m-N} \leq \gamma$. Summing over the cubes comprising E^{k-1} then gives

$$|E^k| \leq \delta|E^{k-1}|,$$

and therefore $|E^k| \le \delta^k$.

Now

$$\int_{Q_0} \omega^{1-\bar{r}} = \int_{Q_0} f^{\bar{r}}\, \omega\, dx \le \int_{Q_0} (M_\omega^\triangle f)^{\bar{r}-1} f\, \omega\, dx$$
$$= \int_{Q_0 \cap \{M_\omega^\triangle f \le 1\}} + \sum_{k=0}^{\infty} \int_{Q_0 \cap E^k \setminus E^{k+1}} .$$

The first integral is majorized by $|Q_0| = 1$, while the k^{th} integral in the sum is majorized by

$$2^{N(k+1)(\bar{r}-1)} |E^k| \le 2^{N(k+1)(\bar{r}-1)} \delta^k,$$

which gives a convergent series if $\bar{r}$ is sufficiently close to 1, because $\delta < 1$. We have therefore proved (46) and, as we have already indicated, this proves (45) with $r = \bar{r}$ and shows that $\omega \in A_{\bar{r}'}$, completing the proof of Theorem 3.

5.2 Another characterization of A_1. We conclude our treatment of the A_p classes with two propositions. These give further insight into the nature of the limiting class A_1 and its relation with the other A_p classes.

First, we present a simple construction of the "general" element of A_1.

We start with a nonnegative function f with the property that $Mf(x) < \infty$ for almost every x.[†] We then fix an exponent q with $0 < q < 1$. Our claim is then that $\omega(x) = (Mf(x))^q$ belongs to A_1, i.e.,

$$M\omega(x) \le c\omega(x), \tag{49}$$

with the bound c depending only on q (and not on f!).

We can, after suitable dilations, translations, and multiplications, reduce (49) to the special case

$$\int_B \omega(x)\, dx \le c, \tag{50}$$

where $\omega(0) = (Mf)^q(0) = 1$ and B is the unit ball centered at the origin. To show (50), let B_2 be the ball of radius 2 centered at the origin and write $f = f_1 + f_2$, where $f_1 = \chi_{B_2} f$ and $f_2 = \chi_{{}^cB_2} f$. Let λ_1 denote the distribution function of Mf_1 on B, i.e.,

$$\lambda_1(\alpha) = |\{x \in B : Mf_1(x) > \alpha\}|.$$

[†] It is not difficult to verify that this condition is equivalent with the requirements that f be locally integrable and that $\int_{|x| \le R} f(x)\, dx$ be majorized by a constant multiple of R^n as $R \to \infty$.

Then

$$\int_B (Mf_1)^q\, dx = \int_0^\infty q\lambda_1(\alpha)\alpha^{q-1}\, d\alpha = \int_0^1 + \int_1^\infty .$$

For the first integral, use the fact that $\lambda_1(\alpha) \le |B| = c$; hence this integral is bounded by a constant. For the second integral, we call on the fundamental weak-type inequality for the maximal function (as in Chapter 1, §3). Thus

$$\lambda_1(\alpha) \le \frac{c}{\alpha} \int f_1 = \frac{c}{\alpha} \int_{B_2} f \le \frac{2c}{\alpha},$$

since $\int_{B_2} f \le |B_2| \cdot Mf(0) \le c \cdot 1$. This results in the convergence of the second integral, and we have that

$$\int_B (Mf_1)^q\, dx \le c.$$

Since f_2 vanishes in $B_2 = B(0,2)$, we see that whenever $x \in B = B(0,1)$ and B' is any ball centered at x

$$\frac{1}{|B'|} \int_{B'} f_2\, dx \le \frac{c}{|B''|} \int_{B''} f\, dx,$$

where B'' is the ball centered at the origin whose radius is 1 greater than that of B'. This shows that

$$Mf_2(x) \le c \cdot Mf(0), \quad \text{for all } x \in B,$$

and hence that

$$\int_B (Mf_2)^q\, dx \le c.$$

Of course, $Mf \le Mf_1 + Mf_2$; combining our estimates gives that $\int_B (Mf)^q\, dx \le c$. Since $\omega(x) = (Mf(x))^q$, we have shown (50) and with it (49), thereby proving our claim. A converse to this statement also holds, and we formulate both in the following proposition.

PROPOSITION 8. *Let f be a nonnegative function with $Mf(x) < \infty$ almost everywhere. Then, for $0 < q < 1$, the function $(Mf)^q$ belongs to A_1. Conversely, given an $\omega \in A_1$, there exists a nonnegative f and a q, $0 < q < 1$, so that the quotient $\omega/(Mf)^q$ is bounded above and below by positive constants.*

To prove the converse, suppose $\omega \in A_1$. Then, as we have seen in §3, it satisfies a reverse Hölder inequality: there is a fixed $r > 1$ so that

$$\left[\frac{1}{|B|} \int_B \omega^r\, dx\right]^{1/r} \le \frac{c}{|B|} \int_B \omega\, dx$$

for all balls B. The A_1 property implies that the right side of this inequality is bounded by $c\omega(\bar{x})$, where $\bar{x}$ is the center of B. Taking the supremum over all B centered at $\bar{x}$ then gives

$$(M(\omega^r))^{1/r} \le c\omega. \tag{51}$$

Now set $f = \omega^r$ and $q = 1/r$ ($q < 1$ since $r > 1$). Then $(Mf)^q = (M(\omega^r))^{1/r}$, and (51) shows that $(Mf)^q \le c\omega$. Now $(M(\omega^r))^{1/r} \ge \omega$ automatically; thus $(Mf)^q \ge \omega$, and the proposition is proved.

5.3 Factorization of A_p weights. We show how to construct a general weight in A_p in terms of elements of A_1.

PROPOSITION 9. *Suppose that ω_1 and ω_2 are A_1 weights. If $1 \le p < \infty$, then $\omega = \omega_1\omega_2^{1-p}$ belongs to A_p. Conversely, suppose that ω is an A_p weight. Then there exist ω_1 and ω_2 in A_1 so that $\omega = \omega_1\omega_2^{1-p}$.*

Proof. Let us consider first the case $p = 2$, where the play of exponents is most transparent. If $\omega = \omega_1/\omega_2$, then

$$\int_B \omega\,dx \le \int_B \omega_1\,dx \cdot [\inf_{x\in B} \omega_2(x)]^{-1}.$$

Similarly

$$\int_B \omega^{-1}\,dx \le \int_B \omega_2\,dx \cdot [\inf_{x\in B} \omega_1(x)]^{-1},$$

and hence

$$\left(\frac{1}{|B|}\int \omega\,dx\right)\cdot\left(\frac{1}{|B|}\int \omega^{-1}\,dx\right) \le \prod_{i=1}^{2}\left[\frac{1}{|B|}\int_B \omega_i\,dx \cdot [\inf_{x\in B}\omega_i(x)]^{-1}\right],$$

and each term in the product is bounded, since $\omega \in A_1$.

A less mechanical argument, which also motivates the converse, can be sketched as follows. Consider the family of convolution operators T_ε that are the subject of Proposition 1 in §2.1. The fact that $\omega_i \in A_1$ is equivalent to the uniform boundedness (as mappings from $L^1(dx)$ to itself) of the operators $\omega_1 T_\varepsilon \omega_1^{-1}$ and $\omega_2 T_\varepsilon \omega_2^{-1}$ (as ε varies). By duality, $\omega_i^{-1} T_\varepsilon \omega_i$ are uniformly bounded from $L^\infty(dx)$ to itself. Then by convexity properties of operators,‡ one sees that $\omega_1^{1/2}\omega_2^{-1/2} T_\varepsilon \omega_1^{-1/2}\omega_2^{1/2}$ are uniformly bounded from $L^2(dx)$ to itself, which by §2.1 implies that ω_1/ω_2 belongs to A_2.

‡ See, e.g., *Fourier Analysis*, Chapter 5, §4.

To prove the converse, we pass from T_ε to the standard maximal operator M, and consider T defined by

$$Tf = \omega^{-1/2}M(\omega^{1/2}f) + \omega^{1/2}M(\omega^{-1/2}f).$$

While the operator T is not linear, it is subadditive, i.e., $T(f_1 + f_2) \le Tf_1 + Tf_2$. If ω belongs to A_2 then so does ω^{-1} and we can assert that both $\omega^{-1/2}M\omega^{1/2}$ and $\omega^{1/2}M\omega^{-1/2}$ are bounded from $L^2(dx)$ to itself. Thus

$$\|Tf\|_2 \le A\|f\|_2,$$

for some $A > 0$. Fix now a nonnegative f with $\|f\|_2 = 1$ and write

$$\eta = \sum_{k=1}^{\infty}(2A)^{-k}T^k(f), \tag{52}$$

where $T^k(f) = T(T^{k-1}(f))$. Then $\|\eta\|_2 \le \sum_{k=1}^{\infty} 2^{-k} = 1$, and $\eta \in L^2$. Furthermore, since T is positivity-preserving and subadditive, we have the pointwise inequality

$$T\eta \le \sum_{k=1}^{\infty}(2A)^{-k}T^{k+1}(f) = \sum_{k=2}^{\infty}(2A)^{1-k}T^k(f) \le (2A)\eta.$$

Thus, if $\omega_1 = \omega^{1/2}\eta$, then

$$M(\omega_1) \le T(\eta)\omega^{1/2} \le 2A\eta\omega^{1/2} = 2A\omega_1$$

and $\omega_1 \in A_1$. Similarly, if $\omega_2 = \omega^{-1/2}\eta$, then $M(\omega_2) \le 2A\omega_2$, so $\omega_2 \in A_1$. Now $\omega = \omega_1/\omega_2$, so the converse is proved for $p = 2$.

The modifications needed to deal with general p are as follows. The proof of the first part of the proposition is similar to what was already shown for $p = 2$, and is in any case a straightforward exercise. For the converse direction, consider first $p \ge 2$, let $\omega \in A_p$, and define T by

$$Tf = [\omega^{-1/p}M(f^{p/p'}\omega^{1/p})]^{p'/p} + \omega^{1/p}M(f\omega^{-1/p}).$$

Because $\omega^{-p'/p} \in A_{p'}$, Theorem 2 in §3 shows that T is bounded on L^p. Also, since $p \ge 2$, $p/p' \ge 1$, and Minkowski's inequality gives $T(f_1+f_2) \le Tf_1 + Tf_2$. We can now form η as in (52), writing $\omega_1 = \omega^{1/p}\eta^{p/p'}$ and $\omega_2 = \omega^{-1/p}\eta$. The fact that $T\eta \le 2A\eta$ then shows that both ω_1 and ω_2 belong to A_1. Moreover,

$$\omega = \omega_1\omega_2^{1-p} = \omega^{1/p}\eta^{p/p'}\cdot(\omega^{-1/p}\eta)^{1-p},$$

since $p/p' = p - 1$, finishing the proof for $p \ge 2$.

The case $p \le 2$ then follows immediately by writing down the corresponding factorization for $\omega^{-p'/p}$ (which belongs to $A_{p'}$) and raising the product to the $(-p/p')^{\text{th}}$ power.

6. Further results

A. The class A_p and reverse Hölder inequalities

6.1 (a) Suppose $\omega_1 \in A_{p_1}$, $\omega_2 \in A_{p_2}$, with $1 \leq p_1, p_2 < \infty$. Then for all $\theta \in [0,1]$, if we take

$$\frac{1}{p} = \frac{1-\theta}{p_1} + \frac{\theta}{p_2}, \qquad \omega^{1/p} = \omega_1^{(1-\theta)/p_1} \cdot \omega_2^{\theta/p_2},$$

we have $\omega \in A_p$.

(b) As a result, if $\omega \in A_p$ and $0 \leq \theta = \dfrac{q-1}{p-1} \leq 1$, then $\omega^\theta \in A_q$; in particular, $\omega^\theta \in A_p$.

(c) If $\omega \in A_p$ then there is an $\varepsilon > 0$ for which $\omega^{1+\varepsilon} \in A_p$.

Statements (a) and (b) follow directly from Hölder's inequality. The assertion (c) is deeper; it is a consequence of the fact that ω and $\omega^{-p'/p}$, being in A_∞, satisfy a reverse Hölder inequality.

6.2 (a) As is seen in §1.8, if $\omega \in A_p$ then $\log \omega \in \mathrm{BMO}$. Conversely, if $f \in \mathrm{BMO}$ is real-valued, and $p > 1$ is fixed, then $f = c \log \omega$ for some $\omega \in A_p$.

(b) An essentially equivalent assertion is that $f \in \mathrm{BMO}$ if and only if $f = c \log \omega$ for some ω that satisfies a reverse Hölder inequality.

To prove that every BMO function f can be represented as in (a), one uses the inequality $|B|^{-1} \int_B e^{\mu|f-f_B|}\,dx \leq c$, which occurs in §1.3.1 of the previous chapter.

6.3 A weight ω belongs to A_∞ if and only if

$$\frac{1}{|B|}\int_B \omega\,dx \cdot \exp\left(\frac{1}{|B|}\int_B \log(1/\omega)\,dx\right) \leq A, \qquad (*)$$

for all balls B. Note that since $p'/p = 1/(p-1)$, this inequality is exactly the limit, as $p \to \infty$, of the inequality (3) that defines A_p.

Indeed, if $\omega \in A_\infty$, then $\omega \in A_p$ for some $p < \infty$, and $(*)$ is a consequence of Jensen's inequality. To prove the converse, one shows that $(*)$ implies the existence of constants $c_1, c_2 > 0$ so that $e^{-c_1|B|/|F|} \leq c_2\omega(F)/\omega(B)$ whenever B is a ball and $F \subset B$; see Hruščev [1984].

6.4 The function $\omega(x) = |x|^a$ belongs to A_p, $p > 1$, if and only if $-n < a < n(p-1)$. Notice however that ω is a doubling measure in the larger range $a > -n$. More subtle examples of doubling measures that are not A_p weights are given in Chapter 1, §8.8: One can find a doubling measure that is totally singular (with respect to Lebesgue measure) and, alternatively, an absolutely continuous doubling measure whose density vanishes on a set of positive measure. See also C. Fefferman and Muckenhoupt [1974], Strömberg [1979b].

6.5 Let P be a polynomial on $\mathbf{R}^n$ having degree d.

(a) The function $|P|$ satisfies a reverse Hölder inequality. Hence, $\log|P| \in \mathrm{BMO}$.

(b) More precisely, we have that $|P|^a \in A_p$, $p > 1$, whenever $-1 < ad < p-1$. Moreover, if n, d, and a are fixed, then the A_p bound of $|P|^a$ remains bounded as P varies over the polynomials of degree d on $\mathbf{R}^n$.

The result follows from the inequality

$$\int_B |P(x)|^{-\mu}\,dx \leq c_{\mu,d}\left(\int_B |P(x)|\,dx\right)^{-\mu},$$

where $\mu d < 1$ and B is the unit ball, which is proved in Ricci and Stein [1987].

6.6 (a) Suppose ω satisfies the reverse Hölder inequality

$$\left(\frac{1}{|B|}\int_B \omega^r\,dx\right)^{1/r} \leq \frac{c}{|B|}\int_B \omega\,dx$$

for all balls B, and some fixed $r > 1$. Then ω automatically satisfies the same inequality for some larger r.

(b) Suppose that $M(\omega) \leq c\omega$ (i.e., $\omega \in A_1$). Then there is an $r > 1$ so that $[M(\omega^r)]^{1/r} \leq c'\omega$, i.e., so that $\omega^r \in A_1$.

To prove (a), one notes that our assumption implies that $\omega^r \in A_\infty$, and hence that ω^r itself satisfies a reverse Hölder inequality. Similarly, if ω is as in (b), then $\omega \in A_\infty$ and ω satisfies a reverse Hölder inequality. For the original proof of (a), see Gehring [1973].

6.7 Suppose $d\mu_1$ and $d\mu_2$ are finite nonnegative measures on $\mathbf{R}^n$. Then $d\mu_1$ is absolutely continuous with respect to $d\mu_2$ if, for any $\delta > 0$, there is a $\gamma > 0$ so that $\mu_1(E) \leq \delta$ whenever $\mu_2(E) \leq \gamma$. We now consider a variant of this relation between $d\mu_1$ and $d\mu_2$, which will hold uniformly at every scale. That is, we suppose that for every $\delta > 0$, there is a $\gamma > 0$ so that $\mu_1(E)/\mu_1(B) \leq \delta$ if $\mu_2(E)/\mu_2(B) \leq \gamma$, whenever $B \subset \mathbf{R}^n$ is a ball and $E \subset B$. When this condition is satisfied, we write $d\mu_1 \preceq d\mu_2$.

The significant fact is that $\preceq$ is an equivalence relation; in particular, $d\mu_1 \preceq d\mu_2$ implies that $d\mu_2 \preceq d\mu_1$.

In this connection we make three remarks. First, if $d\mu_2 = dx$ and $d\mu_1 = \omega(x)\,dx$, then the condition $d\mu_1 \preceq d\mu_2$ is equivalent to the condition $\omega \in A_\infty$ (see §3.1). Second, the fact that $\preceq$ is an equivalence relation is implicit in the argument of §3.1 and its "inverse" in §5.1. Third, we have $d\mu_1 \preceq d\mu_2$ exactly when $d\mu_1$ satisfies a reverse Hölder inequality with respect to $d\mu_2$ (in the sense of §5.1). Further details are in Coifman and C. Fefferman [1974].

6.8 Suppose F is a homeomorphism from $\mathbf{R}^1$ onto itself, which we assume to be orientation-preserving (i.e., increasing). Then $f \mapsto f \circ F$ maps $\mathrm{BMO}(\mathbf{R}^1)$ onto itself exactly when $F' \in A_\infty$. See P. Jones [1983]. The analogue for $n > 1$ requires F to be quasi-conformal; see §6.9 of the previous chapter.

6.9 Let $n > 1$ and suppose that $F : \mathbf{R}^n \to \mathbf{R}^n$ is a quasi-conformal homeomorphism. Then $|\det \nabla F|$ satisfies a reverse Hölder inequality and, in particular, $\log|\det \nabla F| \in \mathrm{BMO}(\mathbf{R}^n)$. See Gehring [1973], Reimann [1974].

6.10 One can consider a weakened version of the reverse Hölder inequality of §6.6(a), namely, suppose there is a fixed $r > 1$ and a weight ω so that

$$\left(\frac{1}{|B|}\int_B \omega^r \, dx\right)^{1/r} \leq \frac{c}{|B^*|}\int_{B^*} \omega \, dx$$

for all balls $B \subset \mathbf{R}^n$; here B^* is the double of B. Then ω satisfies a similar weak reverse Hölder inequality for some $r' > r$.

Inequalities of the kind play a role in nonlinear partial differential equations related to the calculus of variations. In particular, if u is the solution of a second-order equation of this type on $\mathbf{R}^n$, then $\omega = |\nabla u|^{2n/(n+2)}$ satisfies an inequality akin to the above, with $r = (n+2)/n$. For an exposition of these topics, see the monograph of Giaquinta [1983].

6.11 We consider the following generalization of questions treated in §2 and §3 above: For appropriate operators T, characterize the pairs of nonnegative measures $d\mu_1$ and $d\mu_2$ so that either $T : L^p(d\mu_2) \to L^p(d\mu_1)$ is bounded, or is of weak-type (p,p).

(a) Suppose p is fixed, $1 < p < \infty$. Then†

$$\sup_{\Phi\in\mathcal{R}} \int |T_\Phi f|^p \, d\mu_1 \leq c \int |f|^p \, d\mu_2, \qquad T_\Phi f = f * \Phi,$$

exactly when $d\mu_1 = \omega(x)\,dx$ is absolutely continuous, $d\mu_2 = \omega_2(x)\,dx + d\nu$ (here $d\nu$ is totally singular), and

$$\frac{1}{|B|}\int_B \omega_1 \, dx \cdot \left(\frac{1}{|B|}\int_B \omega_2(x)^{-p'/p} \, dx\right)^{p/p'} \leq A,$$

for all balls B. A similar limiting result holds for $p = 1$.

(b) The same conditions are equivalent to the weak-type inequality for the maximal operator

$$\mu_1\{x : Mf(x) > \alpha\} \leq \frac{c}{\alpha^p}\int_{\mathbf{R}^n} |f|^p \, d\mu_2.$$

The proofs of these assertions are parallel to those of the propositions in §2. In particular, one shows that these conditions are equivalent to the property that $(f_B)^p \mu_1(B) \leq c' \int_B f^p \, d\mu_2$ for all nonnegative functions f. For (b), see Muckenhoupt [1972].

† See §2.1 for the definition of $\mathcal{R}$.

6.12 However, when $1 < p < \infty$, the conditions above are not sufficient to guarantee the strong inequality $\|Mf\|_{L^p(d\mu_1)} \leq c_p \|f\|_{L^p(d\mu_2)}$. A counterexample for $p = 2$, $n = 1$ may be constructed by taking (near the origin):

$$f(x) = |x|^{-1}\log(1/|x|)^{-2}, \quad \omega_1(x) = |x|\log(1/|x|), \quad \omega_2(x) = |x|\log(1/|x|)^2.$$

An actual necessary and sufficient condition is that

$$\int_B [M(\chi_B \omega_2^{1-p'})]^p \, \omega_1 \, dx \leq c \int_B \omega_1^{1-p'} \, dx$$

for all balls B. This result is due to E. Sawyer [1982]. When $\omega_1 = \omega_2$, the above is of course equivalent to the A_p condition; for a direct proof of this, see Hunt, Kurtz, and Neugebauer [1983].

B. Singular integrals and maximal operators

6.13 The result given in Theorem 2 extends to non-translation-invariant operators as follows: Let T be a bounded operator from $L^2(\mathbf{R}^n)$ to itself that is associated to a kernel K in the sense that

$$(Tf)(x) = \int_{\mathbf{R}^n} K(x,y)\, f(y)\, dy$$

for all compactly supported functions $f \in L^2$ and all x outside the support of f. Suppose that, for some $\gamma > 0$ and some $A > 0$, K satisfies the inequalities

$$|K(x,y)| \leq A|x-y|^{-n},$$

and

$$|K(x,y) - K(x',y)| \leq A\frac{|x-x'|^\gamma}{|x-y|^{n+\gamma}}, \quad \text{whenever } |x-x'| \leq |x-y|/2,$$

as well as the symmetric version of the second inequality in which the roles of x and y are interchanged.‡ Writing

$$(T_\varepsilon f)(x) = \int_{|x-y|>\varepsilon} K(x,y)\, f(y)\, dy, \qquad (T_* f)(x) = \sup_{\varepsilon>0} |(T_\varepsilon f)(x)|,$$

we have that

$$\int [(T_* f)(x)]^p \, \omega(x)\, dx \leq A_{p,\omega} \int [(Mf)(x)]^p \, \omega(x)\, dx$$

whenever f is bounded and has compact support, $\omega \in A_\infty$, and $1 < p < \infty$.

The proof of this extension requires only a minor modification of the argument given in §4.4. One point worth noting is the proof of the assertion that $\int [(T_* f)(x)]^p \, \omega(x)\, dx < \infty$ whenever f is bounded with compact support and $\int (Mf)^p \, \omega \, dx < \infty$, without the use of the Fourier transform. Indeed, let f be supported in $\{x : |x| \leq R\}$. Then clearly $|(T_* f)(x)| \leq A|x|^{-n}$ for $|x| \geq 2R$, while on the complementary set $|x| \leq 2R$, we use the fact (see (19)) that there is an $r > 1$ for which $\omega^r \cdot \chi_{\{|x|\leq 2R\}}$ is integrable, while $T_* f \in L^{r'}$, $1/r' + 1/r = 1$, since T_* is bounded on $L^{r'}$ (see Chapter 1, §7.4).

‡ Such operators are treated in detail in Chapter 7, §3.

6.14 Let $ML^p(\omega)$ denote the set of locally integrable functions f for which

$$\|f\|_{ML^p} = \left(\int_{\mathbf{R}^n} [(Mf)(x)]^p \, \omega(x)\, dx\right)^{1/p} < \infty.$$

When $1 \le p < \infty$, $ML^p(\omega)$ forms a Banach space with the indicated norm. Our assertion is that if $\int_{\mathbf{R}^n} \omega(x)\, dx = \infty$ and $ML^p(\omega)$ is nonempty, then C_0^∞ is dense in $ML^p(\omega)$. A particular case arises when $\omega \in A_\infty$; since $\omega(x)\, dx$ is a doubling measure, it follows that $\|\omega\|_{L^1(\mathbf{R}^n)} = \infty$ (see Chapter 1, §8.6(a)). Observe also that $ML^p(\omega) \ne \emptyset$ exactly when

$$\int (1 + |x|)^{-np} \, \omega(x)\, dx < \infty,$$

because $(Mf)(x) \ge c(1 + |x|)^{-n}$ whenever f is not identically zero.

To prove that C_0^∞ is dense in $ML^p(\omega)$, notice first that if $f \in ML^p(\omega)$, then

$$[M(f\chi_{{}^cB_R})](x) \to 0 \qquad \text{as } R \to \infty$$

for every x, otherwise there would be a $c > 0$ with $(Mf)(x) \ge c$ for all $x \in \mathbf{R}^n$; here B_R denotes the ball of radius R about the origin. Using the dominated convergence theorem, this shows that functions in ML^p can be approximated by those having compact support. Suppose now that $f \in ML^p$ is compactly supported. Let $\Phi \in C_0^\infty$, $\int \Phi\, dx = 1$; then $f * \Phi_t \in C_0^\infty$ for all $t > 0$. Moreover, $f * \Phi_t \to f$ in ML^p norm, as $t \to 0$, because $M(f * \Phi_t) \le cM(f)$, and $M(f - f * \Phi_t) \to 0$ almost everywhere, as $t \to 0$.

6.15 The primordial weighted inequality for the maximal operator, given by (9) of Chapter 2, has an analogue for the operators T and T_* considered in §4 (and more generally in §6.13): For any $1 < q < \infty$,

$$\int_{\mathbf{R}^n} [(T_*f)(x)]^q \, \omega(x)\, dx \le A_{q,r} \int_{\mathbf{R}^n} |f(x)|^q \, M_r(\omega)\, dx,$$

where $(M_r\omega)(x) = [M(\omega^r)(x)]^{1/r}$, $r > 1$, and M is the usual maximal operator.

Indeed, $\omega \le M_r(\omega)$; while by §5.2, we know that $M_r(\omega) \in A_1 \subset A_\infty$. Our assertion then follows from Theorem 2 in §4.2. For the original proof, see Córdoba and C. Fefferman [1976].

C. Other topics

6.16 There is a general underlying principle that clarifies the connection between weighted inequalities such as (9) of Chapter 2 or §6.15 above, and vector-valued inequalities such as in Chapter 2, §1.1. An illustrative special case of this principle can be stated as follows.

Suppose $\{T_j\}$ is a collection of sublinear operators that are uniformly bounded on $L^p(\mathbf{R}^n)$, for some fixed $p > 2$. Then

$$\|(\sum_j |T_j f_j|^2)^{1/2}\|_{L^p(\mathbf{R}^n)} \le A \|(\sum_j |f_j|^2)^{1/2}\|_{L^p(\mathbf{R}^n)}$$

is and only if, for each $u \in L^r(\mathbf{R}^n)$, $r = 2/p'$, there is a $v \in L^r(\mathbf{R}^n)$, with $\|v\|_{L^r} \le \|u\|_{L^r}$, so that

$$\int_{\mathbf{R}^n} |(T_j f)(x)|^2 \, |u(x)|\, dx \le A^2 \int_{\mathbf{R}^n} |f(x)|^2 \, |v(x)|\, dx$$

for all j.

There is a similar result for $p < 2$, but the inequality now becomes

$$\int |T_j f|^2 \, |v|^{-1}\, dx \le A^2 \int |f|^2 \, |u|^{-1}\, dx,$$

and r is replaced by $2/p$.

For this and various extensions, see Rubio de Francia [1980], García-Cuerva and Rubio de Francia [1985]; earlier ideas are in Maurey [1974].

6.17 It is a significant fact that the weighted theory for one exponent q, $1 < q < \infty$, already implies the full results for all p, $1 < p < \infty$. This is contained in the following "extrapolation theorem".

Suppose T is a sublinear operator and that for some q, $1 < q < \infty$, we have

$$\int |Tf|^q \, \omega\, dx \le c \int |f|^q \, \omega\, dx$$

for all $\omega \in A_q$, where c is independent of f and depends only on the A_q bound of ω. Then it follows that the same inequality, with q replaced by an arbitrary p, $1 < p < \infty$, holds for all $\omega \in A_p$.

Rubio de Francia [1982], [1984]; see also García-Cuerva [1983].

6.18 Much of the theory of weights presented here extends to the more general real-variable setting of Chapter 1. For example, one way to obtain a general version of Theorem 1 of the present chapter (the maximal theorem with weights) is to use the martingale maximal operator (arising in §6.24(b) of the previous chapter) as a substitute for M^Δ.

General systematic treatments are given in Jawerth [1986], Strömberg and Torchinsky [1989]. The latter work also develops a weighted H^p theory for $p \le 1$.

6.19 The theory of weights plays an important role in the study of boundary-value problems for Laplace's equation (and its generalizations) in the context where "minimal" smoothness is assumed. There are two alternative (but not identical) settings for these problems. In the former, Laplace's equation is given in a (bounded) domain whose boundary is assumed only to be Lipschitz; the latter is concerned with the case of a smooth boundary, but where the Laplacian is replaced by a generalization having "rough" coefficients. We consider the first situation here, §6.20 deals with the second.

Let Ω be a bounded region in $\mathbf{R}^{n+1}$ whose boundary is given locally as the graph of a Lipschitz function. Our object is to study the functions u with $\Delta u = 0$ in Ω and $u|_{\delta\Omega} = f$; in particular, we are concerned with the behavior of the nontangential maximal function u^*. In this connection, it is useful to observe that the space $b\Omega$, equipped with the quasi-distance $\rho(x,y) = |x-y|$ and the induced Lebesgue measure $d\sigma$, satisfies all the axioms in Chapter 1, §1. Note that $B(x,\delta)$ is the intersection of the corresponding Euclidean ball with $b\Omega$.

A key object of the theory is "harmonic measure" dh on $b\Omega$, defined by $u(\bar{x}) = \int_{b\Omega} f(x)\,dh(x)$, where $\bar{x}$ is a fixed point in Ω (and dh is associated to $\bar{x}$). It can be shown that:

(a) dh and $d\sigma$ are mutually absolutely continuous; in particular $dh = \omega\,d\sigma$.

(b) ω satisfies a reverse Hölder inequality, namely

$$\left(\frac{1}{\sigma(B(x,\delta))}\int_{B(x,\delta)} \omega^2\,d\sigma\right)^{1/2} \le \frac{C}{\sigma(B(x,\delta))}\int_{B(x,\delta)} \omega\,d\sigma.$$

(c) As a result, we have $\|u^*\|_{L^p(d\sigma)} \le A\|f\|_{L^p(d\sigma)}$ for $2 \le p < \infty$.

(d) If $b\Omega$ is of class C^1, then the reverse Hölder inequality for ω holds with exponent r, for all $r < \infty$; moreover, $\log\omega \in$ VMO. Also, one has that $\|u^*\|_{L^p(d\sigma)} \le A_p\|f\|_{L^p(d\sigma)}$ for all p, $1 < p \le \infty$.

The theory described here has developed further, so as to be able to encompass the "Neumann" problem and other regularity properties in this setting. This advance has required that harmonic measure cede some of its crucial role to other methods, in particular to ideas involving "layer potentials". For all of this, see Dahlberg [1977], [1979], and the further work described in Jerison and Kenig [1981], [1982], Kenig [1985], [1986]. The theory of layer potentials is based in part on the papers of Fabes, Jodeit, and Rivière [1978], and Coifman, McIntosh, and Y. Meyer [1982].

6.20 The second setting for the problem is where Ω is a smooth domain (e.g., the half-space $x_{n+1} > 0$), but where the Laplacian Δ is replaced by an operator

$$L(u) = \sum_{i,j} \frac{\partial}{\partial x_i}\left(a_{ij}\frac{\partial u}{\partial x_j}\right),$$

where $\{a_{ij}(x)\}$ is bounded, measurable, symmetric, and strictly positive definite. Again one considers the corresponding Dirichlet problem: $L(u) = 0$ in Ω, $u|_{b\Omega} = f$, and an attached harmonic measure of L, determined by

$$u(\bar{x}) = \int_{b\Omega} f(x)\,dh^L(x)$$

where, as before, $\bar{x}$ is a fixed point of Ω and dh^L is associated to $\bar{x}$.

There is a connection between this problem and that considered in §6.19 that comes about by mapping a Lipschitz graph domain Ω, given by $x_{n+1} > \phi(x_1,\dots,x_n)$, to the half-space $\mathbf{R}^{n+1}_+$ via the correspondence $x \mapsto \tilde{x}$, where $\tilde{x}_j = x_j$ for $1 \le j \le n$, and $\tilde{x}_{n+1} = x_{n+1} - \phi(x_1,\dots,x_n)$. This correspondence takes Δ to an L of the above kind.

However, it is no longer true, in the general case, that dh^L is absolutely continuous with respect to boundary measure (see §8.9 of Chapter 1), and so a significant problem is that of finding conditions for which this holds, and the further study of dh^L. While it is beyond our scope to pursue these matters here, we do want to indicate an additional insight concerning A_p theory that has arisen from these considerations. It involves the following point of view.

Suppose we are in $\mathbf{R}^1$. Then $\omega(x)\,dx$ is a doubling measure exactly when $\int_I \omega(x)$ is uniformly comparable to $\int_{\tilde{I}} \omega(x)$ for all intervals I; here $\tilde{I}$ denotes either of the intervals that are congruent to I and adjacent to it. This may be thought of as the multiplicative version of the additive inequality

$$\int_I f\,dx - \int_{\tilde{I}} f\,dx = O(|I|),$$

which is in fact Zygmund's condition $F(x+t)+F(x-t)-2F(x) = O(|t|)$ when $F(x) = \int^x f(u)\,du$. However this condition on F is not sufficient to guarantee that F be differentiable a.e.; instead, a square-function condition on F is required.† Thus, in passing back to the multiplicative picture, one may seek a square-function condition that characterizes weights satisfying some reverse Hölder inequality (i.e., those that belong to A_∞). This possibility is given further credence by the fact that $\omega \in A_\infty$ implies $\log\omega \in$ BMO, and because there is a square-function characterization of BMO (given in Chapter 4, §4.3). These heuristics are vindicated by the following.

Suppose $\omega\,dx$ is a doubling measure on $\mathbf{R}^n$, $\Phi \in \mathcal{S}$, $\Phi \ge 0$, and $\int \Phi\,dx = 1$. Then $\omega \in A_\infty$ exactly when

$$d\mu(x,t) = \frac{|\nabla_x(\omega * \Phi_t)|^2}{|\omega * \Phi_t|^2}\,t\,dx\,dt$$

is a Carleson measure on $\mathbf{R}^{n+1}_+$. There are also variants of this condition that characterize the exponent r of the reverse Hölder inequality satisfied by ω, or the A_p class to which it belongs.

These results, and applications to the Dirichlet problem for L, are in R. Fefferman, Kenig, and Pipher [1991], which also cites the earlier literature.

6.21 Methods involving analytic functions and operator theory can be used to give further insights into questions dealing with weighted inequalities for the Hilbert transform H.‡ Indeed, the first characterization of weighted inequalities for H was obtained using such ideas. In stating the results that have been obtained along these lines, it is more convenient to use the periodic Hilbert transform $\tilde{H}$, defined as follows. Let $\mathbf{T} = [-\pi,\pi]$ (with the usual identification with the unit circle), and set

$$\tilde{H}(e^{in\theta}) = \frac{\operatorname{sign} n}{i}e^{in\theta}, \quad n \ne 0, \qquad \tilde{H}(1) = 0;$$

† For this point, which is valid in $\mathbf{R}^n$, see *Singular Integrals*, Chapter 8.

‡ H is defined in Chapter 1, §6.2.

by linearity, $\tilde{H}$ extends to a bounded operator on $L^2(\mathbf{T})$. We next consider the mapping

$$z \mapsto w = \frac{i-z}{i+z}$$

which sends the upper half-plane $\Im(z) > 0$ to the unit disc $|w| < 1$, and maps $\mathbf{R}$ to $\mathbf{T}$ by $e^{i\theta} = \frac{i-x}{i+x}$. If f denotes a function on the unit circle, we let

$$(Jf)(x) = \frac{1}{i+x} f\left(\frac{i-x}{i+x}\right)$$

be the corresponding function on $\mathbf{R}$. A calculation then shows that

$$H = J(\tilde{H} + P_0)J^{-1},$$

where P_0 is the orthogonal projection on the one-dimensional subspace of constants in $L^2(\mathbf{T})$. This allows one to obtain weighted inequalities for H from those for $\tilde{H}$.

(a) A necessary and sufficient condition for a measure $d\mu$ on $\mathbf{T}$ to satisfy

$$\int_{\mathbf{T}} |\tilde{H}f|^2\, d\mu \le A \int_{\mathbf{T}} |f|^2\, d\mu \qquad (*)$$

is that $d\mu = \omega(\theta)\, d\theta$ be absolutely continuous, where $\omega = e^{u+\tilde{H}(v)}$ with $u, v \in L^\infty(\mathbf{T})$, and $\|v\|_{L^\infty} < \pi/2$.

(b) The condition $(*)$ on ω is equivalent to the existence of a nonnegative ω_1 so that $c^{-1} \le \omega/\omega_1 \le c$ and $|\tilde{H}(\omega_1)| \le c\omega_1$, with $c = c_A$.

(c) Another equivalent condition is that there is an $h = \mathcal{H}^1(\mathbf{T})^\dagger$ so that

$$2\omega \le \mathrm{Re}(h), \qquad |h| \le 2A\omega, \qquad \text{and} \quad |\arg(h)| \le \frac{\pi}{2} - \varepsilon_A.$$

For (a), see Helson and Szegö [1960]. The characterizations in (b) and (c) are in Cotlar and Sadosky [1979]; here the proofs use "lifting theorems" for Toeplitz-like operators. It is remarkable that there seems to be no direct link between these characterizations and the A_2 condition resulting from §4.6; in this connection, see Garnett and P. Jones [1978].

6.22 The techniques used in the proofs of (b) and (c) above can also be used to determine the size of the bound A in $(*)$ and, in addition, allow one to deal with the situation where different weights occur on either side of the inequality. Thus

$$\int_{\mathbf{T}} |Hf|^2\, d\mu_1 \le A \int_{\mathbf{T}} |f|^2\, d\mu_2$$

holds if and only if $d\mu_1 = \omega\, d\theta$, $d\mu_1 \le A\, d\mu_2$, and there exists an $h \in \mathcal{H}^1$ so that

$$|A\, d\mu_2 + d\mu_1 - h\, d\theta| \le |A\, d\mu_2 - d\mu_1|.$$

Cotlar and Sadosky [1979].

† Here $\mathcal{H}^1(\mathbf{T})$ denotes the restrictions to the boundary of holomorphic H^1 functions on the unit disc.

6.23 There is an L^p analogue of §6.21 (c). If $1 < p < \infty$, then

$$\int_{\mathbf{T}} |\tilde{H}f|^p\, d\mu \le A \int_{\mathbf{T}} |f|^p\, d\mu$$

holds if and only if $d\mu = \omega\, d\theta$ for a weight $\omega \ge 0$ such that, for every nonnegative $g \in L^{p^*}(\mathbf{T})$, $\frac{1}{p^*} = \left|1 - \frac{2}{p}\right|$, with $\|g\|_{L^{p^*}} \le 1$, there is an $a = a_g \in L^{p^*}$ (with $\|a\|_{L^{p^*}} \le c_A$) and an $h = h_g \in \mathcal{H}^1$ so that

$$g\omega^{2/p} \le \mathrm{Re}(h), \qquad |h| \le a\omega^{2/p}.$$

This is in Cotlar and Sadosky [1983]. Further results, involving multiparameter Hilbert transforms, and extensions to the Heisenberg group, can be found in Cotlar and Sadosky [1990].

Notes

§1 and §2. The A_p classes, $p > 1$, and a special case of §2.1 originate in Rosenblum [1962]. The class A_1 appears, at least implicitly, in C. Fefferman and Stein [1971].

§3. Theorem 1, as well as the corollary concerning the inclusion property of A_p, is due to Muckenhoupt [1972]. Our presentation follows more closely that of Coifman and C. Fefferman [1974], which exploits the reverse Hölder inequality. This idea in turn is based on Gehring [1973], where the concept arose in connection with quasi-conformal mappings.

§4. Theorem 2 is due to Coifman and C. Fefferman [1974]; see also the earlier results in Hunt, Muckenhoupt, and Wheeden [1973]. Another approach is via the sharp function; for this see Muckenhoupt and Wheeden [1976], which uses some ideas in Coifman [1972].

§5. The characterization of A_∞ is in Muckenhoupt [1974] and Coifman and C. Fefferman [1974]; that of A_1 is due to Coifman and Rochberg [1980]. The quotient representation of A_p via A_1 goes back to P. Jones [1980a]; the proof given here is taken from Coifman, P. Jones, and Rubio de Francia [1983].

For a more complete account of the theory of weighted inequalities and its ramifications, see García-Cuerva and Rubio de Francia [1985].

CHAPTER VI

Pseudo-Differential and Singular Integral Operators: Fourier Transform

The second part of this monograph is devoted to an exposition of concepts that belong to the L^2 theory, when this is taken in its broadest sense. We will describe a number of interrelated ideas that are quite different in character from those studied in the previous chapters. Nevertheless, they will serve to complement and expand significantly a major portion of the results we have already obtained. Stated briefly, these ideas center around the following: first, the Fourier transform; next, orthogonality methods generally; and then, the techniques and applications of oscillatory integrals.

We begin, in this chapter, with the Fourier transform, and we study that part of the theory of singular integrals and pseudo-differential operators most intimately connected with it. The close relation we have in mind comes about as follows.

A translation-invariant operator T on $\mathbf{R}^n$ can be represented by a multiplication operator on the Fourier transform side; that is, there is a multiplier $a(\xi)$ (called the *symbol* of T) so that, on a formal level, with T acting on functions of x, we have

$$T(e^{2\pi i x\cdot\xi}) = a(\xi)\, e^{2\pi i x\cdot\xi}$$

for all $\xi \in \mathbf{R}^n$. Note that T is bounded on $L^2(\mathbf{R}^n)$ exactly when a is a bounded function; also, the composition of two such operators corresponds to the product of their symbols.

Pseudo-differential operators may, to begin with, be viewed as generalizations of such operators that are "approximately" translation invariant. If T is one of these more general operators, it is characterized by its symbol $a(x,\xi)$, now a function of x as well as of ξ, so that (formally)

$$T(e^{2\pi i x\cdot\xi}) = a(x,\xi)\, e^{2\pi i x\cdot\xi}.$$

This can be expressed (using the Fourier inversion formula) as

$$(Tf)(x) = \int_{\mathbf{R}^n} a(x,\xi)\, e^{2\pi i x\cdot\xi}\, \widehat{f}(\xi)\, d\xi.$$

The main question then becomes that of making these ideas precise. We need to fix the conditions required of our symbols and see how they affect the properties of the corresponding operators. A variety of considerations come into play here (for example, the sense in which our operators are to be approximately translation-invariant). Experience has shown that the simplest and most useful class of symbols satisfying such (and other reasonable) requirements is the one given by the conditions

$$|\partial_x^\beta \partial_\xi^\alpha a(x,\xi)| \le A_{\alpha,\beta}(1+|\xi|)^{m-|\alpha|}$$

for all multi-indices α and β. The fixed number m is the *order* of the symbol.

We now list some of the main features of this class of symbols and their corresponding operators.

(a) We note first that all partial differential operators (whose coefficients, together with all their derivatives, are bounded) belong to this class. In this particular circumstance, the symbol is a polynomial in ξ, essentially the "characteristic polynomial" of the operator.

(b) The general operators of this class have a parallel description in terms of their kernels. That is, in a suitable sense,

$$(Tf)(x) = \int_{\mathbf{R}^n} K(x,y)\, f(y)\, dy;$$

besides enjoying a cancellation property, K is here characterized by differential inequalities "dual" to those for $a(x,\xi)$. In the key case where the order $m = 0$, this kernel representation makes T a singular integral operator.

(c) The crucial L^2 estimate, when $m = 0$, is a relatively simple consequence of Plancherel's theorem for the Fourier transform. With this, the L^p theory of Chapter 1 is therefore applicable.

(d) The product identity that holds in the translation-invariant case generalizes to the situation treated here as a symbolic calculus for the composition of operators. That is, there is an asymptotic formula for the composition of two such operators, whose main term is the pointwise product of their symbols.

(e) The succeeding terms of this formula are of decreasing orders. These orders measure not only the size properties of the symbols, but determine also the increasing smoothing properties of the corresponding operators. The smoothing properties are most neatly expressed in terms of the Sobolev spaces L^p_k and the Lipschitz spaces Λ_α.

1. Pseudo-differential operators

1.1 One of the principal motivations for the study of pseudo-differential operators is their wide applicability to partial differential equations. We shall briefly illustrate this by presenting the underlying heuristics in terms of what we may call the "freezing principle".

Suppose we are interested in studying the solutions of the classical elliptic second-order equation

$$(Lu)(x) = \sum a_{ij}(x) \frac{\partial^2 u(x)}{\partial x_i \partial x_j} = f(x),$$

where the matrix $\{a_{ij}(x)\}$ is assumed to be real, symmetric, positive definite, and smooth in x. A main goal is then to understand the operator P that is the inverse to L, or (more realistically) a P that inverts L up to a controllable error term. Put another way, we are looking for a P so that $LP = I + E$, where I is the identity operator and E is "small" in an appropriate sense.

To do this, we first fix an arbitrary point x_0, and freeze the operator L at x_0, obtaining a constant coefficient operator

$$L_{x_0} = \sum a_{ij}(x_0) \frac{\partial^2}{\partial x_i \partial x_j}.$$

Using the Fourier transform, L_{x_0} can be exactly inverted by the operator with multiplier

$$\Big(-4\pi^2 \sum_{i,j} a_{ij}(x_0)\xi_i\xi_j\Big)^{-1},$$

which is the inverse of the characteristic polynomial of L_{x_0}. To avoid the awkward (and largely irrelevant) singularity of this multiplier at $\xi = 0$, we introduce a smooth cut-off function η that vanishes near the origin and equals 1 for large ξ. We then define the operator P_{x_0} by

$$(P_{x_0}f)(x) = \int_{\mathbf{R}^n} \Big(-4\pi^2 \sum_{i,j} a_{ij}(x_0)\xi_i\xi_j\Big)^{-1} \eta(\xi)\widehat{f}(\xi)e^{2\pi i x\cdot\xi}\,d\xi$$

and observe that

$$L_{x_0}P_{x_0} = I + E_{x_0},$$

where the error term E_{x_0} is actually an infinite smoothing operator, because it is given by convolution with a fixed test function.

It is reasonable to suppose that the P we are looking for should be well approximated by P_{x_0} when x is near x_0. To make this precise, we unfreeze x_0 and define P by $(Pf)(x) = (P_x f)(x)$, i.e.,

$$(Pf)(x) = \int_{\mathbf{R}^n} \Big(-4\pi^2 \sum_{i,j} a_{ij}(x)\xi_i\xi_j\Big)^{-1} \eta(\xi)\widehat{f}(\xi)e^{2\pi i x\cdot\xi}\,d\xi.$$

The operator P so given is a prototype of a pseudo-differential operator. Moreover, one has $LP = I + E_1$, where the error operator E_1 is "smoothing of order 1". That this is indeed the case is the main point of the symbolic calculus described in §3 below.

1.2 Formally, a pseudo-differential operator is a mapping $f \mapsto T(f)$ given by

$$(Tf)(x) = \int_{\mathbf{R}^n} a(x,\xi)\widehat{f}(\xi)e^{2\pi i x\cdot\xi}\,d\xi, \tag{1}$$

with

$$\widehat{f}(\xi) = \int_{\mathbf{R}^n} f(x)e^{-2\pi i x\cdot\xi}\,dx$$

being the Fourier transform of f, and where $a(x,\xi)$ is the *symbol* of T. In order to emphasize the role of the symbol a, we will often write T as T_a; it is sometimes suggestive to also use the shorthand $T_a = a(x, D)$.

If L is a partial differential operator,

$$L = \sum_{|\alpha|\le m} a_\alpha(x)\Big(\frac{\partial}{\partial x}\Big)^\alpha,$$

then L is of the form (1), $L = T_a$, with

$$a(x,\xi) = \sum_{|\alpha|\le m} a_\alpha(x)(2\pi i\xi)^\alpha;$$

so the symbol in this case equals the characteristic polynomial of L. The desire to describe inverses or other functions of L in a similar fashion—in brief to develop a "functional calculus" for L—leads to the definition (1). That functions of L might be expressible in this form is suggested by the fact that this can be done directly in the constant coefficient case (i.e., when $a(x,\xi)$ is independent of x), as has been indicated above.

Still proceeding formally, we note that if the symbol a is independent of x, $a(x,\xi) = a_1(\xi)$, then T_a is a *multiplier* operator:

$$\widehat{(T_af)}(\xi) = a_1(\xi)\widehat{f}(\xi);$$

while if a is independent of ξ, $a(x,\xi) = a_2(x)$, then T_a is a *multiplication* operator:

$$(T_af)(x) = a_2(x)f(x).$$

Notice also that when $a(x,\xi) = a_2(x)a_1(\xi)$, then T_a is the ordered product $T_{a_2}T_{a_1}$; this ordering is consistent with the usual convention in writing partial differential operators, in which differentiation precedes multiplication. For general a, T_a may be thought of as a limit of linear combinations of operators corresponding to product symbols of this kind.

1.3 To pass from merely formal statements we need to make precise the definition of the pseudo-differential operator (1) and the class of symbols that is used. We shall consider in this chapter the standard symbol class, denoted by S^m, which is the most common and useful of the general symbol classes. A function a belongs to S^m (and is said to be of *order* m) if $a(x,\xi)$ is a C^∞ function of $(x,\xi) \in \mathbf{R}^n \times \mathbf{R}^n$ and satisfies the differential inequalities

$$|\partial_x^\beta \partial_\xi^\alpha a(x,\xi)| \leq A_{\alpha,\beta}(1+|\xi|)^{m-|\alpha|}, \tag{2}$$

for all multi-indices α and β. For instance, (2) holds when a is a polynomial in ξ (independent of x) of degree m.

Given a symbol $a \in S^m$, the operator T_a will initially be defined on the Schwartz class of testing functions $\mathcal{S}$. In fact, it is immediate that the integral (1) converges absolutely and is infinitely differentiable. An integration by parts argument shows that $T_a(f)$ is a rapidly decreasing function. Indeed, observe that

$$(I-\Delta_\xi)e^{2\pi i x\cdot\xi} = (1+4\pi^2|x|^2)e^{2\pi i x\cdot\xi},$$

and define the operator

$$L_\xi = (1+4\pi^2|x|^2)^{-1}(I-\Delta_\xi);$$

thus $(L_\xi)^N e^{2\pi i x\cdot\xi} = e^{2\pi i x\cdot\xi}$. Inserting this in (1) and carrying out the repeated integrations by parts gives us

$$(T_a f)(x) = \int (L_\xi)^N[a(x,\xi)\widehat{f}(\xi)]e^{2\pi i x\cdot\xi}\,d\xi,$$

from which the rapid decrease of $T_a(f)$ is evident. Since a similar argument works for any partial derivative of $T_a(f)$, we see that T_a maps $\mathcal{S}$ to $\mathcal{S}$, and that this mapping is continuous. It is also useful to note that if $\{a_k\}$ is a *pointwise* convergent sequence of symbols in S^m that satisfy the inequalities (2) uniformly in k, then $T_{a_k}(f) \to T_a(f)$ in $\mathcal{S}$ whenever $f \in \mathcal{S}$; we shall make use of this fact later.

1.4 An alternative way of writing the definition (1) is as a repeated integral

$$(T_a f)(x) = \int\int a(x,\xi)e^{2\pi i\xi\cdot(x-y)}f(y)\,dy\,d\xi. \tag{3}$$

However, the integral (3) does not necessarily converge absolutely, even when $f \in \mathcal{S}$. To deal with this, and also to facilitate other manipulations with pseudo-differential operators, it is convenient to approximate a given symbol by symbols of compact support. For this purpose, fix a function $\gamma \in C_0^\infty(\mathbf{R}^n \times \mathbf{R}^n)$ with $\gamma(0,0)=1$. Set $a_\varepsilon(x,\xi) = a(x,\xi)\gamma(\varepsilon x, \varepsilon\xi)$. Notice that if a belongs to the symbol class S^m, then so do the a_ε, and they satisfy the defining inequalities (2) uniformly in ε, for $0<\varepsilon\leq 1$.

Next observe that, as was remarked at the end of §1.3, $T_{a_\varepsilon} \to T_a$, in the sense that $T_{a_\varepsilon}(f) \to T_a(f)$ in $\mathcal{S}$, whenever $f \in \mathcal{S}$, as $\varepsilon \to 0$. Moreover, since the alternate definition (3) clearly holds when a has compact support, we get that

$$(T_a f)(x) = \lim_{\varepsilon\to 0}\int\int a_\varepsilon(x,\xi)e^{2\pi i\xi\cdot(x-y)}f(y)\,dy\,d\xi.$$

In addition, one can verify directly the duality relation

$$\langle T_a f, g\rangle = \langle f, T_a^* g\rangle, \tag{4}$$

whenever $f, g \in \mathcal{S}$; here

$$(T_a^* g)(y) = \lim_{\varepsilon\to 0}\int\int \bar{a}_\varepsilon(x,\xi)e^{2\pi i\xi\cdot(y-x)}g(x)\,dx\,d\xi, \tag{5}$$

and $\langle f,g\rangle$ denotes $\int_{\mathbf{R}^n} f(x)\bar{g}(x)\,dx$. Imitating our proof that T_a maps $\mathcal{S}$ to $\mathcal{S}$, we set

$$L_x = (1+4\pi^2|x|^2)^{-1}(I-\Delta_x),$$

and note that the right side of (5) equals

$$\lim_{\varepsilon\to 0}\int\int (L_x)^N[\bar{a}_\varepsilon(x,\xi)g(x)]e^{2\pi i\xi\cdot(y-x)}\,dx\,d\xi,$$

from which it is clear that T_a^* also maps $\mathcal{S}$ to $\mathcal{S}$.

Thus the pseudo-differential operator T_a, initially defined as a mapping from $\mathcal{S}$ to $\mathcal{S}$, extends via the duality (4) to a mapping from the space of tempered distributions to itself. Notice also that T_a is automatically continuous on this space.

1.5 We note that our symbols (at least when the order $m=0$) satisfy conditions for large ξ of the kind enjoyed by the standard multipliers arising for the translation-invariant singular integrals described in Chapter 1, §6.2. However, our symbols are, in addition, always assumed to be smooth for *all* ξ.[†] This restriction has the advantage of retaining the local behavior of such operators, while eliminating a variety of problems for large x that are not always relevant in applications such as partial differential equations. An additional restriction of this type which is sometimes made is that the symbol $a(x,\xi)$ have compact support in x. While this assumption often simplifies an argument, we shall see below that it is not essential, and we shall not be bound by it.

[†] In this connection, see also §7.15 and §7.16 below.

2. An L^2 theorem

After these preliminaries, we state the first main result.

THEOREM 1. *Suppose a is a symbol of order 0, i.e., that $a \in S^0$. Then the operator T_a, initially defined on $\mathcal{S}$, extends to a bounded operator from $L^2(\mathbf{R}^n)$ to itself.*

To prove the theorem, it suffices to show that

$$\|T_a(f)\|_{L^2} \leq A\|f\|_{L^2}, \qquad \text{whenever } f \in \mathcal{S}, \tag{6}$$

with A independent of f. In fact, suppose that $f \in L^2$, and let $\{f_n\}$ be a sequence in $\mathcal{S}$ so that $f_n \to f$ in L^2. Then because of (6), $T_a(f_n)$ converges in L^2 norm, while, in view of the remarks in §1.4 above, $T_a(f_n)$ converges to $T_a(f)$ in the sense of distributions.

2.1 Turning to the proof of (6), we establish it first under the assumption that the symbol $a(x,\xi)$ has compact support in x. This restriction makes it possible to use the Fourier transform in the x variable, and expand the symbol in a simple way using products of symbols depending on x or ξ only.

We write

$$a(x,\xi) = \int_{\mathbf{R}^n} \widehat{a}(\lambda,\xi)e^{2\pi i\lambda\cdot x}\,d\lambda,$$

where

$$\widehat{a}(\lambda,\xi) = \int_{\mathbf{R}^n} a(x,\xi)e^{-2\pi i x\cdot\lambda}\,dx.$$

Since a has compact x-support, the integral defining $\widehat{a}(\lambda,\xi)$ converges. An integration by parts shows that, for each multi-index α,

$$(2\pi i\lambda)^\alpha \widehat{a}(\lambda,\xi) = \int_{\mathbf{R}^n} [\partial_x^\alpha a(x,\xi)]\, e^{-2\pi i x\cdot\lambda}\,dx$$

and that $|(2\pi i\lambda)^\alpha\widehat{a}(\lambda,\xi)| \leq c_\alpha$, uniformly in ξ. As a result we have

$$\sup_\xi |\widehat{a}(\lambda,\xi)| \leq A_N(1+|\lambda|)^{-N} \tag{7}$$

for arbitrary $N \geq 0$. Now

$$\begin{aligned}(T_a f)(x) &= \int a(x,\xi)\widehat{f}(\xi)e^{2\pi i x\cdot\xi}\,d\xi \\ &= \int\int \widehat{a}(\lambda,\xi)e^{2\pi i\lambda\cdot x}\widehat{f}(\xi)e^{2\pi i x\cdot\xi}\,d\xi\,d\lambda = \int (T^\lambda f)(x)\,d\lambda,\end{aligned}$$

where

$$(T^\lambda f)(x) = e^{2\pi i x\cdot\lambda}(T_{\widehat{a}(\lambda,\xi)} f)(x).$$

Since for each λ, $T_{\widehat{a}(\lambda,\xi)}$ is a multiplier operator on the Fourier transform side, by Plancherel's theorem we have that

$$\|T_{\widehat{a}(\lambda,\xi)} f\|_{L^2} \leq \sup_\xi |\widehat{a}(\lambda,\xi)| \cdot \|\widehat{f}\|_{L^2} = \sup_\xi |\widehat{a}(\lambda,\xi)| \cdot \|f\|_{L^2}.$$

Therefore, by (7), we have $\|T^\lambda\| \leq A_N(1+|\lambda|)^{-N}$.

Because $T_a = \int T^\lambda\,d\lambda$, this yields

$$\|T_a\| \leq A_N \int (1+|\lambda|)^{-N}\,d\lambda < \infty,$$

if we choose $N > n$. Thus (6) is proved when a has compact support in x.

2.2 The proof for general symbols requires that we also use the singular integral realization of the operator T_a. That is, we shall write

$$(T_a f)(x) = \int_{\mathbf{R}^n} k(x,z)f(x-z)\,dz \tag{8}$$

with the following understanding. First, for each x, $k(x,\cdot)$ is the distribution whose Fourier transform is the function $a(x,\cdot)$; written formally this is the identity

$$a(x,\xi) = \int_{\mathbf{R}^n} k(x,z)e^{-2\pi i z\cdot\xi}\,dz.$$

Thus (8) can be interpreted as the convolution of the distribution $k(x,\cdot)$ with the function $f \in \mathcal{S}$, evaluated at the point x.

We are about to prove that, away from the origin, the distribution $k(x,\cdot)$ agrees with a function that is rapidly decreasing at infinity. Therefore, (8) can be interpreted as a convergent integral for x lying outside the support of f, if we assume that the complement of that support is nonempty. We will also use the more standard notation

$$(T_a f)(x) = \int K(x,y)f(y)\,dy,$$

again for $x \notin \operatorname{supp}(f)$, where we have set $K(x,y) = k(x,x-y)$.

The precise study of the singular integral kernel $k(x,\cdot)$ will be taken up in §4 below. For our immediate purposes a very crude estimate will suffice, namely that, away from the origin, $k(x,\cdot)$ is a function that satisfies the inequality

$$|k(x,z)| \leq A_N|z|^{-N}, \quad \text{for all } |z| \geq 1 \text{ and all } N > 0, \tag{9}$$

uniformly in x. To see this, write

$$(T_a f)(x) = \int a(x,\xi)\widehat{f}(\xi)e^{2\pi i x\cdot\xi}\,d\xi.$$

By the standard properties of convolutions of test functions and distributions, this integral equals $[k(x,\cdot) * f](x)$, where $k(x,\cdot)$ is the distribution whose Fourier transform is the function $a(x,\cdot)$, as we have written in (8). Next, $(-2\pi i z)^\alpha k(x,\cdot)$, with the distribution $k(x,\cdot)$ thought of as acting on functions of z, equals the inverse Fourier transform of $\partial_\xi^\alpha a(x,\xi)$; by (2), $\partial_\xi^\alpha a(x,\xi)$ is integrable in ξ whenever $|\alpha| \geq n+1$. This shows that $k(x,\cdot)$ equals a function $k(x,z)$ away from the origin, and that $|z|^N|k(x,z)| \leq A_N$ whenever $N > n$. Thus (9) is also proved.

2.3 We return to the proof of Theorem 1, without assuming that $a(x,\xi)$ has compact x-support. To begin with, we intend to show that, for each $x_0 \in \mathbf{R}^n$,

$$\int_{|x-x_0|\le 1} |(T_a f)(x)|^2\,dx \le A_N \int_{\mathbf{R}^n} \frac{|f(x)|^2\,dx}{(1+|x-x_0|)^N}, \quad \text{all } N \ge 0. \quad (10)$$

This estimate has an interest in its own right because it illustrates a pseudo-locality feature of the operators T_a: the main contribution of $T_a f$ in the unit ball about x_0 comes from the values of $f(x)$ for x near that ball, in view of the rapidly decaying term $(1+|x-x_0|)^{-N}$.

We prove (10) first when $x_0 = 0$. To do this we split f by $f = f_1 + f_2$, with f_1 supported in $B(3)$,[‡] f_2 supported outside $B(2)$, f_1 and f_2 smooth, and with $|f_1|, |f_2| \le |f|$. We fix $\eta \in C_0^\infty$ so that $\eta \equiv 1$ in $B(1)$. Then $\eta T_a(f_1) = T_{\eta a}(f_1)$, and the symbol $\eta(x)a(x,\xi)$ has compact x-support, so the previous result applies. Hence

$$\int_{B(1)} |T_a f_1|^2 \le \int_{\mathbf{R}^n} |T_{\eta a} f_1|^2 \le A \int_{\mathbf{R}^n} |f_1|^2 \le A \int_{B(3)} |f|^2. \quad (11)$$

However, if $x \in B(1)$, since f_2 is supported away from $B(2)$, the representation (8) holds:

$$(T_a f_2)(x) = \int_{{}^cB(2)} k(x, x-z) f_2(z)\,dz.$$

Since $|x-z| \ge 1$ when $x \in B(1)$, $z \notin B(2)$, we can invoke the estimate (9) and obtain that

$$\begin{aligned} |(T_a f_2)(x)| &\le A \int_{{}^cB(2)} |f(z)||z|^{-N}\,dz \\ &\le A \int |f(z)|(1+|z|)^{-N}\,dz. \end{aligned}$$

Therefore by Schwarz's inequality, we get (provided $N > n$),

$$\int_{B(1)} |(T_a f_2)(x)|^2\,dx \le A \int |f(x)|^2 (1+|x|)^{-N}\,dx. \quad (12)$$

Now $T_a(f) = T_a(f_1) + T_a(f_2)$, and so combining (11) with (12) proves (10) when $x_0 = 0$.

[‡] Here $B(r)$ denotes the ball of radius r about the origin.

2.3.1 The passage to (10) for general x_0 can be achieved by noting that, while an individual pseudo-differential operator is not (in general) translation-invariant, the *class* defined by (2), in fact, is. To see this, let τ_h, $h \in \mathbf{R}^n$, denote the unitary translation operator given by

$$(\tau_h f)(x) = f(x-h).$$

Then, as is easily verified,

$$\tau_h T_a \tau_{-h} = T_{a_h},$$

where

$$a_h(x,\xi) = a(x-h,\xi).$$

Note that the symbols a_h satisfy the same estimates (2) that a does, uniformly in h. Hence (10) holds for $x_0 = 0$, but with a replaced by a_h, with a bound independent of h. If we set $h = x_0$, we see that (10) is established (with the bound A_N independent of x_0).

Now it is only a matter of integrating (10) with respect to x_0, choosing $N > n$, and interchanging orders of integration. The result is (6), and the theorem is proved.

2.4 By combining the L^2 result just obtained with the theory described in Chapter 1, we can also prove the L^p boundedness of these operators. These matters, as well as other regularity properties of pseudo-differential operators, will be taken up in §5 below.

3. The symbolic calculus

The main result describing composition of pseudo-differential operators can be stated as follows.

THEOREM 2. *Suppose a and b are symbols belonging to S^{m_1} and S^{m_2} respectively. Then there is a symbol* c *in $S^{m_1+m_2}$ so that*

$$T_c = T_a \circ T_b.$$

Moreover,

$$c \sim \sum_\alpha \frac{(2\pi i)^{-|\alpha|}}{\alpha!} (\partial_\xi^\alpha a)\cdot(\partial_x^\alpha b)$$

in the sense that

$$c - \sum_{|\alpha|<N} \frac{(2\pi i)^{-|\alpha|}}{\alpha!} \partial_\xi^\alpha a \cdot \partial_x^\alpha b \in S^{m_1+m_2-N}, \quad (13)$$

for all $N > 0$.

Remark. The key point of the calculus is already contained in the assertion concerning the main term of the asymptotic development of the symbol, namely that $c - a\cdot b$ is a symbol of lower order (i.e., of order $m_1 + m_2 - 1$). Note that each term $(\partial_\xi^\alpha a)\cdot(\partial_x^\alpha b)$ is a symbol in the class $S^{m_1+m_2-|\alpha|}$.

3.1 For the proof of the theorem we first calculate $T_a \circ T_b$ formally; we assume that a and b have compact support so that our manipulations are justified. We use the alternate formula (3) to write

$$(T_b f)(y) = \int b(y,\xi)\, e^{2\pi i\xi\cdot(y-z)} f(z)\, dz\, d\xi;$$

then we apply T_a, again in the form (3), but here with the variable η replacing ξ in the integration. The result is

$$T_a(T_b f)(x) = \int a(x,\eta)\, b(y,\xi)\, e^{2\pi i\eta\cdot(x-y)} e^{2\pi i\xi\cdot(y-z)} f(z)\, dz\, d\xi\, dy\, d\eta.$$

Now $e^{2\pi i\eta\cdot(x-y)} \cdot e^{2\pi i\xi\cdot(y-z)} = e^{2\pi i(x-y)\cdot(\eta-\xi)} \cdot e^{2\pi i(x-z)\cdot\xi}$, so

$$T_a(T_b f)(x) = \int c(x,\xi)\, e^{2\pi i(x-z)\cdot\xi} f(z)\, dz\, d\xi,$$

with

$$c(x,\xi) = \int a(x,\eta)\, b(y,\xi)\, e^{2\pi i(x-y)\cdot(\eta-\xi)}\, dy\, d\eta. \tag{14}$$

We can also carry out the integration in the y-variable. This leads to the corresponding Fourier transform of b in that variable, and allows us to rewrite (14) as

$$c(x,\xi) = \int a(x,\xi+\eta)\, \widehat{b}(\eta,\xi) e^{2\pi i x\cdot\eta}\, d\eta. \tag{15}$$

3.1.1 We can now proceed using formulae (14) and (15), by replacing a and b with a_ε and b_ε respectively; here $a_\varepsilon(x,\xi) = a(x,\xi)\cdot\gamma(\varepsilon x, \varepsilon\xi)$, $b_\varepsilon(x,\xi) = b(x,\xi)\cdot\gamma(\varepsilon x,\varepsilon\xi)$, as in §1.4. We note that a_ε and b_ε satisfy the same differential inequalities that a and b do, uniformly in ε, $0 < \varepsilon \le 1$. In what follows, the explicit dependence on ε will be suppressed, but all estimates will be made independent of ε. The passage to the limit as $\varepsilon \to 0$ will then give us our desired result, as will become clear below.

3.2 It is simpler to prove the asymptotic formula (13) when the symbol $b(x,\xi)$ is assumed to have compact support in the x-variable, and we consider this case first. Together with the differential inequalities (2) satisfied by b, this extra assumption allows us to assert that

$$|\widehat{b}(\eta,\xi)| \le A_{\bar N}(1+|\eta|)^{-\bar N}(1+|\xi|)^{m_2}, \qquad \text{for all } \bar N \ge 0, \tag{16}$$

and so we try to exploit the formula (15). In that formula, we are tempted to replace $a(x,\xi+\eta)$ with $a(x,\xi)$; indeed, using Taylor's formula, we are justified in substituting

$$\sum_{|\alpha|<N} \frac{1}{\alpha!}\partial_\xi^\alpha a(x,\xi)\, \eta^\alpha + R_N(x,\xi,\eta) \quad \text{for } a(x,\xi+\eta), \tag{17}$$

with a suitable remainder R_N.

Observe that each α in the sum contributes

$$\frac{1}{\alpha!}\int \partial_\xi^\alpha a(x,\xi)\cdot \eta^\alpha\, \widehat{b}(\eta,\xi) e^{2\pi i x\cdot\eta}\, d\eta$$

to (15), which, upon using the Fourier inversion formula, equals

$$\frac{(2\pi i)^{-|\alpha|}}{\alpha!}(\partial_\xi^\alpha a(x,\xi))\cdot(\partial_x^\alpha b(x,\xi)),$$

and is precisely the term corresponding to α in the formula (13).

3.2.1 For the remainder term in Taylor's formula (17), we use the fact that R_N is majorized by a multiple of $|\eta|^N$ times the maximum of derivatives of order N of $a(x,\xi)$ with respect to ξ, taken on the line segment joining ξ to $\xi+\eta$. This gives the following estimates, which are easy consequences of (2):

$$|R_N(x,\xi,\eta)| \le A_N|\eta|^N(1+|\xi|)^{m_1-N}, \quad \text{for } |\xi| \ge 2|\eta|, \tag{18a}$$

and

$$|R_N(x,\xi,\eta)| \le A_N|\eta|^N, \quad \text{for all } \xi \text{ and } \eta, \tag{18b}$$

provided $N \ge m_1$.

Now the difference in (13) equals

$$\int R_N(x,\xi,\eta)\, \widehat{b}(\eta,\xi)\, e^{2\pi i x\cdot\eta}\, d\eta. \tag{19}$$

In estimating the size of (19) we can now invoke (16) and obtain the majorization

$$A_{N,\bar N}(1+|\xi|)^{m_1+m_2-N}\int (1+|\eta|)^{-\bar N}|\eta|^N\, d\eta$$
$$+\, A_{N,\bar N}(1+|\xi|)^{m_2}\int_{2|\eta|\ge|\xi|}(1+|\eta|)^{-\bar N}|\eta|^N\, d\eta,$$

the first term representing the contribution of the η with $2|\eta| \le |\xi|$, and the second term coming from the complementary set. Since $\bar N$ is at our disposal, we choose it sufficiently large, getting an estimate that is bounded by $A(1+|\xi|)^{m_1+m_2-N}$. This is the correct size estimate for the difference in the asymptotic formula (13).

Similarly, if we apply $\partial_x^\beta\partial_\xi^\alpha$ to this difference, we get that it is majorized by the corresponding estimate, namely $A_{\alpha,\beta}(1+|\xi|)^{m_1+m_2-N-|\alpha|}$. In view of (15), this is because $\partial_x^\beta a(x,\xi)$ is still a symbol of order m_1, while $\partial_\xi^\alpha b(x,\xi)$ is now a symbol of order $m_2 - |\alpha|$. Thus we have proved the asymptotic formula (13), under the assumption that b has compact x-support.

3.3 To consider the case when we do not assume that b has compact support, we observe that it suffices to establish the inequalities involved in (13) for x near an arbitrary but fixed point x_0.

We let $\eta(x)$ be a C^∞ function that equals 1 for $|x-x_0| \le 1$ and is supported in $|x-x_0| \le 2$, and write $b = \eta b + (1-\eta)b = b_1 + b_2$. The asymptotic formula for the symbol of $T_a T_{b_1}$ has been established above and, for x near x_0, it is the same as the claimed formula for $T_a T_b$.

It remains therefore to show that if c' is the symbol of $T_a T_{b_2}$, then

$$|c'(x,\xi)| \le A_N(1+|\xi|)^{m_1+m_2-N}, \qquad \text{all } N \ge 0, \tag{20}$$

as long as x is near x_0, say $|x-x_0| \le 1/2$. To see this, we use the formula (14), which gives

$$c'(x,\xi) = \int a(x,\eta)\, b_2(y,\xi)\, e^{2\pi i(x-y)\cdot(\eta-\xi)}\, dy\, d\eta. \tag{21}$$

In the above integral, first integrate by parts in the η-variable, using the identity

$$\Delta_\eta^{N_1}(e^{2\pi i(x-y)\cdot(\eta-\xi)}) = (-4\pi^2)^{N_1} \cdot |x-y|^{2N_1} \cdot e^{2\pi i(x-y)\cdot(\eta-\xi)}.$$

The result is the formula (21) with $a(x,\eta)$ replaced by $\dfrac{\Delta_\eta^{N_1} a(x,\eta)}{(-4\pi^2|x-y|^2)^{N_1}}$.

Next, integrate by parts with respect to the y-variable, using the identity

$$(I-\Delta_y)^{N_2}(e^{2\pi i(x-y)\cdot(\eta-\xi)}) = (1+4\pi^2|\eta-\xi|^2)^{N_2} \cdot e^{2\pi i(x-y)\cdot(\eta-\xi)}.$$

In carrying out the indicated differentiations, we note that if $|x-x_0| \le 1/2$, then $|x-y| \ge 1/2$ in (21), because $b_2(y,\xi)$ is supported outside the set where $|y-x_0| \le 1$.

Thus we obtain

$$|c'(x,\xi)| \le A_{N_1,N_2} \int \frac{(1+|\eta|)^{m_1-2N_1}\,(1+|\xi|)^{m_2}}{(1+|\eta-\xi|)^{2N_2}\,(1+|x-y|)^{2N_1}}\, dy\, d\eta.$$

From this it is clear that by choosing N_1 and N_2 sufficiently large we can achieve (20), concluding the estimates involved in the asymptotic formula (13).

3.4 The final point in the proof is the passage to the limit, $\varepsilon \to 0$. With $a_\varepsilon(x,\xi) = a(x,\xi)\cdot\gamma(\varepsilon x, \varepsilon\xi)$, $b\varepsilon(x,\xi) = b(x,\xi)\cdot\gamma(\varepsilon x,\varepsilon\xi)$, define c_ε by

$$T_{c_\varepsilon} = T_{a_\varepsilon} T_{b_\varepsilon}.$$

What we have shown is that c_ε belongs uniformly to $S^{m_1+m_2}$, and satisfies the formula (13) (with c_ε, a_ε, and b_ε, replacing c, a, and b) uniformly in ε. Moreover, it is clear from the above arguments that c_ε converges pointwise to a limit c. Thus, c belongs to $S^{m_1+m_2}$, satisfies (13), and by the continuity properties described in §1.4, we have $T_c = T_a T_b$. Theorem 2 is therefore proved.

3.5 An example. By a simple example, we show how the symbolic calculus of Theorem 2 can be applied, making precise the heuristics described in §1.1 above. Suppose a is a symbol of order m, which we assume is *elliptic*, in the sense that

$$|a(x,\xi)| \ge A|\xi|^m, \qquad \text{for large } \xi,$$

uniformly in x. Then, with $L = T_a$, we can find an approximate inverse to L as follows.

Set $b(x,\xi) = \eta(\xi)[a(x,\xi)]^{-1}$, where η is smooth, vanishes near the origin, and $\eta(\xi) \equiv 1$ for large $|\xi|$; one checks that $b \in S^{-m}$. If we set $P = T_b$ then, by Theorem 2, $LP = I + E_1$; also $PL = I + E_2$, where the E_i are pseudo-differential operators, each with symbol of order -1. Such error operators have one degree of smoothing, as we shall see below in §5; in fact, a refinement of this construction allows us to modify P so that the error terms are smoothing of infinite order, giving us what is usually called a *parametrix* for L. See also §7.11–§7.13 below.

4. Singular integral realization of pseudo-differential operators

We continue the study (begun in §2.2) of integral representations of pseudo-differential operators T_a, given in the form

$$(T_a f)(x) = \int_{\mathbf{R}^n} k(x,z) f(x-z)\, dz, \tag{22}$$

where, for each $x \in \mathbf{R}^n$, $k(x,\cdot)$ is the distribution that is the inverse Fourier transform (in the ξ-variable) of the function $a(x,\xi)$. We have already observed that, for each x, $k(x,\cdot)$ agrees with a function $k(x,z)$ for z away from the origin, and that (22) holds (as a convergent integral) for each x outside the support of f. The precise version of the estimate (9) is then contained in the following.

PROPOSITION 1. *Suppose $a \in S^m$. Then the kernel $k(x,z)$ is in $C^\infty(\mathbf{R}^n \times (\mathbf{R}^n \backslash \{0\}))$, and satisfies*

$$|\partial_x^\beta \partial_z^\alpha k(x,z)| \le A_{\alpha,\beta,N} |z|^{-n-m-|\alpha|-N}, \qquad z \ne 0, \tag{23}$$

for all multi-indices α and β, and all $N \ge 0$ so that $n+m+|\alpha|+N > 0$.

4.1 Dyadic decomposition. The proof of this proposition, as well as many other arguments that involve explicitly (or implicitly) the Fourier transform, makes use of the division of the dual (frequency) space into "dyadic" spherical shells. This decomposition—whose ideas originated in the work of Bernstein, Littlewood and Paley, and others—will now be described in the form most suitable for us.

We begin by fixing η, a C^∞ function of compact support, defined in the ξ-space $\mathbf{R}^n$, with the properties that $\eta(\xi) = 1$ for $|\xi| \le 1$, and $\eta(\xi) = 0$ for $|\xi| \ge 2$. Together with η, we define another function δ, by $\delta(\xi) = \eta(\xi) - \eta(2\xi)$. Then we have the following two "partitions of unity" of the ξ-space:

$$1 = \eta(\xi) + \sum_{j=1}^{\infty} \delta(2^{-j}\xi), \qquad \text{all } \xi, \tag{24}$$

and

$$1 = \sum_{j=-\infty}^{\infty} \delta(2^{-j}\xi). \qquad \text{all } \xi \ne 0. \tag{25}$$

In fact,

$$\eta(\xi) + \sum_{j=1}^{\ell} \delta(2^{-j}\xi) = \eta(2^{-\ell}\xi) \to 1,$$

as $\ell \to \infty$, for all ξ; while

$$\sum_{j=\ell'}^{\ell} \delta(2^{-j}\xi) = \eta(2^{-\ell}\xi) - \eta(2^{-\ell'+1}\xi) \to 1,$$

if $\ell \to \infty$, $\ell' \to -\infty$, and $\xi \ne 0$. Note also that $\delta(\xi)$ is supported in the shell $1/2 \le |\xi| \le 2$, so that the $\delta(2^{-j}\xi)$ are supported in the shells $2^{j-1} \le |\xi| \le 2^{j+1}$. It follows that for each ξ there are at most two nonzero terms in the sums (24) and (25).

We shall now consider the expression of this decomposition in the x-space. Starting again with our fixed function η, we let Φ be the inverse Fourier transform of η, i.e., $\widehat{\Phi}(\xi) = \eta(\xi)$. Then $\Phi \in \mathcal{S}$, and $\int \Phi\, dx = 1$. Let us also define Ψ by

$$\widehat{\Psi}(\xi) = \delta(\xi) = \eta(\xi) - \eta(2\xi).$$

Then $\int \Psi\, dx = 0$. Writing $\Phi_t(x) = t^{-n}\Phi(x/t)$ (and $\Psi_t(x) = t^{-n}\Psi(x/t)$), we have $\Psi_{2^{-j}} = \Phi_{2^{-j}} - \Phi_{2^{-j+1}}$, while $\widehat{\Phi_{2^{-j}}}(\xi) = \eta(2^{-j}\xi)$, $\widehat{\Psi_{2^{-j}}}(\xi) = \delta(2^{-j}\xi)$.

We now define the "partial sum" operator S_j by†

$$S_j(f) = f * \Phi_{2^{-j}}, \tag{26}$$

and the corresponding difference operator Δ_j by

$$\Delta_j(f) = S_j(f) - S_{j-1}(f) = f * \Psi_{2^{-j}}. \tag{27}$$

† Of course, the sequence (26) may be viewed in terms of the approximations of the identity already discussed in Chapter 1, §6.1 and Chapter 2, §2.1.

In parallel with (24) and (25), we have the operator identities

$$I = S_0 + \sum_{j=1}^{\infty} \Delta_j \tag{24a}$$

and

$$I = \sum_{j=-\infty}^{\infty} \Delta_j. \tag{25a}$$

Note that if f is a tempered distribution, $S_j(f)$ is well defined, and

$$S_0(f) + \sum_{j=1}^{\ell} \Delta_j(f) = S_\ell(f) \to f, \qquad \text{as } \ell \to \infty,$$

in the sense of distributions; this assertion is easily verified upon testing it with an arbitrary element of $\mathcal{S}$, and establishes the identity (24a). However, it is not true that $S_\ell(f) \to 0$ as $\ell \to -\infty$, for arbitrary f. While this convergence does take place when f is suitably small at infinity (say if $f \in L^p$, $p < \infty$), it fails when $f \equiv 1$. Thus (25a) holds only under some restriction on f.

4.2 We now return to our operator T_a; using the identity (25a), we shall write it as

$$T_a = TS_0 + \sum_{j=1}^{\infty} T\Delta_j = \sum_{j=0}^{\infty} T_{a_j},$$

where $a_0(x,\xi) = a(x,\xi)\eta(\xi)$, and $a_j(x,\xi) = a(x,\xi)\delta(2^{-j}\xi)$ for $j = 1, 2, \ldots$.‡

Each of the pseudo-differential operators T_{a_j} will now be written in its singular integral form (22):

$$(T_{a_j}f)(x) = \int k_j(x,z) f(x-z)\, dz;$$

but, since the a_j have compact ξ-support and are smooth, the kernels k_j will also be smooth, and the integrals above will converge for all x. The kernels k_j are given by

$$k_j(x,z) = \int a_j(x,\xi) e^{2\pi i \xi \cdot z}\, d\xi,$$

and one has the following estimates for them.

‡ Notice, incidentally, that the symbols $\delta(2^{-j}\xi)$ are in S^0 uniformly in j.

LEMMA. *Suppose a belongs to the symbol class S^m. Then*

$$|\partial_x^\beta \partial_z^\alpha k_j(x,z)| \le A_{M,\alpha,\beta} \cdot |z|^{-M} \cdot 2^{j[n+m-M+|\alpha|]}, \tag{28}$$

for all α, β, and $M \ge 0$, where the bound $A_{M,\alpha,\beta}$ is independent of $j \ge 0$.

Observe that

$$(-2\pi i z)^\gamma \partial_x^\beta \partial_z^\alpha [k_j(x,z)] = \int \partial_\xi^\gamma [(2\pi i \xi)^\alpha \partial_x^\beta a_j(x,\xi)] e^{2\pi i \xi\cdot z}\, d\xi.$$

We now make the most direct estimates on the above integral. First, the integrand is supported in the ball $|\xi| \le 2^{j+1}$, which has volume bounded by a multiple of 2^{nj}. Second, since the support is also limited by $2^{j-1} \le |\xi|$ (when $j \ne 0$), the differential estimates (2) satisfied by the symbols, and the corresponding estimates for the factors $\delta(2^{-j}\xi)$, show that the integrand is bounded by a multiple of $2^{j[m+|\alpha|-|\gamma|]}$. The result of this is

$$|z^\gamma \partial_x^\beta \partial_z^\alpha k_j(x,z)| \le A_{M,\alpha,\beta} \cdot 2^{j[n+m-M+|\alpha|]}, \qquad \text{whenever } |\gamma| = M.$$

Taking the supremum over all γ, $|\gamma| = M$, gives (28), and proves the lemma.

4.3 We now finish the proof of Proposition 1.

Since $k(x,\cdot) = \sum_{j=0}^\infty k_j(x,\cdot)$, with the sum converging for every x (in the sense of distributions), it will suffice to show that

$$\sum_{j=0}^\infty |\partial_x^\beta \partial_z^\alpha k_j(x,z)|$$

satisfies the estimates given by the right side of (23).

Let us first consider the case when $0 < |z| \le 1$. We break the above sum into two parts: the first where $2^j \le |z|^{-1}$, the second where $2^j > |z|^{-1}$. For the first sum we use the estimate (28) with $M = 0$, which is then majorized by a multiple of

$$\sum_{2^j \le |z|^{-1}} 2^{j(n+m+|\alpha|)}.$$

This in turn is

$$O(|z|^{-n-m-|\alpha|}) \qquad \text{when } n+m+|\alpha| > 0,$$

or

$$O(\log(|z|^{-1}) + 1) \qquad \text{when } n+m+|\alpha| \le 0.$$

In either case we get the estimate

$$O(|z|^{-n-m-|\alpha|-N}),$$

under the restrictions that $|z| \le 1$, $N \ge 0$, and $n+m+|\alpha|+N \ge 0$.

Next, for the second sum, we choose $M > n+m+|\alpha|$. Because of (28), we get the estimate

$$O(|z|^{-M}) \sum_{2^j > |z|^{-1}} 2^{j[n+m+|\alpha|-M]} = O(|z|^{-n-m-|\alpha|}).$$

The last term is certainly $O(|z|^{-n-m-|\alpha|-N})$, if $N \ge 0$, since $|z| \le 1$.

Finally, we consider the situation when $|z| \ge 1$. If M is sufficiently large (in particular if $M > n+m+|\alpha|+N$), then (28) shows that the sum is majorized by $O(|z|^{-M})$, which is $O(|z|^{-n-m-|\alpha|-N})$ for every N, since $|z| \ge 1$. The proof of the proposition is therefore concluded.

Remarks. (i) The proof of (23) shows that if $m < 0$, then for each x, the distribution $k(x,\cdot)$ is actually a locally integrable function.

(ii) Proposition 1 has a converse, allowing us to characterize the kernels for pseudo-differential operators with symbols in S^m; see §7.4.

4.4 Multipliers in $\mathbf{R}^n$ and kernels. We shall now deal with the connected problem of estimating the kernels of the multiplier operators that arose in Chapter 1, §6.2. These are the versions of the singular integral operators of that chapter that are related to the pseudo-differential operators of order 0 studied above, and are translation-invariant. The arguments needed are closely related to those just given but, in contrast with the symbol $a(x,\xi)$, our multiplier $m(\xi)$ is now allowed to have a singularity at $\xi = 0$ (of the kind consistent with scale invariance).

In keeping with the notations of Chapter 1, we denote by K the distribution whose Fourier transform is m. Our assumptions on m are as follows: we suppose that it is a bounded function that is in $C^\infty(\mathbf{R}^n \backslash \{0\})$ and satisfies

$$|\partial_\xi^\alpha m(\xi)| \le A_\alpha |\xi|^{-|\alpha|}, \qquad \text{for all } \alpha. \tag{29a}$$

Alternatively, we suppose only that $m \in C^\ell(\mathbf{R}^n \backslash \{0\})$ and satisfies

$$|\partial_\xi^\alpha m(\xi)| \le A_\alpha |\xi|^{-|\alpha|}, \qquad \text{for } 0 \le |\alpha| \le \ell, \tag{29b}$$

where ℓ is the smallest integer $> n/2$.

PROPOSITION 2. (a) *Under the assumption* (29a), *we have that K agrees with a function $K(x)$ away from the origin that is C^∞ there and satisfies*

$$|\partial_x^\alpha K(x)| \le A'_\alpha |x|^{-n-|\alpha|}, \qquad \text{all } \alpha. \tag{30a}$$

(b) *Under the assumption (29b), K agrees with a function $K(x)$ away from the origin that is locally integrable there and satisfies*

$$\int_{|x|\geq 2|y|} |K(x-y) - K(x)|\,dx \leq A, \qquad \text{all } y \neq 0. \tag{30b}$$

4.4.1 To prove (30a), we use the dyadic decomposition (25) (instead of (24)), and write

$$m(\xi) = \sum_{j=-\infty}^{\infty} m(\xi)\delta(2^{-j}\xi) = \sum_{j=-\infty}^{\infty} m_j(\xi).$$

We set

$$K_j(x) = \int e^{2\pi i x\cdot\xi} m_j(\xi)\,d\xi.$$

Since $\sum m_j(\xi)$ converges to $m(\xi)$ for $\xi \neq 0$, and does so boundedly, it follows that $\sum K_j$ converges to K in the sense of distributions. So it will suffice to estimate $\sum_j |\partial_x^\alpha K_j(x)|$ for $x \neq 0$.

One sees easily (as in (28)) that

$$|\partial_x^\alpha K_j(x)| \leq A_{M,\alpha} \cdot |x|^{-M} \cdot 2^{j[n-M+|\alpha|]}, \qquad \text{all } j, \text{ all } M \geq 0. \tag{31}$$

Using (31) for $M = 0$ gives

$$\sum_{2^j \leq |x|^{-1}} |\partial_x^\alpha K_j(x)| \leq A_\alpha \sum_{2^j \leq |x|^{-1}} 2^{j[n+|\alpha|]} \leq A'_\alpha |x|^{-n-|\alpha|}.$$

Similarly, with $M > n + \alpha$, one gets

$$\sum_{2^j > |x|^{-1}} |\partial_x^\alpha K_j(x)| \leq A_{M,\alpha}\,|x|^{-M} \sum_{2^j > |x|^{-1}} 2^{j[n+|\alpha|-M]} \leq A'_\alpha\,|x|^{-n-|\alpha|},$$

and (30a) is therefore established.

4.4.2 Turning to the second part of Proposition 2, we shall observe that the application of Plancherel's theorem, instead of the crude estimates for the Fourier transform we have used, will also allow us to deal with the K_j, while not requiring as much regularity for m.

First, one notes that

$$\int |(-2\pi i x)^\gamma K_j(x)|^2\,dx = \int |\partial_\xi^\gamma m_j(\xi)|^2\,d\xi, \qquad |\gamma| = M,$$

which implies

$$\int |\,|x|^M K_j(x)|^2\,dx \leq A 2^{nj} \cdot 2^{-2jM}, \tag{32}$$

for $0 \leq M \leq l$, on the basis of assumption (29b). From this one gets

$$\int_{|x|\leq a} |K_j(x)|\,dx \leq A\,2^{jn/2} a^{n/2}, \qquad a > 0, \tag{33}$$

by Schwarz's inequality:

$$\int_{|x|\leq a} |K_j(x)|\,dx \leq \left(\int |K_j(x)|^2\,dx\right)^{1/2} \left(\int_{|x|\leq a} dx\right)^{1/2},$$

using (32) with $M = 0$. Similarly we have

$$\int_{|x|\geq a} |K_j(x)|\,dx \leq A\,2^{jn/2} a^{n/2} 2^{-j\ell} a^{-\ell}, \qquad a > 0, \tag{34}$$

if we take $M = \ell =$ smallest integer $> n/2$ in (32). Combining these two, with $a = 2^{-j}$, yields

$$\int_{\mathbf{R}^n} |K_j(x)|\,dx \leq A, \qquad \text{all } j. \tag{35}$$

In the same way, we can also prove that

$$\int |\partial_x^\alpha K_j(x)|\,dx \leq A_\alpha 2^{j|\alpha|}, \qquad \text{all } j \text{ and } \alpha;$$

and what is essentially the same as this when $|\alpha| = 1$,

$$\int_{\mathbf{R}^n} |K_j(x+h) - K_j(x)|\,dx \leq A|h|2^j, \qquad \text{all } h, j. \tag{36}$$

Now write

$$\sum_j \int_{|x|\geq 2|y|} |K_j(x-y) - K_j(x)|\,dx = \sum\nolimits_1 + \sum\nolimits_2,$$

where $\sum_1$ is taken over the j with $2^j \leq |y|^{-1}$, and $\sum_2$ is taken over the j with $2^j > |y|^{-1}$. For $\sum_1$ we estimate the summands by

$$\int_{\mathbf{R}^n} |K_j(x-y) - K_j(x)|\,dx,$$

which by (36) gives us the estimate

$$A|y| \sum_{2^j \leq |y|^{-1}} 2^j \leq A'.$$

In the second sum, we majorize the summands by

$$2\int_{|x|\geq|y|} |K_j(x)|\,dx,$$

which, by (34) and the fact that $\ell > n/2$, leads to the estimate

$$A|y|^{n/2-\ell} \sum_{2^j > |y|^{-1}} 2^{j(n/2-\ell)} \leq A'$$

for $\sum_2$. This completes the proof of Proposition 2.

4.5 We now consider a converse question. Given a distribution K that is a function away from the origin that satisfies the differential inequalities (30a), we wish to find the additional conditions that need to be imposed on K so that $\widehat{K}$ is a bounded function, i.e., so that $f \mapsto f * K$ is a bounded operator on $L^2(\mathbf{R}^n)$. That additional restrictions ("cancellation conditions") must be required can be seen by the example $K(x) = |x|^{-n}$. This function satisfies the inequalities (30a), yet there is no distribution agreeing with K away from the origin whose Fourier transform is bounded (see Chapter 7, §3.1).

The result presented below, which is quite simple, will provide a clue to the more general problem of describing which "Calderón-Zygmund kernels" are associated to bounded operators on $L^2(\mathbf{R}^n)$; this latter point will be treated in Chapter 7. Also, a variant will allow us to prove a converse to Proposition 1, thereby specifying the kernels $k(x,z)$ that arise from symbols $a(x,\xi) \in S^m$ (see §7.4).

Our characterization needs the notion of a *standard* or *normalized* bump function, a notion already used in Chapter 3, §1.8. Suppose we fix a large integer N.[†] For a smooth function ϕ to be such a normalized bump function, we require that it be supported in the unit ball $|x| < 1$, and that it satisfy the inequalities

$$\left|\frac{\partial^\alpha \phi}{\partial x^\alpha}\right| \leq 1, \qquad 0 \leq |\alpha| \leq N. \tag{37}$$

Given ϕ, we denote by the ϕ^R the function defined by $\phi^R(x) = \phi(x/R)$, so that ϕ^R is associated with the ball of radius R. The additional conditions on K can then be expressed by testing the distribution K on such ϕ^R.

PROPOSITION 3. *Suppose K is a distribution that equals a function K away from the origin, with*

$$|\partial_x^\alpha K(x)| \leq A_1 |x|^{-n-|\alpha|}, \qquad \textit{for } |\alpha| \leq 1. \tag{38}$$

Then $\widehat{K}$ is a bounded function if and only if there is an A so that

$$|K(\phi^R)| \leq A, \tag{39}$$

for all normalized bump functions ϕ and all $R > 0$.

4.5.1 Let us assume that $\widehat{K}$ is a bounded function. Then, since the Fourier transform of ϕ^R is $R^n \cdot \widehat{\phi}^{R^{-1}}$, we see that

$$K(\phi^R) = \int \widehat{K}(\xi)\, \widehat{\phi}(-R\xi)\, R^n \, d\xi = \int \widehat{K}(-R^{-1}\xi)\, \widehat{\phi}(\xi)\, d\xi.$$

[†] The exact value of N is usually of no great importance; in the present context the condition $N > n/2$ will do.

Thus $|K(\phi^R)| \leq \|\widehat{K}\|_{L^\infty} \cdot \|\widehat{\phi}\|_{L^1} \leq A$, since (37) implies, by Plancherel's formula, that

$$\int (1 + |\xi|)^{2N} \, |\widehat{\phi}(\xi)|^2 \, d\xi \leq A',$$

which gives $\int |\widehat{\phi}(\xi)| \, d\xi \leq A'$.

To prove the converse, decompose K as $K = K^0 + K^\infty$, where $K^0 = K \cdot \eta$ and $K^\infty = K \cdot (1 - \eta)$; here η is a C^∞ function, supported in $|x| \leq 1$, that equals 1 near the origin. Now $\widehat{K} = \widehat{K^0} + \widehat{K^\infty}$, and both $\widehat{K^0}$ and $\widehat{K^\infty}$ are locally integrable functions: the first since it is the Fourier transform of a distribution of compact support (see also the proof of (42) below); the second because it is the Fourier transform of the L^2 function $K \cdot (1 - \eta)$. Thus $\widehat{K}$ is a function. We next claim that it will suffice to prove that

$$|\widehat{K}(\xi)| \leq B, \qquad \text{for } 1 \leq |\xi| \leq 2, \tag{40}$$

where B depends only on the bounds A_1 and A appearing in (38) and (39).

Indeed, if (40) holds, then we can replace the distribution K with K_ε for any $\varepsilon > 0$, where $K_\varepsilon(\phi) = K(\phi^{\varepsilon^{-1}})$. The function $K_\varepsilon(x)$ associated to K_ε then satisfies $K_\varepsilon(x) = \varepsilon^{-n} K(x/\varepsilon)$. Therefore, if K satisfies the hypotheses (38) and (39), then K_ε satisfies these hypotheses with the same bounds. Then (40) implies that

$$|\widehat{K}(\varepsilon\xi)| \leq B, \qquad \text{for } 1 \leq |\xi| \leq 2,$$

and any $\varepsilon > 0$, in other words that $\widehat{K}$ is bounded.

To establish (40), we again use the decomposition $K = K^0 + K^\infty$. Since

$$(2\pi i \xi_j)\widehat{K^\infty}(\xi) = \int \frac{\partial K^\infty}{\partial x_j} e^{-2\pi i x \cdot \xi} \, dx$$

and $|\partial K^\infty / \partial x_j| \leq A'(1 + |x|)^{-n-1}$, we get

$$|\widehat{K^\infty}(\xi)| \leq B, \qquad 1 \leq |\xi| \leq 2. \tag{41}$$

Next, since K^0 is a distribution of compact support,

$$\widehat{K^0}(\xi) = K^0(e^{-2\pi i x \cdot \xi}) = K(\psi),$$

where $\psi(x) = \eta(x) e^{-2\pi i x \cdot \xi}$. Now the ψ satisfy the conditions (37) for normalized bump functions when $1 \leq |\xi| \leq 2$, after multiplication by a suitable constant (which is bounded away from 0). Therefore (39), with $\delta = 1$, implies

$$|\widehat{K^0}(\xi)| \leq B, \qquad 1 \leq |\xi| \leq 2. \tag{42}$$

Putting this together with (41), and noting that the bound B depends only on the constants A_1 and A, we see that the proof of Proposition 3 is complete.

5. Estimates in L^p, Sobolev, and Lipschitz spaces

We now take up the regularity properties of our pseudo-differential operators as expressed in terms of the standard function spaces. We begin with the L^p boundedness of an operator of order 0.

5.1 L^p estimates. Suppose a belongs to the symbol class S^0. Then, as we saw in §2.2 and §4, we can express $T = T_a$ as

$$(Tf)(x) = \int K(x,y)f(y)\,dy = \int k(x, x-y)f(y)\,dy. \tag{43}$$

By Proposition 1 (inequality (23)),

$$|K(x,y)| \leq A|x-y|^{-n},$$

so that the integral (43) converges whenever $f \in \mathcal{S}$ and x is away from the support of f. Since we know that T is bounded on $L^2(\mathbf{R}^n)$, this representation extends to all $f \in L^2(\mathbf{R}^n)$ for almost every $x \notin \operatorname{supp} f$.

More generally, Proposition 1 says that

$$|\partial_x^\alpha \partial_y^\beta K(x,y)| \leq A_{\alpha,\beta}|x-y|^{-n-|\alpha|-|\beta|}; \tag{44}$$

hence K satisfies

$$\int_{|x-y|\geq 2\delta} |K(x,y) - K(x,\bar{y})|\,dx \leq A, \qquad \text{if } |y-\bar{y}| \leq \delta, \text{ all } \delta > 0, \tag{45}$$

as well as the inequality symmetric to (45), in which the roles of x and y are reversed.

We are now in a position to invoke the general theory of singular integrals developed in Chapter 1. In the present context we use the usual structure on $\mathbf{R}^n$: the balls $B(x,\delta)$ are $\{y : |x-y| < \delta\}$ and the quasi-distance ρ is given by $\rho(x,y) = |x-y|$. Here $|\cdot|$ denotes the standard Euclidean norm; also $d\mu(x)$ is the usual Lebesgue measure dx.

The basic estimates (9) (for $q = 2$) and (10) in Chapter 1 §5 are satisfied for our T. Moreover, the smoothness condition (29) (of Chapter 1, §7) also holds. As a consequence all the relevant results in Chapter 1 apply to T, in particular Corollary 1 in §7.1, Proposition 2 in §7.3, and the concluding remarks in §7.4. We summarize the main consequences as follows:

PROPOSITION 4. *Suppose T_a is the pseudo-differential operator corresponding to a symbol a in S^0. Then T_a extends to a bounded operator on $L^p(\mathbf{R}^n)$ to itself, for $1 < p < \infty$.*

Note that the definition of T_a on all of L^p, given in Chapter 1, which in effect is an extension of the restriction of T_a to $L^2 \cap L^p$, is consistent with the extension of T_a to distributions (which was given in §1.4 of the present chapter).

5.2 Sobolev spaces. We first recall the definition of the Sobolev spaces L_k^p, where k is a positive integer. A function f belongs to $L_k^p(\mathbf{R}^n)$ if $f \in L^p(\mathbf{R}^n)$ and the partial derivatives $\partial_x^\alpha f$, taken in the sense of distributions, belong to $L^p(\mathbf{R}^n)$, whenever $0 \leq |\alpha| \leq k$. The norm in L_k^p is given by

$$\|f\|_{L_k^p} = \sum_{|\alpha|\leq k} \|\partial_x^\alpha f\|_{L^p}.$$

PROPOSITION 5. *Suppose T_a is a pseudo-differential operator whose symbol a belongs to S^m. If m is an integer and $k \geq m$, then T_a is a bounded mapping from L_k^p to L_{k-m}^p, whenever $1 < p < \infty$.*

The proof is a direct consequence of Proposition 4 (in §5.1 above), once we verify the operator identity

$$\left(\frac{\partial}{\partial x}\right)^\gamma T_a = \sum_{|\alpha|\leq k} T_{a_\alpha}\left(\frac{\partial}{\partial x}\right)^\alpha, \tag{46}$$

whenever $|\gamma| \leq k - m$, where the a_α are appropriate symbols of order 0.

Notice first that $\partial_x^\gamma T_a$ is an operator of the form $T_{\tilde{a}}$, corresponding to a symbol $\tilde{a} \in S^k$, since $a \in S^m$.

Next, observe that we can write

$$1 = \alpha_0(\xi) + \sum_{j=1}^n \alpha_j(\xi)\xi_j, \tag{47}$$

where $\alpha_0, \alpha_1, \ldots, \alpha_n$ are symbols in S^{-1}. In fact, if η is a smooth function of compact support that equals 1 near $\xi = 0$, then we can take

$$\alpha_0(\xi) = \eta(\xi), \qquad \alpha_j(\xi) = (1-\eta(\xi))\frac{\xi_j}{|\xi|^2},$$

which gives (47). Using this, we can set

$$\tilde{a}(x,\xi) = \tilde{a}(x,\xi)\Big[\alpha_0(\xi) + \sum_{j=1}^n \alpha_j(\xi)\xi_j\Big]^k,$$

which, upon combining terms, can be rewritten as

$$\tilde{a}(x,\xi) = \sum_{|\alpha|\leq k} a_\alpha(x,\xi)(2\pi i\xi)^\alpha.$$

This establishes (46) and proves the proposition.

5.2.1 The definition of the spaces L^p_k, $1 < p < \infty$, and the results of Proposition 5 can be extended to cover the case when k is an arbitrary real number. For this purpose we consider the one-parameter family of pseudo-differential operators $(I - \Delta)^{k/2}$, defined as the operators T_a, with $a(x,\xi) = (1 + 4\pi^2|\xi|^2)^{k/2}$. Observe that when k is an even integer, then $(I - \Delta)^{k/2}$ agrees with its usual definition as a partial differential operator. Now we say that a distribution f belongs to L^p_k if

$$(I - \Delta)^{k/2} f \in L^p(\mathbf{R}^n);$$

we take $\|f\|_{L^p_k} = \|(I - \Delta)^{k/2} f\|_{L^p}$.

If we apply the above proposition to the case $T_a = (I - \Delta)^{m/2}$, $k = m$, we get that

$$\|(I - \Delta)^{k/2} f\|_{L^p} \le A \sum_{|\alpha| \le k} \|\partial_x^\alpha f\|_{L^p}.$$

Similarly, applying the proposition with

$$T_a = \partial_x^\alpha (I - \Delta)^{-k/2}, \qquad |\alpha| \le k, \qquad m = 0,$$

gives us

$$\sum_{|\alpha| \le k} \|\partial_x^\alpha f\|_{L^p} \le A \|(I - \Delta)^{k/2} f\|_{L^p}.$$

Therefore the new definition just given for L^p_k is consistent with the previous definition, and the two norms are equivalent.

With this new definition of L^p_k, we obtain:

COROLLARY. *The results in the previous proposition remain valid for arbitrary real k and m.*

For the proof we need only the following substitute for (46):

$$(I - \Delta)^{(k-m)/2} T_a = T_{\tilde{a}} \cdot (I - \Delta)^{k/2},$$

where $\tilde{a}$ is a symbol in S^0, assuming that $a \in S^m$. In fact, by the symbolic calculus (Theorem 2 in §3)

$$(I - \Delta)^{(k-m)/2} T_a = T_c, \qquad \text{where } c \in S^k,$$

which determines $\tilde{a}$ by $c(x,\xi) = \tilde{a}(x,\xi)(1 + 4\pi^2|\xi|^2)^{k/2}$, and our assertion is established.

5.3 Lipschitz spaces. We now take up the Lipschitz (Hölder) space Λ_γ, $\gamma > 0$. According to the usual definition (for $0 < \gamma < 1$), a function f belongs to Λ_γ if there exists a constant A so that $|f(x)| \le A$ almost everywhere and

$$\sup_x |f(x - y) - f(x)| \le A|y|^\gamma$$

for all y.‡ For our purposes it will be useful to adopt as our definition an equivalent property, which can be stated as follows. We say that $f \in \Lambda_\gamma$ if there exists a constant A so that

$$\|f\|_{L^\infty} \le A \qquad \text{and} \qquad \|\Delta_j(f)\|_{L^\infty} \le A 2^{-j\gamma}, \quad j = 1, 2, \ldots. \tag{48}$$

Here the Δ_j are the "dyadic difference operators", given by

$$\Delta_j(f) = f * \Psi_{2^{-j}},$$

as in §4.1.

The smallest A in (48) can be taken to be the norm in the space Λ_γ. That this definition is equivalent to the usual one will be shown in §5.3.4 below. The continuity property of pseudo-differential operators in the Λ_γ spaces is contained in the following proposition.

PROPOSITION 6. *Suppose a is a symbol in S^m. Then the operator T_a is a bounded mapping from Λ_γ to $\Lambda_{\gamma - m}$, whenever $\gamma > m$.*

5.3.1 Proof. Suppose first that $f \in \Lambda_\gamma$. Then

$$f = S_0 f + \sum_{j=1}^\infty \Delta_j f = \sum_{j=0}^\infty f_j,$$

and $T_a = \sum_{j=0}^\infty T_{a_j}$, with $T_{a_0} = T_a S_0$ and $T_{a_j} = T_a \Delta_j$ for $j \ge 1$; all this is by (28) and the discussion in §4.2. Now the Δ_j are essentially projection operators—not in the strict sense, since $\Delta_j \ne \Delta_j^2$—but we do have that

$$\Delta_j = \Delta_j \cdot (\Delta_{j-1} + \Delta_j + \Delta_{j+1}), \tag{49}$$

in view of the support properties of the $\widehat{\Psi}(2^{-j}\xi) = \delta(2^{-j}\xi)$; here we use the convention that $\Delta_0 = S_0$ and $\Delta_{-1} = 0$.† From this it follows that

$$Tf = \sum_{j=0}^\infty T_{a_j} f'_j, \qquad \text{with } \|f'_j\|_{L^\infty} \le A 2^{-j\gamma}, \tag{50}$$

if $f'_j = (\Delta_{j-1} + \Delta_j + \Delta_{j+1}) f$.

‡ Note that the boundedness of f makes this inequality trivial for $|y| \ge 1$.

† Unlike the notation of (25a).

We now need the following observation, which is interesting in its own right.

LEMMA. *Suppose the symbol a belongs to S^m, and define $T_{a_j} = T_a\Delta_j$. Then, as operators from $L^\infty(\mathbf{R}^n)$ to itself, the T_{a_j} have norms that satisfy*

$$\|T_{a_j}\| \leq A2^{jm}. \tag{51}$$

To prove the lemma, recall that the T_{a_j} may be realized in terms of kernels k_j, with

$$(T_{a_j}f)(x) = \int k_j(x, x-z)\, f(z)\, dz,$$

where the k_j satisfy the estimates (28). According to these estimates

$$\int_{\mathbf{R}^n} |k_j(x, x-z)|\, dz = A\int_{|z|\leq 2^{-j}} 2^{j[n+m]}\, dz + \int_{|z|\geq 2^{-j}} |z|^{-n-1}2^{j[m+1]}\, dz,$$

the first coming from (28) for $M = 0$, and the second from the case $M = n+1$. If we make the obvious calculations we get

$$\int_{\mathbf{R}^n} |k_j(x, x-z)|\, dz \leq A2^{jm},$$

which proves (51). An immediate consequence of (51) is that the norm of the operator $\partial_x^\alpha T_{a_j}$, as a mapping from $L^\infty(\mathbf{R}^n)$ to itself, has an estimate

$$\|\partial_x^\alpha T_{a_j}\|_{L^\infty, L^\infty} \leq A_\alpha 2^{j|\alpha|}\cdot 2^{jm}. \tag{52}$$

This is because the operator $\partial_x^\alpha T_{a_j}$ has symbol

$$e^{-2\pi i x\cdot\xi}\partial_x^\alpha[a_j(x,\xi)e^{2\pi i x\cdot\xi}],$$

which belongs to $S^{m+|\alpha|}$.

5.3.2 Returning to (50), we see that

$$Tf = \sum T_{a_j}f_j' = \sum F_j,$$

where, because of (52),

$$\|\partial_x^\alpha F_j\|_{L^\infty} \leq A_\alpha 2^{j[|\alpha|+m-\gamma]}. \tag{53}$$

We shall show that this implies that $\sum F_j \in \Lambda_{\gamma-m}$.

In fact, using (53) with $\alpha = 0$ shows that $\sum F_j$ converges in the L^∞ norm, since $\gamma > m$, and therefore $\sum F_j$ satisfies the first condition in (48). Next we shall verify that

$$\|\Delta_j(\Sigma F_\ell)\|_{L^\infty} \leq A2^{-j(\gamma-m)}, \qquad j = 1, 2, \ldots. \tag{54}$$

It is useful to observe that we can write, for any fixed positive integer ℓ,

$$I = \sum_{|\alpha|\leq\ell} T^{(\alpha)}\partial_x^\alpha$$

where $T^{(\alpha)}$ are pseudo-differential operators, with symbols belonging to $S^{-\ell}$, and the symbols are independent of x. In fact this identity is an immediate consequence of (47) in the form

$$1 = \Big(\alpha_0(\xi) + \sum_{j=1}^m \alpha_j(\xi)\xi_j\Big)^\ell,$$

once we collect terms. The above identity allows us to write

$$\Delta_j(F_i) = \Delta_j\sum_\alpha T^{(\alpha)}(\partial_x^\alpha)F_i = \sum_{|\alpha|\leq\ell} T^{(\alpha)}\Delta_j(\partial_x^\alpha)F_i$$

since the $T^{(\alpha)}$ and Δ_j commute. We can now invoke (53) and (51) (with $T_{a_j} = T^{(\alpha)}\Delta_j$, $m = -\ell$) to get

$$\|\Delta_j(F_i)\|_{L^\infty} \leq A2^{-j\ell}\cdot 2^{i[\ell+m-\gamma]}.$$

When $i \leq j$, we use this with ℓ chosen to be the smallest integer $> \gamma - m$. This gives the estimate

$$\|\Delta_j(\sum_{i\leq j} F_i)\|_{L^\infty} \leq A2^{-j(\gamma-m)}.$$

Similarly, when $i > j$, we use the above with $\ell = 0$ and get

$$\|\Delta_j(\sum_{i>j} F_i)\|_{L^\infty} \leq A2^{-j(\gamma-m)}.$$

Combining these two majorizations yields (54) and concludes the proof of Proposition 6.

5.3.3 We shall now point out a very simple but useful alternative characterization of Λ_γ. This is in terms of approximations by smooth functions; it is also closely connected with the definition of Λ_γ spaces as intermediate spaces, using the "real" method of interpolation.‡

‡ For the latter, see also the references given in §7.8.

COROLLARY 1. *A function f belongs to Λ_γ if and only if there is a decomposition*

$$f = \sum_{j=0}^{\infty} f_j$$

with

$$\|\partial_x^\alpha f_j\|_{L^\infty} \leq A 2^{-j\gamma} \cdot 2^{j|\alpha|} \tag{55}$$

for all $0 \leq |\alpha| \leq \ell$, where ℓ is the smallest integer $> \gamma$.

When $f \in \Lambda_\gamma$, the argument proving (53), with $T_a = I$, $f_j = F_j = \Delta_j(f)$, gives the required estimate for the f_j. To get the converse, one need only use the argument that (53) implies (54), with $m = 0$, $F_j = f_j$.

A second consequence of Proposition 6 is the following:

COROLLARY 2. *The operator $(I - \Delta)^{m/2}$ gives an isomorphism from Λ_γ to $\Lambda_{\gamma-m}$, whenever $\gamma > m$.*

This is clear because $(I - \Delta)^{m/2}$ is continuous from Λ_γ to $\Lambda_{\gamma-m}$, and its inverse, $(I - \Delta)^{-m/2}$, is continuous from $\Lambda_{\gamma-m}$ to Λ_γ.

5.3.4 Next, we shall show that our definition (given at the beginning of §5.3.) of Λ_γ agrees with the usual definition when $0 < \gamma < 1$.

Suppose first that f is bounded and satisfies

$$\sup_x |f(x - y) - f(x)| \leq A|y|^\gamma.$$

Then

$$\begin{aligned}\Delta_j f(x) = (f * \Psi_{2^{-j}})(x) &= \int f(x - y)\Psi_{2^{-j}}(y)\, dy \\ &= \int [f(x - y) - f(x)]\Psi_{2^{-j}}(y)\, dy,\end{aligned}$$

since $\int \Psi(y)\, dy = 0$. Thus

$$\|\Delta_j(f)\|_{L^\infty} \leq A \int |y|^\gamma \cdot |\Psi_{2^{-j}}(y)|\, dy = A2^{-j\gamma},$$

and condition (48) holds.

Conversely, if (48) is satisfied, according to Corollary 1, we can write

$$f = \sum_{j=0}^{\infty} f_j,$$

with the f_j satisfying the differential inequalities given in (55). Using (55) with $\alpha = 0$, we see that $\sum f_j$ converges in L^∞ norm. Next, if $|y| \leq 1$,

$$f(x - y) - f(x) = \sum_{2^j \leq |y|^{-1}} [f_j(x - y) - f_j(x)] + \sum_{2^j > |y|^{-1}} [f_j(x - y) - f_j(x)].$$

Using the mean-value theorem and (55) for $|\alpha| = 1$, we see that the terms in the first sum are dominated by $A|y|2^{-j\gamma}2^j$, resulting in the estimate

$$A|y| \sum_{2^j \leq |y|^{-1}} 2^{j(1-\gamma)} \leq A|y|^\gamma,$$

since $\gamma < 1$. The terms in the second sum are majorized by $2A2^{-\gamma j}$, and this gives $O(|y|^\gamma)$ as an estimate, since $\gamma > 0$. Thus

$$\sup_x |f(x - y) - f(x)| \leq A|y|^\gamma,$$

and the equivalence of the two definitions is established.

Finally, it is also known that the operators $(I - \Delta)^{m/2}$ are isomorphisms from Λ_γ to $\Lambda_{\gamma-m}$, when the two spaces are given their usual definitions.† Thus the identification of the present Λ_γ spaces with the usual Λ_γ spaces is established.

5.4 Almost orthogonality. Returning to the observation (51), notice that the same proof shows that the operators T_{a_j} have norms bounded by $A2^{mj}$, not only when mapping $L^\infty(\mathbf{R}^n)$ to itself, but also when mapping $L^p(\mathbf{R}^n)$ to itself, for $1 \leq p \leq \infty$. In particular, if we take $m = 0$ and $p = 2$, we see that whenever $a \in S^0$, then, in

$$T_a = \sum_j T_{a_j},$$

the summands are uniformly bounded operators on $L^2(\mathbf{R}^n)$. That the sum is also bounded is then in essence a consequence of the fact that these summands are, in a sense, mutually almost orthogonal. This vague statement can be made precise; moreover its further development and extension is an important principle that is one of the main themes of the next chapter.

† See *Singular Integrals*, Chapter 5.

6. Appendix: Compound symbols

6.1 Another useful device in the study of pseudo-differential operators is the notion of a compound symbol. It has the advantage of combining the idea of a symbol (and its ordering, which puts x-multiplications to the left of ξ-multiplications), together with the reverse order, which appears when taking adjoints.[‡]

A function $c = c(x, y, \xi)$ that is smooth on $\mathbf{R}^n \times \mathbf{R}^n \times \mathbf{R}^n$ will be said to be a *compound symbol* of order m if it satisfies the analogue of (2), namely

$$|\partial_y^\gamma \partial_x^\beta \partial_\xi^\alpha c(x, y, \xi)| \leq A_{\alpha,\beta,\gamma}(1 + |\xi|)^{m-|\alpha|}. \tag{56}$$

To each such c, we assign the operator $T_{[c]}$, defined formally by

$$T_{[c]}(f)(x) = \int c(x, y, \xi)\, e^{2\pi i \xi\cdot(x-y)} f(y)\, dy\, d\xi. \tag{57}$$

This is suggested by (3); like that identity, it can be made precise by replacing $c(x, y, \xi)$ with

$$c_\varepsilon(x, y, \xi) = c(x, y, \xi)\gamma(\varepsilon y, \varepsilon\xi),$$

(where γ is smooth, has compact support, and $\gamma(0, 0) = 1$), and letting $\varepsilon \to 0$. The result is that whenever $f \in \mathcal{S}$, then $T_{[c_\varepsilon]}f$ converges in $\mathcal{S}$ as $\varepsilon \to 0$, and the limiting operator, denoted by $T_{[c]}$, is continuous from $\mathcal{S}$ to itself.

A pseudo-differential operator may have many different representations via compound symbols in the form (57), but only one representation (1), in terms of a usual symbol. This is made precise by the following:

PROPOSITION. *Suppose c is a compound symbol of order m. Then there exists an $a \in S^m$, so that $T_a = T_{[c]}$. Moreover, the asymptotic expansion for a is given by*

$$a(x, \xi) - \sum_{|\alpha|<N} \frac{(2\pi i)^{-|\alpha|}}{\alpha!} \partial_\xi^\alpha \partial_y^\alpha c(x, y, \xi)\big|_{x=y} \in S^{m-N}, \tag{58}$$

which holds for any $N \geq 0$.

The proof of this proposition is, to a large extent, a reprise of the proof of Theorem 2 in §3. First, one shows (formally) that, in analogy with (14),

$$a(x, \xi) = \int c(x, y, \eta)\, e^{2\pi i(x-y)\cdot(\eta-\xi)}\, dy\, d\eta \tag{59}$$

and also, parallel to (15), that

$$a(x, \xi) = \int \widehat{c}(x, \eta, \xi + \eta)\, e^{2\pi i x\cdot\eta}\, d\eta, \tag{60}$$

[‡] Another symmetrization of the roles of x and ξ occurs in the "Weyl calculus", which will be taken up in the context of the Heisenberg group in Chapter 12.

where $\widehat{c}$ denotes the Fourier transform of $c(x, y, \xi)$ with respect to the y-variable. Assuming first that $c(x, y, \xi)$ has compact support in the y-variable, we exploit (60) via Taylor's formula

$$\widehat{c}(x, \eta, \xi + \eta) = \sum_{|\alpha|<N} \frac{1}{\alpha!} \partial_\xi^\alpha \widehat{c}(x, \eta, \xi)\eta^\alpha + R_N(x, \eta, \xi),$$

where

$$|R_N(x, \eta, \xi)| \leq A|\eta|^N(1 + |\eta|)^{-M}(1 + |\xi|)^{m-N}$$

with M large, if $|\xi| \geq 2|\eta|$. Also

$$|R_N(x, \eta, \xi)| \leq A|\eta|^N(1 + |\eta|)^{-M},$$

which holds for N sufficiently large, and M arbitrary, and which is used when $|\xi| \leq 2|\eta|$. Inserting the above in (60) gives the desired estimates for (58), and proves the asymptotic formula under the assumption that $c(x, y, \xi)$ has compact y-support.

To pass to the general case, we argue as in §3.3, by reducing matters to the situation when $c(x, y, \xi)$ vanishes for y near x. Then we can exploit formula (59) and integrate by parts in the y and η variables, as in the argument following (21). The resulting integral is then estimated by

$$A_{N_1,N_2} \int (1 + |\eta|)^{m-2N_1}(1 + |\eta - \xi|)^{-2N_2}(1 + |x - y|)^{-2N_1}\, dy\, d\eta,$$

which is $O((1 + |\xi|)^{m-N})$, as soon as we choose N_1 and N_2 sufficiently large. Thus the contribution of $c(x, y, \xi)$ vanishing for y near x is contained in the remainder term of (58), and the proof is concluded.

We should note in this connection that a pseudo-differential operator determines its symbol uniquely; see §7.1.

6.2 Adjoints. An immediate application of the above is that the class of pseudo-differential operators with symbols in S^m is closed under taking adjoints.

PROPOSITION. *Suppose $a \in S^m$. Then there is an $a^* \in S^m$ so that $T_a^* = T_{a^*}$, where T^* is defined as in (4). Moreover, for every $N \geq 0$, we have that*

$$a^*(x, \xi) - \sum_{|\alpha|<N} \frac{(2\pi i)^{-|\alpha|}}{\alpha!} \partial_\xi^\alpha \partial_x^\alpha \bar{a}(x, \xi) \in S^{m-N}.$$

In fact, from (5) it follows that $T_a^* = T_{[c]}$, where $c(x, y, \xi) = \bar{a}(y, \xi)$, which, by the previous proposition, implies our assertion.

6.3 Change of variables. An important fact about our pseudo-differential operators is that this class is invariant (at least locally) under smooth changes of variable. This can be made precise as follows.

Suppose O and $\tilde{O}$ are a pair of bounded open sets in $\mathbf{R}^n$, and that we are given a diffeomorphism τ from $\tilde{O}$ onto O. For each smooth function f having compact support in $\tilde{O}$, define $\tau(f)$ by $(\tau f)(x) = f(\tau^{-1}(x))$; thus the support of τf is a compact subset of O. Now let $a(x,\xi)$ be a symbol in S^m, which we assume has compact x-support in O. Then there is a $b \in S^m$, with compact x-support in $\tilde{O}$, so that $\tau^{-1}T_a\tau = T_b$, and more precisely, $(\tau^{-1}T_a\tau)f = T_b f$, whenever f has compact support in $\tilde{O}$. Moreover, the following transformation formula holds:

PROPOSITION. $b(x,\xi) = a(\tau(x), [\partial\tau/\partial x]'\xi)$ modulo S^{m-1}.

Here $[\partial\tau/\partial x]'$ denotes the transpose inverse of the Jacobian matrix of the mapping τ.

To prove the proposition, we observe first that

$$(\tau^{-1}T_a\tau)f = \int\int e^{2\pi i(\tau(x)-y)\cdot\xi} a(\tau(x),\xi)\, f(\tau^{-1}(y))\, dy\, d\xi.$$

Make the change of variables $y = \tau(u)$, $dy = |\partial\tau/\partial u|\, du$. Then the integral becomes (replacing u by y)

$$\int e^{2\pi i(\tau(x)-\tau(y))\cdot\xi} a(\tau(x),\xi)\, f(y) \left|\frac{\partial\tau}{\partial y}\right| dy\, d\xi. \tag{61}$$

In the above, insert a cut-off function $\eta(x,y)$ for x near y, splitting (61) into a part supported for x near y, and a remaining part, supported for x away from y. The remaining part corresponds to the operator whose kernel is C^∞ in both variables, as repeated integration by parts in ξ shows, and thus gives an operator that belongs to S^{-N}, for every $N \geq 0$.[†]

To analyze the main part of (61), we observe that when y is near x, we can write $\tau(x) - \tau(y) = L_{x,y}(x-y)$, where for each (x,y), $L_{x,y}$ is an invertible *linear* transformation, depending smoothly on x and y. If we insert this in (61), make the linear change of ξ-variables indicated by $L_{x,y}$, and note that $L_{x,x} = \partial\tau/\partial x$, we see that (61) equals

$$\int e^{2\pi i(x-y)\cdot\xi}\, a[\tau(x), L'_{x,y}(\xi)]\, \eta(x,y) \left|\frac{\partial\tau}{\partial y}\right| |L^{-1}_{x,y}|\, dy\, d\xi.$$

Thus we have an operator with compound symbol c,

$$c(x,y,\xi) = a[\tau(x), L'_{x,y}(\xi)]\, \eta(x,y) \left|\frac{\partial\tau}{\partial y}\right| |L^{-1}_{x,y}|,$$

and the proposition follows as a consequence of the asymptotic formula (58), for $N = 1$.

If we define the "principal symbol" corresponding to an $a \in S^m$ as its equivalence class modulo S^{m-1}, then the proposition ensures a natural transformation law of the principal symbol under a change of variables. This allows one to extend some of the notions of pseudo-differential operators, and their principal symbols, to manifolds.

† The characterization of operators whose symbols belong to $\cap_N S^{-N}$ is described in §7.2 below.

7. *Further results*

A. Symbols and operators

7.1 A pseudo-differential operator determines its symbol uniquely. Indeed, if $a \in S^m$, then for each η, $T_a(e^{2\pi i x\cdot\eta}) = a(x,\eta)\cdot e^{2\pi i x\cdot\eta}$. This can be given the following rigorous meaning: if $\gamma \in C_0^\infty$, $\gamma(0) = 1$, and we set $f_\varepsilon(x) = \gamma(\varepsilon x)e^{2\pi i x\cdot\eta}$, then $T_a(f_\varepsilon) \to a(x,\eta)\cdot e^{2\pi i x\cdot\eta}$ in the sense of distributions, as $\varepsilon \to 0$. Alternatively, the identity is valid if T_a is extended to be a mapping of tempered distributions, as in §1.4.

7.2 The correspondence $a \mapsto T_a$ has the following properties:

(i) We have $a \in \mathcal{S}(\mathbf{R}^n \times \mathbf{R}^n)$ if and only if there is a $K \in \mathcal{S}(\mathbf{R}^n \times \mathbf{R}^n)$ so that $(T_a f)(x) = \int_{\mathbf{R}^n} K(x,y)\, f(y)\, dy$.

(ii) The mapping $a \mapsto T_a$ extends to an isomorphism from $\mathcal{S}'(\mathbf{R}^n \times \mathbf{R}^n)$ to the continuous linear operators mapping $\mathcal{S}(\mathbf{R}^n)$ to $\mathcal{S}'(\mathbf{R}^n)$. In this sense, each such operator corresponds to a unique "generalized" symbol.

(iii) If we define $S^{-\infty} = \bigcap_m S^m$, then $a \in S^{-\infty}$ exactly when the kernel K belongs to $C^\infty(\mathbf{R}^n \times \mathbf{R}^n)$ and

$$\sup_{x,y}(1+|x-y|)^\gamma |\partial_x^\alpha \partial_y^\beta K(x,y)| < \infty$$

for all α, β, γ.

(iv) Alternatively, $T = T_a$ for some $a \in S^{-\infty}$ is and only if T is a continuous mapping from $L_k^2(\mathbf{R}^n)$ to $L_{k'}^2(\mathbf{R}^n)$ for all k, k', $-\infty < k < k' < \infty$.

To prove (i) and (ii), recall that $K(x,y) = k(x, x-y)$ with $\widehat{k}(x,\xi) = a(x,\xi)$, and use the Schwartz kernel theorem.

7.3 A commutator

$$[T_{a_1}, T_{a_2}] = T_{a_1}T_{a_2} - T_{a_2}T_{a_1}$$

of two pseudo-differential operators T_{a_1}, T_{a_2} has lower order than either of the terms $T_{a_1}T_{a_2}$ or $T_{a_2}T_{a_1}$. In particular, if $a \in S^m$, then $[T_a, (\partial/\partial x_j)]$ is the operator with symbol $-\partial a(x,\xi)/\partial x_j$, which belongs to S^m. Also, if M_j denotes the multiplication operator $f \mapsto x_j f$, then the commutator $[T_a, M_j]$ has symbol $-(2\pi i)^{-1}\partial a(x,\xi)/\partial\xi_j$, which has order $m-1$. Such commutation properties in fact characterize the symbol class S^m.

Indeed, let $C_{\partial/\partial x_j}[T] = [T, \partial/\partial x_j]$, and $C_{x_j}[T] = [T, M_j]$. Then one can assert: If $T : \mathcal{S}(\mathbf{R}^n) \to \mathcal{S}(\mathbf{R}^n)$ is linear, then $T = T_a$ for some $a \in S^m$ exactly when

$$C_{\partial/\partial x_{i_1}} \circ \cdots \circ C_{\partial/\partial x_{i_\ell}} \circ C_{x_{j_1}} \circ \cdots \circ C_{x_{j_r}}[T] : L_k^2 \to L_{k+r-m}^2$$

is bounded, for all ℓ, r, and k.

For this and other variants, see R. Beals [1977]; also Coifman and Y. Meyer [1978].

7.4 Suppose that $a \in S^m$. We have already seen that the operator T_a corresponds to the kernel $K(x, y) = k(x, x - y)$, where k is related to a by $k(x, \cdot)\hat{}(\xi) = a(x, \xi)$, and that k satisfies the differential inequalities (23). The converse is true in the following sense:

(a) Let $m < 0$. Suppose $k(x, z) \in C^\infty(\mathbf{R}^n \times \mathbf{R}^n \backslash \{0\})$ and that k satisfies the inequalities (23). Then there is a symbol $a \in S^m$ so that, whenever $f \in \mathcal{S}$, the integral $\int k(x, x - y) f(y)\, dy$ converges absolutely to $(T_a f)(x)$.

(b) Now take $m \geq 0$ and suppose that, for each x, we are given a distribution $k(x, \cdot)$ depending smoothly on x that, away from the origin, agrees with a C^∞ function $k(x, z)$ that satisfies the differential inequalities (23). We must also require the following (necessary) "cancellation" conditions: For each multi-index β, there is a constant A_β so that, whenever ϕ is a normalized bump function (see §4.5), we have

$$\sup_{0<R\leq 1} \sup_{x \in \mathbf{R}^n} R^m |\partial_x^\beta \langle k(x, \cdot), \phi^R \rangle| \leq A_\beta;$$

here, as before, we have written $\phi^R(x) = \phi(x/R)$. It follows from these assumptions that there is an $a \in S^m$ so that $(T_a f)(x) = [k(x, \cdot) * f(\cdot)](x)$, for all $f \in \mathcal{S}$.

The proof is an elaboration of the argument given in §4.5.

7.5 Symbols $a(x, \xi) \in S^m$ have an approximate homogeneity of degree m (in the ξ variable): If $a(x, \xi)$ vanishes for small ξ, then the symbols a_δ belong uniformly to S^m, for $0 < \delta < \infty$; here $a_\delta(x, \xi) = \delta^{-m} a(x, \delta\xi)$. There is a corresponding approximate homogeneity for the kernels $k(x, z)$. These approximate homogeneities can be made exact, in the translation-invariant case, by the following considerations.

Let K be a distribution on $\mathbf{R}^n$. Recalling the definition $\phi_t(x) = t^{-n}\phi(x/t)$, we say that K is *homogeneous of degree d* if $K(\phi_t) = t^d K(\phi)$ for all $\phi \in \mathcal{S}$, and all $t > 0$.

(a) K is homogeneous of degree d exactly when the Fourier transform $\widehat{K}$ is homogeneous of degree $-n - d$.

(b) If K satisfies (a), then K agrees with a C^∞ function away from the origin if and only if the same is true for $\widehat{K}$.

(c) Suppose K satisfies (a) and (b). Let $K_0(x)$ denote the restriction of K to $\mathbf{R}^n \backslash \{0\}$. Then K_0 is a C^∞ function that is homogeneous of degree d; i.e., $K_0(tx) = t^d K_0(x)$ for all $x \in \mathbf{R}^n$, $t > 0$. Moreover, when $d > -n$, K_0 is locally integrable and $K(\phi) = \int_{\mathbf{R}^n} K_0(x)\, \phi(x)\, dx$, for all $\phi \in \mathcal{S}$.

(d) Again suppose that (a) and (b) hold. If $d = -n$, then necessarily K_0 enjoys the cancellation property $\int_{|x|=1} K_0(x)\, dx = 0$. Moreover, $K = c\delta +$ p.v.K_0; i.e.,

$$K(\phi) = c\phi(0) + \text{p.v.} \int_{\mathbf{R}^n} K_0(x)\, \phi(x)\, dx, \qquad \phi \in \mathcal{S},$$

for some constant c. That is, K is a Calderón-Zygmund distribution of the type described in Chapter 1, §8.19.

Assertion (a) follows easily from the definitions. To prove (b), write $K = K_1 + K_\infty$, where K_1 has compact support and K_∞ vanishes near the origin. Then $\widehat{K_1}$ is automatically smooth. Moreover, for each α, the distribution $|x|^{2N} \partial_x^\alpha \widehat{K_\infty}$ is the Fourier transform of an L^1 function if N is a sufficiently large integer. Thus $\widehat{K_\infty}$ is C^∞ away from the origin. The cancellation condition in (d) is a consequence of that given by the proposition in §4.5. Similar results hold when the standard homogeneity is replaced by a homogeneity attached to nonisotropic dilations; for some references to the literature, see Chapter 1, §8.20. A detailed proof of (d) may be found in Folland and Stein [1974].

7.6 In §4.4 we treated multipliers m that satisfy the condition

$$|\partial_\xi^\alpha m(\xi)| \leq A_\alpha |\xi|^{-\alpha}, \qquad \text{for all } \alpha \text{ with } 0 \leq |\alpha| \leq \ell;$$

here ℓ is the smallest integer with $\ell > n/2$. Similar conclusions can be reached with somewhat weaker hypotheses:

(i) Assume instead that for all α, $|\alpha| \leq \ell$, there is an A_α with

$$\sup_{R>0} R^{-n+2|\alpha|} \int_{R<|\xi|<2R} |\partial_\xi^\alpha m(\xi)|^2\, d\xi \leq A_\alpha.$$

(ii) An even weaker hypothesis is as follows. Fix a function $\eta \in C_0^\infty(\mathbf{R}^n)$ with $\eta(\xi) = 1$ for $1 \leq |\xi| \leq 2$, and $\eta(\xi) = 0$ for $|\xi| \leq 1/2$, $|\xi| \geq 4$. Then we require that for some real $k > n/2$, $\eta(\xi)m(R\xi) \in L_k^2(\mathbf{R}^n)$ for each $R > 0$, with

$$\sup_{0<R<\infty} \|\eta(\xi)m(R\xi)\|_{L_k^2(d\xi)} \leq A < \infty.$$

Under either of these conditions, if we let K be the distribution given by $\widehat{K} = m$, then, away from the origin, K agrees with a locally integrable function and satisfies

$$\sup_{y \neq 0} \int_{|x| \geq 2|y|} |K(x - y) - K(x)|\, dx \leq A.$$

Moreover, m is bounded, so the assumptions in Chapter 1, §6.2 are satisfied. As a result, the mapping T, given by $\widehat{Tf}(\xi) = m(\xi)\widehat{f}(\xi)$, extends to a bounded operator from $L^p(\mathbf{R}^n)$ to itself when $1 < p < \infty$, and is of weak-type $(1, 1)$.

Multiplier theorems of this type go back to Marcinkiewicz and Mihlin; the version in (i) is due to Hörmander. An alternative approach, using square functions, is in *Singular Integrals*, Chapter 4, where references to the literature can be found.

B. Function spaces

7.7 (a) If $a \in S^0$, then the operator T_a maps the Hardy space $H^p(\mathbf{R}^n)$ to the local Hardy space $H^p_{\text{loc}}(\mathbf{R}^n)$‡ for all p, $0 < p < \infty$.

(b) In particular, such a T_a maps $H^1(\mathbf{R}^n)$ to $L^1(\mathbf{R}^n)$, and $L^\infty(\mathbf{R}^n)$ to $\text{BMO}(\mathbf{R}^n)$; see also Chapter 4, §6.3^a. A stronger version of the second assertion is as follows: Define $\text{BMO}_{\text{loc}}(\mathbf{R}^n)$ to be the set of locally integrable f for which

$$\sup_{\text{radius}(B)\le 1} \frac{1}{|B|}\int_B |f - f_B|\,dx \le A \qquad \text{and} \qquad \sup_{\text{radius}(B)>1} \frac{1}{|B|}\int_B |f|\,dx \le A,$$

for some constant A. Then T_a maps $\text{BMO}_{\text{loc}}(\mathbf{R}^n)$ to $\text{BMO}(\mathbf{R}^n)$.

For (a), see Goldberg [1979]; the assertion concerning BMO_{loc} is the dual of conclusion (a) for $p = 1$.

7.8 (a) The continuity properties of the operators T_a, $a \in S^m$, acting on the Λ_γ spaces studied in §5.3, extend to the more general spaces $\Lambda^{p,q}_\gamma$ treated in *Singular Integrals*, Chapter 5. Indeed, if $a \in S^m$, then T_a extends to a bounded mapping from $\Lambda^{p,q}_\gamma$ to $\Lambda^{p,q}_{\gamma-m}$, whenever $1 \le p, q \le \infty$. This follows (as in §5.3) from the following characterization of an $f \in \Lambda^{p,q}_\gamma$, when $\gamma > 0$: we have $\|f\|_{L^p} \le A$ and

$$\left(\sum_{j=1}^{\infty}(2^{j\gamma}\|\Delta_j f\|_{L^p})^q\right)^{1/q} \le A.$$

(b) There is a simple relation between the Sobolev space L^p_γ and $\Lambda^{p,q}_\gamma$ only when $p = 2$. In general, the most that we can say is that $L^p_\gamma \subset \Lambda^{p,\max(p,2)}_\gamma$ and $\Lambda^{p,\min(p,2)}_\gamma \subset L^p_\gamma$ for $1 < p < \infty$, and all γ (see *Singular Integrals*, Chapter 5).

The $\Lambda^{p,q}_\gamma$ were studied systematically in Taibleson [1964], [1965], [1966]. For the interpolation theory of these spaces, one may also consult Bennett and Sharpley [1988].

7.9 Another family of function spaces that contains the Hardy spaces H^p_{loc} and the Sobolev spaces L^p_k (but not, in general, the Lipschitz spaces $\Lambda^{p,q}_\gamma$) are the spaces $F^{p,q}_\gamma$ of Triebel and Lizorkin. Essentially, $f \in F^{p,q}_\gamma$ when

$$\left(|(S_0 f)(x)|^q + \sum_{j=1}^{\infty}[2^{j\gamma}(\Delta_j f)(x)]^q\right)^{1/q} \in L^p(\mathbf{R}^n).$$

Again, it can be shown that if $a \in S^m$, then $T_a : F^{p,q}_\gamma \to F^{p,q}_{\gamma-m}$.

For an exposition of the properties of these spaces, see Peetre [1976], Triebel [1983], Frazier, Jawerth, and G. Weiss [1991]. The action of pseudo-differential operators on these spaces is studied in, e.g., Frazier, Torres, and G. Weiss [1988], Torres [1992].

‡ See Chapter 3, §5.17 for a definition.

7.10 A significant refinement of the classical L^p spaces are the Lorentz spaces $L^{p,q}$. We have that $f \in L^{p,q}(\mathbf{R}^n)$ when

$$\left(\frac{q}{p}\int_0^\infty [t^{1/p} f^*(t)]^q \frac{dt}{t}\right)^{1/q} = \|f\|^*_{p,q} < \infty;$$

here f^* is the nonincreasing function on $(0,\infty)$ that is equimeasurable with $|f|$. When $p = q$, $\|f\|^*_{p,q} = \|f\|_{L^p}$ and $L^{p,p} = L^p$. When $q = \infty$, $L^{p,q}$ is the "weak-type" L^p space, i.e., $f \in L^{p,\infty}$ if and only if $|\{x : |f(x)| > \alpha\}| \le A\alpha^{-p}$, for all $\alpha > 0$. One always has $L^{p,q_1} \subset L^{p,q_2}$ if $q_1 < q_2$; in particular, $L^{p,1} \subset L^{p,p} \subset L^{p,\infty}$ (when $p \ge 1$).

These spaces are important not only because of the ubiquitous nature of weak-type inequalities, but also because experience shows that sharp versions of results initially involving the L^p spaces are expressible in terms of the Lorentz spaces. We give two examples of this.

(a) The Hardy-Littlewood-Sobolev inequality can be stated as the inclusion

$$L^p_k \subset L^q, \qquad \text{where } q^{-1} = p^{-1} - kn^{-1} \text{ and } 1 < p < q < \infty;$$

here L^p_k are the Sobolev spaces defined in §5.2; see also Chapter 1, §8.21 and Chapter 8, §4.2. In its sharp form, the inclusion is $L^p_k \subset L^{q,p}$.

(b) Differentiability almost everywhere: Suppose F is a locally integrable function on $\mathbf{R}^n$, and $\nabla F \in L^{p,1}(\mathbf{R}^n)$ in the sense of distributions, with $p = n$. Then F can be redefined on a set of measure zero so that F is continuous on $\mathbf{R}^n$ and moreover, for almost all $x \in \mathbf{R}^n$, there is a vector $(\nabla F)(x)$ so that

$$F(x+h) - F(x) - \nabla F(x)\cdot h = o(|h|), \qquad \text{as } h \to 0.$$

For an exposition of the properties of the $L^{p,q}$ spaces, see *Fourier Analysis*, Chapter 5. Assertion (a) follows from the Marcinkiewicz interpolation theorem in the setting of these spaces. Assertion (b) is in Stein [1981].

C. Elliptic estimates

7.11 The asymptotic behavior of a symbol can be prescribed as follows. Suppose $a_j \in S^{m_j}$, $j = 0, 1, 2, \ldots$, with $m_0 > m_1 > m_2 > \cdots$ and $m_j \to -\infty$ as $j \to \infty$. Then there is a symbol $a \in S^{m_0}$ with $a \sim a_0 + a_1 + a_2 + \cdots$, in the sense that

$$a - \sum_{j=0}^{k} a_j \in S^{m_k}, \qquad k = 1, 2, \ldots.$$

Note that $a \sim 0$ according to this definition exactly when $a \in S^{-\infty}$.

The symbol a can be formed in much the same way as E. Borel constructed a C^∞ function with a preassigned Taylor development at a given point: We fix a C^∞ cut-off function η with $\eta(\xi) = 1$ when $|\xi| \ge 2$, and $\eta(\xi) = 0$

when $|\xi| \leq 1$. Then for an appropriate sequence ε_j, converging to zero sufficiently rapidly, we can take

$$a(x,\xi) = \sum_{j=0}^{\infty} a_j(x,\xi)\eta(\xi/\varepsilon_j).$$

In particular, it suffices to take ε_j so that

$$|\partial_x^\beta \partial_\xi^\alpha [a_j(x,\xi)\eta(\xi/\varepsilon_j)]| \leq 2^{-j}(1+|\xi|)^{m_j+1-|\alpha|},$$

when $|\alpha| \leq j$, $|\beta| \leq j$.

7.12 The inversion of elliptic operators via the symbolic calculus introduced in §3.5 can be put in the following more general form.

(i) Suppose $a \in S^m$ and that a is elliptic on an open set $U \subset \mathbf{R}^n$; that is, there is an $A > 0$ so that, for all $x \in U$, $|a(x,\xi)| \geq A|\xi|^m$ for large ξ. Suppose we are given another symbol $c \in S^{m'}$ so that $c(x,\xi)$ is supported for x in a compact subset of U. Then there exists a $b \in S^{m'-m}$ so that

$$T_b T_a + T_e = T_c$$

for some $e \in S^{-\infty}$.

(ii) As a result, if $L = \sum_{|\alpha|\leq m} a_\alpha(x)\partial_x^\alpha$ is an elliptic operator in U, and $\zeta_1 \prec \zeta_2 \prec \zeta_3$ are C_0^∞ cut-off functions,† then there is a $b \in S^{-m}$ so that

$$T_b(\zeta_2 L) = \zeta_1 I + E\zeta_3,$$

where the error E is an operator whose symbol is in $S^{-\infty}$.

In describing the construction of b, it is convenient to use the shorthand $a_1 \circ a_2 = a_3$ to mean $T_{a_1}T_{a_2} = T_{a_3}$. We determine $b \sim b_0 + b_1 + b_2 + \cdots$ as follows. Let $b_0 = c/a$. Then by Theorem 2, $b_0 \circ a = c + e_0$, with $e_0 \in S^{m'-1}$. By induction, suppose that $b_0, \ldots, b_{j-1}$ and $e_0, \ldots, e_{j-1}$ have been determined. Then let $b_j = e_{j-1}/a$ and set $e_j = b_j \circ a - b_j a$. Note that $b_j \in S^{-m+m'-j}$ and that $e_j \in S^{m'-j-1}$.

Systematic treatments of the symbolic calculus and its applications may be found in the monographs of Taylor [1981], Treves [1982], Hörmander [1985].

7.13 Suppose $L = \sum_{|\alpha|\leq m} a_\alpha(x)\partial_x^\alpha$ is an elliptic operator with smooth coefficients given in an open domain $U \subset \mathbf{R}^n$. As a consequence of the parametrix exhibited in §7.12, we can now describe the "interior" estimates valid for L: If $Lu = f$ in U and that f belongs locally to L_k^p, then u belongs locally to L_{k+m}^p.

To be more precise, let $\zeta_1, \zeta_2 \in C_0^\infty(U)$ with $\zeta_2 = 1$ on the support of ζ_1. Then

$$\|\zeta_1 u\|_{L_{k+m}^p} \leq A(\|\zeta_2 f\|_{L_k^p} + \|\zeta_2 u\|_{L^p}).$$

This is a consequence of §7.12 and the results in §5. Similar inequalities hold for the Λ_γ spaces, as well as for the function spaces arising in §7.8 and §7.9. In each case, there is a gain of m in the degree of smoothness (of u relative to f), which is characteristic of the elliptic situation.

† Here $\zeta_i \prec \zeta_{i+1}$ means that $\zeta_{i+1} = 1$ on the support of ζ_i.

D. Other topics

7.14 The dyadic decomposition of the frequency space of the Fourier transform (treated in §4.1) has a close connection with square functions; this relation was first stressed by Littlewood and Paley. We describe variants of square function inequalities (as in §6.3 and §8.23 of Chapter 1) appropriate to the present context.

(a) With Δ_k as in §4.1, one has

$$\Big\|\Big(\sum_k |\Delta_k f|^2\Big)^{1/2}\Big\|_{L^p} \approx \|f\|_{L^p}, \qquad 1 < p < \infty.$$

(b) As a consequence, suppose $\{f_k\}$ is a sequence of functions with $\widehat{f_k}$ supported in $a2^k \leq |\xi| \leq b2^k$, where $0 < a \leq b < \infty$ are fixed. Then

$$\Big\|\sum_k f_k\Big\|_{L^p} \leq \Big\|\Big(\sum_k |f_k|^2\Big)^{1/2}\Big\|_{L^p}.$$

The proof of the direct inequality in (a), namely

$$\Big\|\Big(\sum_k |\Delta_k f|^2\Big)^{1/2}\Big\|_{L^p} \leq A_p\|f\|_{L^p},$$

follows the lines set down in Chapter 1, §6.4 and §8.23. An alternate approach is to show that the linear operator

$$f \mapsto \sum_k \varepsilon_k \Delta_k(f)$$

is bounded on L^p, uniformly in all choices of $\varepsilon_k = \pm 1$, by using Chapter 2, §5.15(b). For this approach see, e.g., *Singular Integrals*, Chapter 4, §5. To prove the converse (i.e., (b)), one fixes $\ell \geq 1$ and defines

$$\tilde{\Delta}_k = \sum_{|j|\leq \ell} \Delta_{k-j}.$$

Then $\sum_k \Delta_j \tilde{\Delta}_k = I$, and one can argue as in §8.23 of Chapter 1, then choosing ℓ so that $2^{\ell-1} \geq b/a$.

7.15 We want to comment on the condition

$$|\partial_x^\beta \partial_\xi^\alpha a(x,\xi)| \leq A_{\alpha,\beta}|\xi|^{-\alpha},$$

as a simpler-looking version of the defining condition for the class S^0; here the factors $|\xi|^{-|\alpha|}$ replace $(1+|\xi|)^{-|\alpha|}$. A further motivation for considering this class comes from the conditions (29), where a is independent of x. Despite these considerations, neither L^2 boundedness holds for the corresponding operators, nor is there a symbolic calculus (i.e., the operators of "order 0" we have defined are not closed under composition). In fact:

(a) Suppose $\phi \in C_0^\infty$, $\phi(\xi) = 1$ when $1/2 \le |\xi| \le 2$, and ϕ vanishes near the origin. If

$$a(x,\xi) = \sum_{k\le 0} e^{2\pi i 2^k x}\,\phi(2^{-k}\xi),$$

then a satisfies the differential inequalities; but a variant of the construction in Chapter 7, §1.2 shows that the corresponding operator is unbounded.

(b) On $\mathbf{R}^1$, take $a_1(x,\xi) = \mathrm{sign}(\xi)$, $a_2(x,\xi) = e^{2\pi i x}$. Then

$$T_{a_1}T_{a_2} = T_a \qquad \text{with} \quad a(x,\xi) = e^{2\pi i x}\,\mathrm{sign}(\xi - 1),$$

and the symbol a does not satisfy the differential inequalities.

7.16 Nevertheless, there is a satisfactory theory if we assume that $a(x,\xi)$ is *homogeneous* in ξ of degree 0. More precisely:

(a) If we suppose $|\partial_\xi^\alpha a(x,\xi)| \le A_\alpha |\xi|^{-\alpha}$ for all α (which follows from the case $|\xi| = 1$), then T_a is bounded from $L^p(\mathbf{R}^n)$ to itself, $1 < p < \infty$.

(b) If we assume that $|\partial_x^\beta \partial_\xi^\alpha a_i(x,\xi)| \le A_{\alpha,\beta}|\xi|^{-\alpha}$ for all α, β (again with $a_i(x,\xi)$ homogeneous of degree 0 in ξ), then $T_{a_1} \circ T_{a_2} - T_{a_1\cdot a_2}$ is smoothing of order 1, in the sense that this difference maps L_k^p to L_{k+1}^p.

These results represent the symbolic calculus developed in Calderón and Zygmund [1957], and are proved by expanding $a(x,\xi)$ in spherical harmonics in ξ.

Notes

In the development of the theory, singular integrals came first. It was there that the notion of the symbol of an operator, as well as a related calculus, originated. See Mihlin [1948], Calderón and Zygmund [1957], Unterberger and Bokobza [1964], Seeley [1965]. An account of the early history of that aspect of singular integrals may be found in Seeley [1967]. The theory of pseudo-differential operators as such was hinted at by these developments but was formalized in Kohn and Nirenberg [1965], Hörmander [1965], Unterberger and Bokobza [1965], Systematic treatments dealing with applications to partial differential equations are given in Taylor [1981], Treves [1982], Hörmander [1985]. The ideas related to compound symbols are due to Kuranishi; see the exposition in Nirenberg [1968].

A detailed account of the theory of Sobolev and Lipschitz spaces may be found in *Singular Integrals*, Chapter 5. See also Bergh and Löfström [1976], Triebel [1983], Bennett and Sharpley [1988].

CHAPTER VII

Pseudo-Differential and Singular Integral Operators: Almost Orthogonality

The two kinds of operators considered in the previous chapter, namely pseudo-differential operators and singular integrals, in fact represented essentially the same objects, but emphasized different aspects. The pseudo-differential realization, given in terms of the Fourier transform, was most convenient for the L^2 theory and the symbolic calculus, while the singular integral realization was most closely connected with the real variable theory of Chapter 1, L^p estimates, etc.

The purpose of the present chapter is to try to extend this theory. Here we immediately encounter two difficulties. First, the close link that existed between these two kinds of operators does not persist, and the generalizations of pseudo-differential operators and singular integral operators must be pursued separately. Second, and more importantly, the Fourier transform no longer suffices to give the L^2 theory, and must be supplanted by a more general approach.

The idea we shall use is to appeal to a fundamental principle that has many applications. It states (in a precise form) that an operator T is bounded on L^2 if it can be decomposed as a sum $T = \sum T_j$, in which the components T_j are uniformly bounded, and the different T_j are mutually "almost orthogonal". We shall formulate two versions of this principle in §2, one whose proof is simple, and the other which lies a little deeper.

Our first illustration of this method arises when we consider pseudo-differential operators associated to a general class of symbols, called $S^m_{\rho,\delta}$, whose members are characterized by the differential inequalities

$$|\partial_x^\beta \partial_\xi^\alpha a(x,\xi)| \le A_{\alpha,\beta}(1 + |\xi|)^{m-\rho|\alpha|+\delta|\beta|}$$

for some fixed ρ and δ. Note that when $\rho = 1$ and $\delta = 0$, we have the class S^m studied in the previous chapter.

There are two points to be borne in mind in connection with these symbol classes. The first, regarding $S^m_{1,1}$ with $m = 0$, is that it is the largest class whose corresponding operators T have kernel representations (in an appropriate sense)

$$(Tf)(x) = \int_{\mathbf{R}^n} K(x,y)\,f(y)\,dy,$$

where K satisfies inequalities of Calderón-Zygmund type required for the singular integrals treated in Chapter 1. For these operators, one has in fact that

$$|\partial_x^\alpha \partial_y^\beta K(x,y)| \le A_{\alpha,\beta}|x-y|^{-n-|\alpha|-|\beta|} \tag{1}$$

for all α and β. But, unfortunately, such operators are not (in general) bounded on L^2. On the other hand, for the operators corresponding to $S^0_{\rho,\rho}$ with $0 \le \rho < 1$, the almost-orthogonality principle applies. This is of interest since various operators arising in the study of subelliptic problems (e.g., the heat equation and certain operators on the Heisenberg group) are of this kind. However, the $S_{\rho,\rho}$ conditions are not well suited to L^p, $p \neq 2$, and in applications other specific properties must come into play; the reason for this is that the inequalities determining $S_{\rho,\rho}$, $0 \le \rho < 1$, do not by themselves give precise enough descriptions of the kernels of the corresponding operators.

The situation is much clearer when we focus on operators whose kernels satisfy (1), or variants of the same character that require limited smoothness.† The natural question that arises is as follows: For such a T, what are the additional conditions that must be imposed so that T has a bounded extension on L^2? The answer to this question, which has far-reaching consequences, is that both T and its formal adjoint T^* must be "bounded" when tested on very simple trial functions—the "bump" functions. A related assertion gives directly verifiable "cancellation" conditions that must hold on the kernel K in order to yield the same conclusion. These results are variants of the "$T(1)$" theorem, and are an outgrowth of the study of the Cauchy integral. Again, here the almost-orthogonality principle plays a key role in the proof of the theorem; but, as we shall see, ideas related to the square-function characterization of BMO also need to be used.

1. *Exotic and forbidden symbols*

1.1 We consider first the extension of the differential inequalities (2) for symbols given in the previous chapter that will still guarantee that the corresponding kernels satisfy the inequalities (1) above. The condition will be seen to be

$$|\partial_x^\beta \partial_\xi^\alpha a(x,\xi)| \le A_{\alpha,\beta}(1+|\xi|)^{-|\alpha|+|\beta|}, \tag{2}$$

which is weaker than the requirement that $a \in S^0$, since here each x-differentiation allows a further growth factor of $1+|\xi|$. We generalize the

† When properly defined, the *class* of such operators is both translation and scale invariant.

inequality (2), and say that a function a belongs to the symbol class $S^m_{\rho,\delta}$ if $a = a(x,\xi)$ is smooth for $(x,\xi) \in \mathbf{R}^n \times \mathbf{R}^n$ and

$$|\partial_x^\beta \partial_\xi^\alpha a(x,\xi)| \le A_{\alpha,\beta}(1+|\xi|)^{m-\rho|\alpha|+\delta|\beta|}. \tag{3}$$

Note that $S^m_{1,0}$ is our previously studied standard class S^m; also, (2) is the condition that a belongs to $S^0_{1,1}$.

We summarize our first observations as follows:

PROPOSITION 1. *Let*

$$(T_a f)(x) = \int_{\mathbf{R}^n} a(x,\xi)\,\hat{f}(\xi)\,e^{2\pi i x\cdot\xi}\,d\xi, \qquad \textit{for } f \in \mathcal{S},$$

where $a \in S^0_{1,1}$. Then there is a kernel K satisfying (1) *so that*

$$(T_a f)(x) = \int K(x,y)\,f(y)\,dy,$$

whenever x is outside the support of f.

1.1.1 To prove the proposition, we should remark first that the integral defining $T_a(f)$ converges for each $x \in \mathbf{R}^n$, no matter which $S^m_{\rho,\delta}$ class a belongs to, provided that $f \in \mathcal{S}$. It should also be observed that the argument given in Chapter 6, §1.3 goes through unchanged to show that the operator T_a maps $\mathcal{S}$ to $\mathcal{S}$ continuously.

Next, write

$$T_a = \sum_{j=0}^{\infty} T_{a_j},$$

with $T_{a_0} = TS_0$ and $T_{a_j} = T_a\Delta_j$ for $j \ge 1$; then $a_0(x,\xi) = a(x,\xi)\hat{\Phi}(\xi)$ and

$$a_j(x,\xi) = a(x,\xi)\hat{\Psi}(2^{-j}\xi), \quad j \ge 1,$$

as in §4.2 of the previous chapter.

Now $T_{a_j}f(x) = \int k_j(x,z)\,f(x-z)\,dz$, and for k_j we can make the estimates

$$|\partial_x^\beta \partial_z^\alpha k_j(x,z)| \le A_{M,\alpha,\beta}|z|^{-M}2^{j[n+m-M+|\alpha|+|\beta|]}, \quad M \ge 0. \tag{4}$$

These differ from (28) of Chapter 6 only in that the operators ∂_x^β acting on the $k_j(x,z)$ introduce the further factors $2^{j|\beta|}$; this is because by (2), the effect of ∂_x^β is to replace the symbol of order zero by a symbol of order $|\beta|$. From (4), we get the following estimate for $k(x,z) = \sum k_j(x,z)$, in analogy with (23) of the previous chapter:

$$|\partial_x^\beta \partial_z^\alpha k(x,z)| \le A_{\alpha,\beta}|z|^{-n-|\alpha|-|\beta|}, \quad z \neq 0. \tag{5}$$

Since we have $T_a f(x) = \int k(x,z)\,f(x-z)\,dz$, it follows that we can take $K(x,y) = k(x,x-y)$, and therefore (5) is equivalent to (1).

1.2 Failure of L^2 boundedness. In view of the proposition just proved, and the fact that as a result such operators are bounded on L^p, $1 < p < \infty$, once they are bounded on L^2,‡ we could be tempted to think that the use of the class $S_{1,1}$ would give the proper extension of all results hitherto proved for the standard symbol class. That this hope is misplaced, and that in this sense the class $S_{1,1}$ must remain forbidden fruit, is the thrust of the following observation.

PROPOSITION 2. *There exists an $a \in S^0_{1,1}$ so that T_a is not bounded from $L^2(\mathbf{R}^n)$ to itself.*

1.2.1 A good example of an element of $S^0_{1,1}$, already to a large extent typical, is given by an a of the form

$$a(x,\xi) = \sum_{j=1}^{\infty} a_j(x)\widehat{\Psi}(2^{-j}\xi); \tag{6}$$

the operator T_a is then of the form

$$(T_a f)(x) = \sum_{j=1}^{\infty} a_j(x)(\Delta_j f)(x). \tag{7}$$

Here the condition that $a \in S^0_{1,1}$ is then

$$|\partial_x^\beta a_j(x)| \le A_\beta 2^{j|\beta|}, \qquad \text{for all } \beta. \tag{8}$$

Indeed, notice that since $\widehat{\Psi}(2^{-j}\xi)$ is supported in $2^{j-1} \le |\xi| \le 2^{j+1}$, for each ξ at most two terms in the sum (6) are nonzero; since these occur only where $|\xi| \sim 2^j$, it is clear that (8) implies that $a \in S^0_{1,1}$.

We shall construct our counterexample on $\mathbf{R}^1$. To do this it is convenient to modify (6) slightly, replacing $\widehat{\Psi}$ by another $\widehat{\Psi_1} \in \mathcal{S}$ with $\widehat{\Psi_1}(\xi)$ supported in $2^{-1/2} \le |\xi| \le 2^{1/2}$ and $\widehat{\Psi_1}(\xi) = 1$ for $2^{-1/4} \le |\xi| \le 2^{1/4}$. Using such a $\widehat{\Psi_1}$, we choose $a_j(x) = e^{-2\pi i 2^j x}$, and set

$$a(x,\xi) = \sum_{j=1}^{\infty} e^{-2\pi i 2^j x}\widehat{\Psi_1}(2^{-j}\xi). \tag{9}$$

Noting that $\widehat{\Psi_1}(2^{-j}\xi)$ is supported where $2^{j-1/2} \le |\xi| \le 2^{j+1/2}$, we see that, for each ξ, at most one term in the sum (9) is nonzero. Since our a_j clearly satisfy (8), it follows that the a given by (9) belongs to $S^0_{1,1}$.

‡ See also Chapter 6, §5.1.

We next choose f_0 to be a nonzero element of $\mathcal{S}$ whose Fourier transform is supported in the set $|\xi| \le 1/2$, and let

$$f_N(x) = \sum_{j=4}^{N} (1/j) e^{2\pi i 2^j x} f_0(x).$$

Because of orthogonality we will see that

$$\|f_N\|^2_{L^2} = \Big(\sum_{j=4}^{N} j^{-2}\Big)\|f_0\|^2_{L^2} \le c, \qquad \text{as } N \to \infty. \tag{10}$$

However the operator T_a, with a given by (9), will completely undo this orthogonality. In fact, we will have

$$T_a f_N = \Big(\sum_{j=4}^{N} j^{-1}\Big) f_0, \tag{11}$$

giving $\|T_a f_N\|_{L^2} \ge c \log N$; hence T_a is not bounded on L^2.

It remains only to verify the assertions (10) and (11). Note first that the Fourier transform of $e^{2\pi i 2^j x} f_0(x)$ is $\widehat{f_0}(\xi - 2^j)$; since $\widehat{f_0}(\xi)$ is supported in $|\xi| \le 1/2$, the functions $\widehat{f_0}(\xi - 2^j)$, $j \ge 1$ have disjoint supports and are thus orthogonal. Since the Fourier transform is unitary, this means that the functions $e^{2\pi i 2^j x} f_0(x)$, $1 \le j \le N$, are also pairwise orthogonal, and (10) follows directly.

Next observe that

$$T_a(e^{2\pi i 2^j x} f_0)(x) = f_0(x), \qquad j \ge 4.$$

Indeed, $(e^{2\pi i 2^j x} f_0)\widehat{\ }(\xi) = \widehat{f_0}(\xi - 2^j)$ is supported in $|\xi - 2^j| \le 1/2$ and $\widehat{\Psi_1}(2^{-j}\xi) \equiv 1$ there, since that set is included in the set where $2^{j-1/4} \le |\xi| \le 2^{j+1/4}$; here we use that $j \ge 4$. Hence (11) is also proved, establishing the proposition.

1.2.2 Remarks. We make some additional remarks about operators of the form (7) and their variants.

1. Suppose a is a bounded function on $\mathbf{R}^n$. If we set $a_j = S_j(a)$, then the sequence a_j satisfies (8). Thus the operator $f \mapsto \sum_{j=1}^{\infty} S_j(a)\Delta_j(f)$ or, more generally, the operator

$$f \mapsto \sum_{j=1}^{\infty} S_{j-\ell}(a)\Delta_j(f) \tag{12}$$

(where ℓ is a fixed integer) is an operator corresponding to a symbol in $S^0_{1,1}$. It is worth noting that, if ℓ is chosen to be sufficiently large, then the operator (12) is obviously bounded on $L^2(\mathbf{R}^n)$. Indeed, the spectrum of $S_{j-\ell}(a)\Delta_j(f)$ is contained in $2^{j-2} \le |\xi| \le 2^{j+2}$ (if we fix $\ell \ge 3$), so the individual terms of (12) are essentially mutually orthogonal.

2. Operators such as (12), but with the summation extended over all integers, arise in several problems. Observe that if a and f are (say) respectively bounded and in $\mathcal{S}$, then

$$a \cdot f = \lim_{j \to +\infty} S_{j-\ell}(a) S_j(f) \quad \text{a.e.},$$

while $\lim_{j \to -\infty} S_{j-\ell}(a) S_j(f) \equiv 0$. Thus

$$a \cdot f = \sum_{j=-\infty}^{\infty} [S_{j-\ell}(a) S_j(f) - S_{j-\ell-1}(a) S_{j-1}(f)],$$

and rearranging terms gives

$$a \cdot f = \sum_{j=-\infty}^{\infty} S_{j-\ell}(a) \Delta_j(f) + \sum_{j=-\infty}^{\infty} \Delta_{j-\ell}(a) S_{j-1}(f).$$

The two sums above may be taken as alternate substitutes for the product $a \cdot f$ appearing on the left side; these are usually referred to as *paraproducts*. The first of these, when the roles of a and f are reversed (which is merely a variant of the second sum), is the bilinear operator

$$B(a, f) = \sum_{j=-\infty}^{\infty} \Delta_{j+\ell}(a) S_j(f). \tag{13}$$

It has several remarkable properties: if a is merely in BMO, the operator $f \mapsto B(a, f)$ is bounded on L^2, and has a kernel satisfying the Calderón-Zygmund conditions (1). The proofs of these assertions are postponed until later (§3.3), where they are taken up in connection with the characterization of operators with Calderón-Zygmund kernels that are bounded on L^2.

3. Further discussion of operators of this type can be found in §5.20–§5.21 below.

1.3 The operators in the forbidden class $S_{1,1}$ do have a significant redeeming feature: their regularity on function spaces that involve a *positive* degree of smoothness. A representative example is given by their action on the Lipschitz spaces Λ_γ.

PROPOSITION 3. *Suppose the symbol a belongs to $S^0_{1,1}$. Then T_a extends to a bounded mapping of Λ_γ to itself, for $\gamma > 0$.*

1.3.1 The operator T_a is initially defined on $\mathcal{S}$, and so we will first show that

$$\|T_a(f)\|_{\Lambda_\gamma} \le A_\gamma \|f\|_{\Lambda_\gamma}, \qquad f \in \mathcal{S}. \tag{14}$$

This is essentially a reprise of the argument in Chapter 6, §5.3, which proved the corresponding result for the symbol classes $S^m = S^m_{1,0}$.

We begin by recalling that $T_a = \sum_{j=0}^{\infty} T_{a_j}$, where $T_{a_0} = T_a S_0$ and $T_{a_j} = T_a \Delta_j$ for $j \ge 1$. Moreover,

$$T_a(f) = \sum_{j=0}^{\infty} T_{a_j}(f_j'), \tag{15}$$

with $f_j' = (\Delta_{j-1} + \Delta_j + \Delta_{j+1}) f$, so that

$$\|f_j'\|_{L^\infty} \le A\, 2^{-j\gamma} \|f\|_{\Lambda_\gamma}. \tag{16}$$

Next, because of the estimate (4) for the kernels of the T_{a_j}, we see that

$$\|\partial_x^\alpha T_{a_j}\| \le A_\alpha\, 2^{j|\alpha|}, \qquad \text{for all } \alpha,$$

where the operator norm $\|\cdot\|$ above is that of mappings from L^∞ to itself. Thus, because of (15) and (16), we have $T_a(f) = \sum_{j=0}^{\infty} F_j$, with

$$\|\partial_x^\alpha F_j\|_{L^\infty} \le A_\alpha\, 2^{j(|\alpha| - \gamma)} \|f\|_{\Lambda_\gamma}. \tag{17}$$

Corollary 1 in §5.3.3 of the previous chapter then implies that our conclusion (14) holds for $f \in \mathcal{S}$. Finally, observe that if f is merely in Λ_γ, then the summands in (15) are well defined, and the sum converges in the L^∞ norm; this gives us the natural extension of T_a to Λ_γ (which agrees with its restriction to $\mathcal{S}$). Since again the same estimates hold, the proposition is therefore completely proved.

Similar results hold for other function spaces. For this see §5.6 below.

1.3.2 We can summarize the moral learned from propositions 2 and 3. In the decomposition

$$T_a = \sum T_{a_j},$$

each T_{a_j} was uniformly bounded from L^∞ to itself; that was essentially enough to guarantee the boundedness of T_a on Λ_γ. However, while each T_{a_j} was also uniformly bounded on L^2, the lack of any orthogonality between the summands defeated the L^2 boundedness of the sum T_a.

1.4 Exotic symbols. We have seen that, in general, symbols of the forbidden class $S^0_{1,1}$ do not lead to bounded operators on L^2. Given the situation that holds for the standard symbols, it is natural to inquire as to which further symbol classes are bounded on L^2. In pursuing this matter, we must face the possibility that the kernels of our operators no longer satisfy the Calderón-Zygmund conditions (1). The answer to the question raised, as well as the treatment of other problems, lead us to consider the so-called "exotic classes" $S_{\rho,\rho}$ (see (3), with $\delta = \rho$), where $0 \le \rho < 1$. Setting aside momentarily the issue of L^2 boundedness, we briefly indicate several of the other reasons why the classes $S_{\rho,\rho}$ (and in particular $S_{1/2,1/2}$) are of interest.

1.4.1 The heat equation. We take as our starting point the constructions of approximate inverses to elliptic partial differential operators sketched in Chapter 6, §1.1 and §7.12, and try to extend these considerations to the heat equation. The following notation will be used. Our underlying space will be $\mathbf{R}^{n+1}$ with points $x = (x_0, x_1, \dots, x_n)$. We also write $x = (t, x')$, where $t = x_0$ and $x' = (x_1, \dots, x_n)$. Similarly we split the dual variable ξ as $\xi = (\tau, \xi')$, with $\tau = \xi_0$ dual to $t = x_0$, and $\xi' = (\xi_1, \dots, \xi_n)$ dual to x'.

We consider the operator L given by

$$L(u) = \frac{\partial u}{\partial t} - \sum_{j=1}^{n} \frac{\partial^2 u}{\partial x_j^2}$$

and try to solve the problem $Lu = f$ by constructing an approximate inverse P so that $LP = I + E$, with an appropriately small error term E. We do this be setting $P = T_a$, where the symbol a is given by

$$a(x,\xi) = a(\xi) = (2\pi i\tau + 4\pi^2|\xi'|^2)^{-1}\eta(\xi); \tag{18}$$

here η is a smooth cut-off function that vanishes near the origin and equals 1 for large $\xi = (\tau, \xi')$. Closely connected to the symbol a are the symbols a_{ij}, given by

$$a_{ij}(\xi) = \xi_i\xi_j a(\xi), \qquad \text{for } 1 \le i, j \le n.$$

What inequalities does this symbol satisfy? First, as far as the order of a is concerned, the best that can be said is that

$$|a(\xi)| \le A(1 + |\xi|)^{-1},$$

i.e., that a is of order -1; this is because of the evident order of decrease when ξ is of the form (τ, ξ'), with $|\xi'|^2 \le |\tau|$, $\tau \to \infty$.

Similarly, when differentiating with respect to any of the ξ' variables, the gain is only the factor $(1+|\xi|)^{-1/2}$ (and not $(1+|\xi|)^{-1}$), as is evident

from taking ξ of the form (τ, ξ'), with $|\xi'|^2 \sim |\tau| \to \infty$. Thus the best that can be said for the symbol (18) is that it belongs to the class $S^{-1}_{1/2,0}$; similarly, the symbols a_{ij} belong to $S^0_{1/2,0}$. Finally, if we transform the underlying space by a smooth change of variables (which transforms the heat equation into another "parabolic" equation), the best that we could expect from the change of variables argument in Chapter 6, §6.3 is that the corresponding symbols would be in the classes $S^{-1}_{1/2,1/2}$ and $S^0_{1/2,1/2}$.

1.4.2 Cauchy-Szegö kernel. It is a remarkable fact that considerations that may be thought of as dual to those for the heat equation are the ones appropriate for the Cauchy-Szegö projection operators for certain basic domains in several complex variables. We shall take up some of these matters in detail when we study the Heisenberg group in Chapter 12. For now we shall limit ourselves to citing several of the relevant formulas.

Essentially speaking, the Cauchy-Szegö kernel is a distribution represented, away from the origin, by the function

$$K(x) = c_n(it + |x|^2)^{-n-1}, \qquad (t,x) \in \mathbf{R}^{n+1};$$

the underlying complex space is $\mathbf{C}^{m+1}$ and $n = 2m$. The Fourier transform of K is the function

$$\widehat{K}(\xi) = \begin{cases} c_n' e^{-\pi|\xi'|^2/2\tau}, & \text{if } \tau > 0; \\ 0 & \text{if } \tau \le 0. \end{cases} \tag{19}$$

The symbol corresponding to the Cauchy-Szegö projection is derived from the function $\widehat{K}(\xi) \cdot \eta(\xi)$, with η as in (18) above. This function belongs to $S^0_{1/2,0}$. Here, as in the heat equation (18), the differential inequalities satisfied by (19) are governed by the nonisotropic dilations

$$(\tau, \xi') \mapsto (\delta^2\tau, \delta\xi'), \qquad \delta > 0.$$

To pass from the function $\widehat{K}(\xi)\eta(\xi)$ to the actual symbol of the Cauchy-Szegö projection forces us to consider an implicit change of variables in our formulae. We are thus led again to symbols in the class $S^0_{1/2,1/2}$. See Chapter 12, §7.15.

1.4.3 A simple singularity. A very different source for symbols in the exotic class $S_{\rho,0}$ comes from the Fourier analysis of some basic models of point singularities of functions. The situation is already illustrated in the case of one variable. In $\mathbf{R}^1$, consider first the function $K(x) = |x|^{-\gamma}$ for x small, and with K modified to be smooth away from the origin and to vanish for large x. Then if $\gamma < 1$, the Fourier transform $\widehat{K}(\xi)$ is a standard symbol belonging to the class $S^{\gamma-1}_{1,0}$; moreover, when suitably defined for $\gamma \geq 1$, the corresponding distributions again have symbols whose Fourier transforms belong to $S_{1,0}^{\gamma-1}$.†

Now the function $|x|^{-\gamma}$ represents of course the simplest singularity at $x = 0$. What happens, for example, when the singularity $|x|^{-\gamma}$ is modified so that it also oscillates rapidly as $x \to 0$? A simple instance of this occurs when

$$K(x) = e^{i/|x|}|x|^{-\gamma} \qquad \text{for small } x$$

and again is smooth away from the origin and vanishes for large x. This particular example, among others, was first studied by Riemann; it can be seen in this case that

$$\widehat{K}(\xi) \quad \text{behaves like} \quad c'e^{ic|\xi|^{1/2}}|\xi|^{\gamma/2-3/4}$$

for large ξ; we will prove this in Chapter 8, §1.4.2. It follows that $\widehat{K}(\xi)$ is a symbol belonging to the class $S_{1/2,0}^{\gamma/2-3/4}$.

It is important to emphasize the fundamental difference between this symbol and the two previous types associated to the heat equation and to the Heisenberg group. For the first two classes, when the symbol in question is of order 0, the corresponding pseudo-differential operators will be bounded on all L^p, $1 < p < \infty$; while for the symbols just considered, when $\gamma = 3/2$ (i.e., the operator is of order 0), the operator will be bounded only on L^2. In order for the operator to be bounded on all L^p, $1 < p < \infty$, we must require that $\gamma \leq 1$, so that the operator has order $\leq -1/4$. For these facts, see §5.12 below.

2. Almost orthogonality

After the preliminary observations and remarks in the previous section concerning forbidden and exotic symbols, we return to the main task at hand: to develop a method of wide scope that proves the L^2 boundedness of operators.

The treatment of the $S_{\rho,\rho}$ operators will be the first application of these ideas. Later we shall see that these methods are very useful in several other situations.

2.1 The principle of orthogonality is the key idea underlying the unique nature of L^2. It has several different expressions; probably the simplest is the truism: To prove that an operator T is bounded on L^2, it suffices to do the same for TT^* (or T^*T).

† For these facts, see Chapter 6, §7.4.

The reason that this observation is useful in practice is that if T is (formally) representable by a kernel K, i.e., if

$$Tf(x) = \int K(x,y)\, f(y)\, dy,$$

then T^*T is (formally) representable by the kernel

$$\int \bar{K}(z,x)\, K(z,y)\, dz. \tag{20}$$

The kernel (20) is often better than K because the integration in (20) can have both a smoothing effect and can take into account the cancellation properties of K. Basic examples of this arise when dealing with unitary operators, such as the Fourier transform and the classical Hilbert transform.‡

The idea of using T^*T as we have just described does not always suffice to do the job. A more refined version of the principle of orthogonality is as follows: one has boundedness of an operator T on L^2 if one can decompose it as a sum

$$T = \sum_j T_j,$$

where the norms of the T_j are uniformly bounded, and we have an approximate mutual orthogonality of the summands T_j. The latter can be expressed by the assertion that both $T_iT_j^*$ and $T_i^*T_j$ tend to zero suitably as $|i-j| \to \infty$.

The result we have in mind can be stated precisely as follows: Suppose $\{T_j\}$ is a finite collection of bounded operators on L^2.† We assume that we are given a sequence of positive constants $\{\gamma(j)\}_{j=-\infty}^{\infty}$, with

$$A = \sum_{j=-\infty}^{\infty} \gamma(j) < \infty, \tag{21}$$

and our hypothesis is that

$$\begin{aligned} \|T_i^*T_j\| &\leq [\gamma(i-j)]^2, \\ \|T_iT_j^*\| &\leq [\gamma(i-j)]^2; \end{aligned} \tag{22}$$

here $\|\cdot\|$ denotes the operator norm on L^2.

‡ For the latter, see §5.1; several other examples will appear below in §5.2 and Chapter 9, §1.

† The theorem below is purely a Hilbert space result, but we continue to use the (special) terminology of L^2.

THEOREM 1 (Cotlar's lemma). *Under the assumptions* (21) *and* (22) *above, the operator*

$$T = \sum_j T_j$$

satisfies

$$\|T\| \le A. \tag{23}$$

We have stated the theorem in its *a priori* form, where only finitely many T_j are involved. What is important is that the conclusion (23) gives a bound independent of the number of these T_j.

2.2 Proof. We begin with some preliminary remarks about norms of operators on L^2. First, $\|T^*\| = \|T\|$; to see this note that

$$\|T\| = \sup |\langle Tf, g\rangle|,$$

where the supremum is taken over all $f, g \in L^2$ of norm 1. Since $\langle Tf, g\rangle = \langle f, T^*g\rangle$, the assertion follows. Second, $\|T\|^2 = \|T^*T\|$, because on the one hand

$$\|T^*T\| \le \|T^*\| \cdot \|T\| = \|T\|^2,$$

and on the other hand

$$\|T\|^2 = \sup_{\|f\|_{L^2}=1} \langle Tf, Tf\rangle = \sup \langle T^*Tf, f\rangle \le \|T^*T\|.$$

In particular, when T is self-adjoint, $\|T\|^2 = \|T^2\|$, and by induction $\|T\|^m = \|T^m\|$, at least when m is a power of 2.‡ Using this with T replaced by T^*T (which is visibly self-adjoint), we have

$$\|T\|^{2m} = \|(T^*T)^m\|. \tag{24}$$

In proving the theorem, we shall use (24) because it allows us to most efficiently exploit our hypotheses. Written out in full,

$$(T^*T)^m = \sum_{i_1,\dots,i_{2m}} T^*_{i_1} T_{i_2} T^*_{i_3} \cdots T_{i_{2m}}. \tag{25}$$

We shall estimate this sum by majorizing the norms of the individual summands.

First, associating the factors in each summand as

$$(T^*_{i_1} T_{i_2})(T^*_{i_3} T_{i_4}) \cdots (T^*_{i_{2m-1}} T_{i_{2m}}),$$

‡ This assertion holds for arbitrary positive integers m, but requires a further argument.

and using the first inequality in (22), we get

$$\|T^*_{i_1} T_{i_2} \cdots T_{i_{2m}}\| \le \gamma^2(i_1 - i_2) \cdot \gamma^2(i_3 - i_4) \cdots \gamma^2(i_{2m-1} - i_{2m}). \tag{26}$$

Alternatively, we can associate the factors as

$$T^*_{i_1}(T_{i_2} T^*_{i_3}) \cdots (T_{i_{2m-2}} T^*_{i_{2m-1}}) T_{i_{2m}}.$$

Then since $\|T_{i_1}\| \le \gamma(0) \le A$, and similarly $\|T_{i_{2m}}\| \le A$, we get

$$\|T^*_{i_1} T_{i_2} \cdots T_{i_{2m}}\| \le A^2 \cdot \gamma^2(i_2 - i_3) \cdot \gamma^2(i_4 - i_5) \cdots \gamma^2(i_{2m-2} - i_{2m-1}). \tag{27}$$

We take the geometric mean of (26) and (27) and insert this in (25). The result is

$$\|(T^*T)^m\| \le \sum_{i_1,\dots,i_{2m}} A \cdot \gamma(i_1 - i_2) \cdot \gamma(i_2 - i_3) \cdots \gamma(i_{2m-1} - i_{2m}).$$

In the above, we first sum in i_1 and use the fact that $\sum_{i_1} \gamma(i_1 - i_2) \le A$. Next, we sum in i_2, using $\sum_{i_3} \gamma(i_2 - i_3) \le A$. Continuing in this way for the indices $i_1, \dots, i_{2m-1}$ gives

$$\|(T^*T)^m\| \le A^{2m} \sum_{i_{2m}} 1.$$

We assumed that we had only finitely many nonzero T_i's; say there are N of them. Then, by (24),

$$\|T\| \le A \cdot N^{1/2m}.$$

Finally, we let $m \to \infty$, proving the theorem.

2.3 Remarks.

1. In the theorem above, the summation was taken over the integers. However, we can replace $\mathbf{Z}$ here by $\mathbf{Z}^r$ (for any r) and the result still holds, without any change in the proof. Indeed, if we write $j = (j_1, j_2, \dots j_r) \in \mathbf{Z}^r$, $T = \sum_{j\in\mathbf{Z}^r} T_j$, and require $\sum_{j\in\mathbf{Z}^r} \gamma(j) = A < \infty$, then we can leave the assumptions (22) and conclusion (23) unchanged.

One may also obtain a variant of the theorem where the decomposition of T is given by integrals rather than sums. For such a formulation, see §5.5.

2. A special case of the theorem—actually, its original version—arises when the T_j are self-adjoint and mutually commuting. In that case the T_j can be given a simultaneous spectral resolution and the result thereby reduces to the easy situation where the T_j are scalar multiples of the identity. See §5.4.

3. We now describe a cruder version of the theorem, which can sometimes be used as a substitute in applications and whose proof is much more direct. We assume the T_j satisfy:

$$\|T_j\| \le A,$$

$$\|T_i^* T_j\| \le \gamma(i)\cdot\gamma(j), \quad \text{if } i \ne j, \text{ and} \tag{28}$$

$$T_i T_j^* = 0, \qquad \text{if } i \ne j.$$

As before, we suppose that $\sum_j \gamma(j) = A < \infty$.

PROPOSITION. *Under these assumptions, we have that*

$$\| \sum_j T_j \| \le 2^{1/2}\cdot A.$$

To prove this, observe that, because of (28), the ranges of the T_j^* are mutually orthogonal. In fact

$$\langle T_j^* f, T_i^* g\rangle = \langle T_i T_j^* f, g\rangle = 0, \qquad \text{if } i \ne j.$$

Therefore $T_j^* = E_j T_j^*$, where the E_j are mutually orthogonal projections, and hence also $T_j = T_j E_j$. As a result

$$\sum_j \|T_j(f)\|^2 \le A^2 \sum_j \|E_j(f)\|^2 \le A^2\|f\|^2.$$

Also if $T = \sum_j T_j$, then

$$\|T(f)\|^2 = \sum_{i,j}\langle T_j f, T_i f\rangle = \sum_j \|T_j(f)\|^2 + \sum_{i\ne j}\langle T_i^* T_j f, f\rangle.$$

We now invoke (28) again to get $\|Tf\|^2 \le 2A^2\|f\|^2$, which is the desired conclusion.

2.4 The class $S_{0,0}$. Our first application of almost orthogonality is the following result:

PROPOSITION. *Suppose a is a symbol that belongs to the class $S^0_{0,0}$. Then the operator T_a, initially defined on $\mathcal{S}$, has a bounded extension to an operator from $L^2(\mathbf{R}^n)$ to itself.*

We observe that $T_a = S\mathcal{F}$, where $(\mathcal{F}f)(\xi) = \widehat{f}(\xi)$ is the Fourier transform, and

$$Sf(x) = \int_{\mathbf{R}^n} a(x,\xi) e^{2\pi i x\cdot\xi} f(\xi)\, d\xi. \tag{29}$$

By Plancherel's theorem, it suffices to establish the L^2-boundedness of the operator S. Notice that, in view of the assumptions on a (that is, (3) with $\rho = \delta = 0$), the roles of x and ξ in S are perfectly symmetric.

We now decompose the ξ-space, and, because of the aforementioned symmetry, we also decompose the x-space in the same way. Moreover, the differential inequalities satisfied by the symbol a lead us to make these portions into sets of essentially unit size.

To be precise, we choose a smooth, nonnegative function ϕ that is supported in the unit cube

$$Q_1 = \{x : |x_j| \le 1 \text{ for } j = 1, \dots, n\},$$

and for which

$$\sum_{i\in\mathbf{Z}^n} \phi(x - i) = 1. \tag{30}$$

To construct such a ϕ, simply fix any smooth, nonnegative ϕ_0 that equals 1 on the cube $Q_{1/2} = 1/2\cdot Q_1$, and is supported in Q_1. Noting that $\sum\limits_{i\in\mathbf{Z}^n} \phi_0(x-i)$ converges and is bounded away from 0 for all $x \in \mathbf{R}^n$, we take

$$\phi(x) = \phi_0(x)\left[\sum_{i\in\mathbf{Z}^n}\phi_0(x-i)\right]^{-1}.$$

Next, let $\mathbf{i} = (i, i') \in \mathbf{Z}^{2n} = \mathbf{Z}^n \times \mathbf{Z}^n$ denote an element of $\mathbf{Z}^{2n}$, and similarly write $\mathbf{j} = (j, j')$ for another element of $\mathbf{Z}^{2n}$. We set

$$a_{\mathbf{i}}(x,\xi) = \phi(x - i)\, a(x,\xi)\, \phi(\xi - i'),$$

and write $S_{\mathbf{i}}$ for the operator (29) with $a(x,\xi)$ replaced by $a_{\mathbf{i}}(x,\xi)$. In view of (30) we have that

$$S = \sum_{\mathbf{j}\in\mathbf{Z}^{2n}} S_{\mathbf{j}} \tag{31}$$

where, as is easily verified, for each $f \in \mathcal{S}$, the sum $\sum\limits_{\mathbf{j}\in\mathbf{Z}^{2n}} S_{\mathbf{j}}(f)$ converges in $\mathcal{S}$ to $S(f)$.

The main point is then to verify the almost-orthogonality estimates:

$$\|S_{\mathbf{i}}^* S_{\mathbf{j}}\| \le A(1 + |\mathbf{i} - \mathbf{j}|)^{-2N} \tag{32}$$

and

$$\|S_{\mathbf{i}} S_{\mathbf{j}}^*\| \le A(1 + |\mathbf{i} - \mathbf{j}|)^{-2N}. \tag{33}$$

Here $\|\cdot\|$ denotes the L^2 operator norm, N is sufficiently large, and the bound A is independent of $\mathbf{i}$ and $\mathbf{j}$.

2.4.1 To deal with (32) and (33), it is useful to recall the following simple estimate of the norm of an operator S in terms of the size of its kernel.

LEMMA. *Suppose S is given by*

$$(Sf)(x) = \int s(x,y)\, f(y)\, dy,$$

where the kernel s satisfies

$$\sup_x \int |s(x,y)|\, dy \le 1,$$

and

$$\sup_y \int |s(x,y)|\, dx \le 1.$$

Then

$$\|S\|_{L^2 \to L^2} \le 1. \tag{34}$$

The L^2 operator norm $\|S\|$ equals

$$\sup |\langle Sf, g\rangle| = \sup \left| \int\int s(x,y)\, f(y)\, \bar{g}(x)\, dy\, dx \right|,$$

where the supremum is taken over all f and g with $\|f\|_{L^2} \le 1$ and $\|g\|_{L^2} \le 1$. Since

$$|fg| \le (|f|^2 + |g|^2)/2,$$

the integral in question is dominated by

$$\frac{1}{2}\left\{ \int\int |s(x,y)|\, |f(y)|^2\, dy\, dx + \int\int |s(x,y)|\, |g(x)|^2\, dy\, dx \right\}.$$

In the first integral, we carry out the integration first with respect to x; in the second integral, we integrate first with respect to y. Invoking our hypotheses then gives us the desired conclusion (24).

2.4.2 Observe that if

$$(S_{\mathbf{i}}^* S_{\mathbf{j}})(f)(\xi) = \int s_{\mathbf{ij}}(\xi,\eta)\, f(\eta)\, d\eta,$$

then the kernel of $S_{\mathbf{i}}^* S_{\mathbf{j}}$ is given by

$$s_{\mathbf{ij}}(\xi,\eta) = \int \bar{a}_{\mathbf{i}}(x,\xi)\, a_{\mathbf{j}}(x,\eta) e^{2\pi i x\cdot(\eta-\xi)}\, dx.$$

In the above integral we integrate by parts, using the identity

$$(I - \Delta_x)^N e^{2\pi i x\cdot(\eta-\xi)} = (1 + 4\pi^2|\eta - \xi|^2)^N e^{2\pi i x\cdot(\eta-\xi)}.$$

We also note that $a_{\mathbf{i}}(x,\xi)$ and $a_{\mathbf{j}}(x,\eta)$ are given by

$$\phi(x-i)\, a(x,\xi)\, \phi(\xi - i') \qquad \text{and} \qquad \phi(x-j)\, a(x,\eta)\, \phi(\eta - j')$$

respectively, and so have disjoint x-support, unless $i - j \in Q_1$. These observations lead to the bounds

$$\begin{cases} |s_{\mathbf{ij}}(\xi,\eta)| & \le \dfrac{A_N\, \phi(\xi - i')\, \phi(\eta - j')}{(1 + |\xi - \eta|)^{2N}} \quad \text{if } i - j \in Q_1, \\ s_{\mathbf{ij}}(\xi,\eta) & = 0 \quad \text{if } i - j \notin Q_1. \end{cases}$$

So if we invoke the lemma, we get immediately that

$$\|S_{\mathbf{i}}^* S_{\mathbf{j}}\| \le A(1 + |\mathbf{i} - \mathbf{j}|)^{-2N}.$$

Since, as we have noted, the situation is symmetric in x and ξ, the same proof also shows that

$$\|S_{\mathbf{i}} S_{\mathbf{j}}^*\| \le A(1 + |\mathbf{i} - \mathbf{j}|)^{-2N},$$

which gives (32) and (33).

Now it is only a matter of applying the almost-orthogonality theorem in the version described in the first remark in §2.3. Here $r = 2n$ and $\gamma(\mathbf{j})$ is a multiple of $(1 + |\mathbf{j}|)^{-N}$, with $N > 2n$. The result is that

$$\|\sum S_{\mathbf{j}}\| \le A,$$

where the sum is taken over any finite subset of $\mathbf{Z}^{2n}$, and the bound A is independent of that subset. Because of the convergence (31) of $\sum S_{\mathbf{j}}(f)$ in $\mathcal{S}$ (whenever $f \in \mathcal{S}$), we may conclude that

$$\|S(f)\|_{L^2} \le A\|f\|_{L^2}, \qquad f \in \mathcal{S}.$$

From this, the extendability of S, and therefore of T_a, to a bounded operator on L^2 is evident.

Remark. Note that the above argument gives a bound for T that depends only on the L^∞ norms of the derivatives (of order at most $4n+1$) of the symbol a. Incidentally, more complicated arguments allow one to reduce this order substantially (see §5.13).

2.5 The class $S^0_{\rho,\rho}$. We shall now obtain a similar result for the more general classes $S^0_{\rho,\rho}$, when $0 \leq \rho < 1$.

THEOREM 2. *Suppose a is a symbol that belongs to $S^0_{\rho,\rho}$ with $0 \leq \rho < 1$. Then the operator T_a (initially defined on $\mathcal{S}$), has a bounded extension from $L^2(\mathbf{R}^n)$ to itself.*

Proof. We begin by replacing the symbol $a(x,\xi)$ with

$$a_\varepsilon(x,\xi) = a(x,\xi)\,\gamma(\varepsilon x, \varepsilon y), \qquad 0 < \varepsilon \leq 1;$$

here γ is a fixed smooth function of compact support, with $\gamma(0,0) = 1$. Notice that the symbols a_ε belong to $S^0_{\rho,\rho}$ uniformly in ε and that, for $f \in \mathcal{S}$,

$$T_{a_\varepsilon}(f) \to T_a(f), \quad \text{as } \varepsilon \to 0,$$

in the topology of $\mathcal{S}$. This will allow us to assume that our initial symbol has compact support in both x and ξ, so the manipulations appearing below will be automatically justified. Also, in what follows, the explicit dependence on ε will be suppressed, but all of our estimates will be made independent of ε.

We treat the operator T_a by using the standard dyadic partition of the ξ-space, which occurred already in §4.1 of the previous chapter. That is, we write

$$T_a = \sum_{j=0}^{\infty} T_{a_j} = T_a S_0 + \sum_{j=1}^{\infty} T_a \Delta_j, \tag{35}$$

where $a_0(x,\xi) = a(x,\xi)\,\eta(\xi)$ is supported in $|\xi| \leq 2$, and, for $j \geq 1$, $a_j(x,\xi) = a(x,\xi)\,\widehat{\Psi}(2^{-j}\xi)$ is supported in the shell $2^{j-1} \leq |\xi| \leq 2^{j+1}$. Note that the symbols a_j are uniformly in $S^0_{\rho,\rho}$.

There are two distinct parts to the proof: first, that the operators T_{a_j} are essentially mutually orthogonal, and second, that they are uniformly bounded in norm. To proceed, it will be convenient to break the sum (35) into two parts

$$T_a = \sum_{j \text{ even}} T_{a_j} + \sum_{j \text{ odd}} T_{a_j},$$

so that the summands in each part have disjoint ξ-support; it suffices to prove the boundedness of each sum separately.

Let us first consider the sum taken over the odd j. We have that

$$T_{a_j} T^*_{a_k} = 0 \qquad \text{if } j \neq k \tag{36}$$

because $T_{a_j}(T_{a_k})^* = T_a\Delta_j(T_a\Delta_k)^* = T_a\Delta_j\Delta_k^* T_a$, and the supports of the multipliers corresponding to Δ_j and Δ_k are disjoint. Next, we estimate $T^*_{a_j}T_{a_k}$. Since

$$T_{a_j}(f)(z) = \int_{\mathbf{R}^n\times\mathbf{R}^n} a_j(z,\xi)e^{2\pi i\xi\cdot(z-y)} f(y)\,d\xi\,dy,$$

and

$$T^*_{a_k}(g)(x) = \int_{\mathbf{R}^n\times\mathbf{R}^n} \bar{a}_k(z,\eta)e^{2\pi i\eta\cdot(x-z)} g(z)\,dz\,d\eta,$$

(see §1.4 of the previous chapter), we see that

$$(T^*_{a_j}T_{a_k})f(x) = \int K(x,y)\,f(y)\,dy,$$

with

$$K(x,y) = \int_{\mathbf{R}^n\times\mathbf{R}^n\times\mathbf{R}^n} \bar{a}_k(z,\eta)\,a_j(z,\xi)e^{2\pi i[\xi\cdot(z-y)-\eta\cdot(z-x)]}\,dz\,d\eta\,d\xi. \tag{37}$$

To bound the kernel K, we exploit the oscillatory nature of the exponential (and the relative smoothness of the factors $\bar{a}_k$ and a_j) by integrating by parts with respect to the variables z, η, and ξ. First, one carries this out with respect to the z-variable by writing

$$\frac{(I-\Delta_z)^N}{(1+4\pi^2|\xi-\eta|^2)^N} e^{2\pi i(\xi-\eta)\cdot z} = e^{2\pi i(\xi-\eta)\cdot z},$$

inserting this in (37), and passing the z-differentiations to the factors a_k and a_j. Next, one performs a similar process on the η-variable, beginning with

$$\frac{(I-\Delta_\eta)^N}{(1+4\pi^2|x-z|^2)^N} e^{2\pi i\eta(x-z)} = e^{2\pi i\eta(x-z)},$$

and then passing the differentiations in the η-variable. Finally, an analogous step is carried out for the ξ-variable. If we take into account the differential inequalities for the symbols a_j (see (3)), and the restrictions on their supports, we see that each order of differentiation in the z-variable gives us a factor of order

$$(1+|\xi-\eta|)^{-1} \approx 2^{-\max(k,j)}$$

for every factor of order

$$(1+|\xi|+|\eta|)^\rho \approx 2^{\rho\max(k,j)}$$

that we may lose. As a result, the kernel K is dominated by a constant multiple of

$$2^{\max(k,j)\cdot 2\rho N}\cdot 2^{-\max(k,j)\cdot 2N}\cdot 2^{\max(k,j)\cdot 2n}\int_{\mathbf{R}^n} Q(x-z)\,Q(z-y)\,dz,$$

where $Q(z) = (1+|z|)^{-2N}$, if $k \neq j$.

Now if $K_0(x,y) = \int Q(x-z)\,Q(z-y)\,dz$, then

$$\int K_0(x,y)\,dy = \int K_0(x,y)\,dx = \left(\int (1+|z|)^{-2N}\right)^2 < \infty,$$

if $2N > n$. Thus, invoking the lemma in §2.4.1, we get

$$\|T^*_{a_j} T_{a_k}\| \le A \cdot 2^{\max(k,j)[2\rho N - 2N + 2n]}, \qquad j \ne k,$$

which implies that

$$\|T^*_{a_j} T_{a_k}\| \le \gamma(j)\,\gamma(k), \qquad j \ne k, \tag{39}$$

with $\gamma(j) = A \cdot 2^{-\varepsilon j}$, $\varepsilon > 0$, if we choose N so large that $N > n/(1-\rho)$; then $\varepsilon = N(1-\rho) - n$.

We have therefore satisfied the hypothesis (28) of the proposition in §2.3, save for what is here the most crucial step: that the summands in (35), namely the operators T_{a_j}, are uniformly bounded in norm.

To prove this, note what happens if we re-scale the symbols

$$a_j(x,\xi) = a(x,\xi)\,\Psi(2^{-j}\xi),$$

by setting

$$\tilde{a}_j(x,\xi) = a_j(2^{-j\rho}x, 2^{j\rho}\xi).$$

In view of the $S_{\rho,\rho}$ inequalities satisfied by the a_j, we observe that the $\tilde{a}_j$ belong to $S^0_{0,0}$, uniformly in j.

Next, if Λ_j denote the scaling operators given by

$$\Lambda_j(f)(x) = f(2^{j\rho}x),$$

then, as is easily verified,

$$T_{a_j} = \Lambda_j T_{\tilde{a}_j} \Lambda_j^{-1}.$$

Now $\|\Lambda_j f\|_{L^2} = 2^{nj\rho/2}\|f\|_{L^2}$ and $\|\Lambda_j^{-1} f\|_{L^2} = 2^{-nj\rho/2}\|f\|_{L^2}$; so the boundedness result for operators of the class $S^0_{0,0}$ (given in §2.4), applied to the $\tilde{a}_j$, guarantees that

$$\|T_{a_j}\| \le A. \tag{47}$$

We may therefore conclude that the sum $\sum\limits_{j \text{ odd}} T_{a_j}$ yields a bounded operator on L^2; the sum $\sum\limits_{j \text{ even}} T_{a_j}$ is treated similarly, and the theorem is therefore proved.

3. L^2 theory of operators with Calderón-Zygmund kernels

We now turn to the important problem of determining necessary and sufficient conditions for an operator, whose kernel satisfies differential inequalities of the Calderón-Zygmund type (i.e., of the nature (1)), to be extendable to a bounded operator on $L^2(\mathbf{R}^n)$. The various issues involved in this problem can be formulated in terms of two related questions.

We consider first an operator T, initially defined as a mapping from test functions in $\mathcal{S}$ to distributions in $\mathcal{S}'$, to which is associated a kernel $K(x,y)$, defined when $x \ne y$, that satisfies inequalities of the kind:

$$|\partial_x^\alpha \partial_y^\beta K(x,y)| \le A|x-y|^{-n-|\alpha|-|\beta|}; \tag{42}$$

such kernels are called *Calderón-Zygmund kernels* or *singular integral kernels.*†

The assumed relation between T and K is that, whenever $f \in \mathcal{S}$ has compact support, then the distribution Tf can be identified with the function

$$(Tf)(x) = \int K(x,y)\,f(y)\,dy, \tag{43}$$

for x outside the support of f.‡

QUESTION 1. *What additional conditions must be imposed on T to guarantee that it extends to a bounded operator from $L^2(\mathbf{R}^n)$ to itself?*

In the second version of our problem, our starting point is not a (densely defined) operator T, but rather a kernel $K(x,y)$, given for $x \ne y$.

QUESTION 2. *Suppose the function $K(x,y)$ satisfies the differential inequalities* (42). *What additional ("cancellation") conditions must be imposed on K so that there exists a bounded operator $T : L^2(\mathbf{R}^n) \to L^2(\mathbf{R}^n)$ having kernel K in the sense of* (43)*?*

3.1 Three simple propositions. The following observations will help clarify the nature of the conditions that will be imposed, as well as motivate some aspects of the proof of the main theorem. The first proposition shows graphically that a cancellation condition must be required for the kernel K.

PROPOSITION 1. *Suppose $K(x,y)$ is a function given for $x \ne y$ that satisfies the inequality $K(x,y) \ge c|x-y|^{-n}$, $c > 0$. Then there does not exist an operator T that is bounded on $L^2(\mathbf{R}^n)$ for which K is the kernel, in the sense of* (43).

† This terminology also applies to kernels satisfying a weaker version of these inequalities (such as (48) below).

‡ Note that the operator T is *not* uniquely determined by the kernel $K(x,y)$ (via (43)). A simple example is the identity operator (or, more generally, multiplication by a bounded function), for which the kernel is identically zero.

3.1.1 The proof is by contradiction, which is achieved by testing the operator T (for which (43) is assumed to hold) on a function supported in a large ball. The support is chosen so that both it and its complement are essentially uniformly distributed in that ball, both having measure that is a positive fraction of the measure of the ball.

To be precise, let Q be a small cube (say of side length $1/2$) centered at the origin. For any lattice point j in $\mathbf{R}^n$ (i.e., $j \in \mathbf{Z}^n$), write $Q_j = Q+j$ for the cube Q translated so that its center is j. Let

$$S = \bigcup_{|j|\leq 2R} Q_j ;$$

then observe that

$$|S \cap \{x : |x| < R\}| \geq c_1 R^n, \qquad \text{for } R \geq 1,$$

and also that

$$|{}^c S \cap \{x : |x| < R\}| \geq c_1 R^n, \qquad \text{for } R \geq 1,$$

for some positive constant c_1.

Take $f = \chi_S$. Then if $x \notin S$, we have that

$$Tf(x) = \int K(x,y)\, f(y)\, dy \geq c \sum_{|j|\leq 2R} \int_{Q_j} \frac{dy}{|x-y|^n}.$$

So if, in addition, $|x| \leq R$, then

$$Tf(x) \geq c' \sum_{0<|j|\leq R} |j|^{-n} \geq c'' \log R.$$

Thus,

$$\begin{aligned} \|Tf\|_{L^2}^2 &\geq \int_{{}^c S\cap\{x:|x|<R\}} |Tf(x)|^2\, dx \\ &\geq (c'')^2 \cdot (\log R)^2 \cdot |{}^c S \cap \{x : |x| < R\}| \geq c \cdot (\log R)^2 R^n, \end{aligned}$$

while clearly $\|f\|_{L^2}^2 \leq \bar{c} R^n$, which gives the desired contradiction.

3.1.2 The second observation is a more precise version of the first, but in the translation-invariant setting. It also gives a hint about formulating an answer to Question 2 above. We consider here a distribution K on $\mathbf{R}^n$ that, away from the origin, equals a function (which we denote by $K(x)$) with

$$|K(x)| \leq A|x|^{-n}.$$

PROPOSITION 2. *If the operator $T(f) = f * K$, initially defined for $f \in \mathcal{S}$, extends to a bounded operator from $L^2(\mathbf{R}^n)$ to itself, then there is a constant $\bar{A}$ so that*

$$\left| \int_{\varepsilon<|x|<N} K(x)\, dx \right| \leq \bar{A}, \tag{44}$$

whenever $0 < \varepsilon < N < \infty$.

To prove the proposition, choose a C^∞ function ϕ, supported in the unit ball $|x| \leq 1$, with $0 \leq \phi(x) \leq 1$, and $\phi(x) = 1$ when $|x| \leq 1/2$. Define $\phi^R(x) = \phi(x/R)$. According to our assumption, the Fourier transform of the distribution K is a bounded measurable function, which we denote by m. Then, by the Fourier inversion theorem,

$$T(\phi^R)(0) = (K * \phi^R)(0) = \int m(\xi)\, \widehat{\phi}(R\xi)\, R^n\, d\xi.$$

As a result,

$$|(K * \phi^R)(0)| \leq A \int |\widehat{\phi}(R\xi)|\, R^n\, d\xi = A \int |\widehat{\phi}(\xi)|\, d\xi \leq A'.$$

Thus for all ε and N,

$$|[K * (\phi^N - \phi^\varepsilon)](0)| \leq 2A'. \tag{45}$$

However, the difference

$$\int_{\varepsilon<|x|<N} K(x)\, dx - [K * (\phi^N - \phi^\varepsilon)](0)$$

is majorized by

$$\int_{N/2\leq|x|\leq N} |K(x)|\, dx + \int_{\varepsilon/2\leq|x|\leq\varepsilon} |K(x)|\, dx;$$

so the condition $|K(x)| \leq A|x|^{-n}$, together with (45), gives us (44), and the proposition is proved.

Remarks. (i) This proposition is a variant of the one in Chapter 6, §4.5, which requires some further regularity of the function $K(x)$.

(ii) When this additional regularity holds, the condition (44) is also sufficient, as follows by the arguments in Chapter 6, §4.5, or the more general result (Theorem 4) proved below.

(iii) However, without further regularity of K, the L^2 boundedness may fail; see §5.16 below.

3.1.3 Our last observation is also for translation-invariant operators, and brings to light a special cancellation property that holds in this particular setting. As in the previous proposition, we deal with an operator $Tf = f * K$ of the kind considered there: T extends to a bounded operator on $L^2(\mathbf{R}^n)$, the distribution K equals a function $K(x)$ away from the origin with $|K(x)| \le A|x|^{-n}$, but we also require some additional regularity of K, for instance

$$|\partial_x^\alpha K(x)| \le A_\alpha |x|^{-n-|\alpha|},$$

or more generally,

$$\int_{|x|\ge 2|y|} |K(x-y) - K(x)|\,dx \le A,$$

for all y.

Assume that f is a bounded function, with compact support, that satisfies $\int f\,dx = 0$. Then the L^2 function Tf is also integrable; i.e., $\int |Tf|\,dx < \infty$. This assertion follows from the fact that f is a constant multiple of an atom (see §3.1 of Chapter 3). Moreover, by the theorem in Chapter 3, §3.2, Tf is in the Hardy space H^1, and

$$\int_{\mathbf{R}^n} Tf(x)\,dx = 0, \tag{46}$$

since every H^1 function can be represented as an L^1-convergent sum of atoms.

We summarize our conclusions as follows:

PROPOSITION 3. *Suppose T is a translation-invariant operator of the type described above. Let f be a bounded function with compact support whose integral vanishes. Then $Tf \in L^1(\mathbf{R}^n)$, and (46) holds.*

Incidentally, once we have verified that $Tf \in L^1$,† the proof of (46) can be given directly. We have that

$$(Tf)\hat{}(\xi) = m(\xi)\widehat{f}(\xi), \tag{47}$$

where m is the Fourier transform of K. Next, since f and Tf are in $L^1(\mathbf{R}^n)$, both $(Tf)\hat{}(\xi)$ and $\widehat{f}(\xi)$ are continuous in ξ. Now $\widehat{f}(0) = 0$, so

$$(Tf)\hat{}(0) = \lim_{\xi\to 0} m(\xi)\widehat{f}(\xi) = 0,$$

where the limit is taken over an appropriate sequence in the ξ-space; this is because the identity (47) holds for almost every ξ, and the function m is essentially bounded. Hence (46) is established.

† The verification of this is easy; see the proof of (45) in Chapter 3, or §3.2.1 below.

3.2 Answer to the first question. We formulate in detail the answer to the first question posed at the beginning of §3. We are concerned with a linear operator that is given as a continuous mapping

$$T : \mathcal{S} \to \mathcal{S}'$$

from test functions to distributions. We assume that associated to T there is a kernel $K(x,y)$, defined for $x \ne y$, that satisfies the following weaker version of the differential inequalities (1): For some γ, $0 < \gamma \le 1$, we have

$$|K(x,y)| \le A|x-y|^{-n},$$

$$|K(x,y) - K(x',y)| \le A\,\frac{|x-x'|^\gamma}{|x-y|^{n+\gamma}}, \quad \text{if } |x-x'| \le |x-y|/2, \quad \text{and}$$

$$|K(x,y) - K(x,y')| \le A\,\frac{|y-y'|^\gamma}{|x-y|^{n+\gamma}}, \quad \text{if } |y-y'| \le |x-y|/2 \tag{48}$$

The relation between K and T is that if $f \in \mathcal{S}$ has compact support, then, outside the support of f, the distribution Tf agrees with the function

$$(Tf)(x) = \int_{\mathbf{R}^n} K(x,y)\,f(y)\,dy. \tag{49}$$

Now the operator T will be tested on appropriate "bump functions". Such collections of functions have proved useful in several previous contexts; see Chapter 3, §1.8 and Chapter 6, §4.5.

We fix a positive integer N and recall that the set of *normalized bump functions* consists of those smooth functions ϕ, supported in the unit ball, that satisfy

$$|\partial_x^\alpha \phi(x)| \le 1, \qquad 0 \le |\alpha| \le N. \tag{50}$$

While this collection depends on N, the particular choice of N is of no great importance and will not be made explicit in what follows.

For each ball $B = B(x_0, R)$, of center x_0 and radius R, write

$$\phi^{R,x_0}(x) = \phi\left(\frac{x-x_0}{R}\right).$$

As ϕ ranges over the normalized bump functions, we say that ϕ^{R,x_0} ranges over the *normalized bump functions for the ball* $B(x_0,R)$.

The condition on T we are interested in is that T be *restrictedly bounded*: Whenever ϕ is a normalized bump function, the distribution $T(\phi^{R,x_0})$ belongs to L^2, and the estimate

$$\|T(\phi^{R,x_0})\|_{L^2} \leq A \cdot R^{n/2} \tag{51}$$

holds, with an A that is independent of R, x_0, and ϕ.

Note that (51) yields a simple extension of itself in which the support of ϕ is not necessarily bounded: The inequality continues to hold for any $\phi \in \mathcal{S}$, as long as $\|\phi\|_{\alpha,\beta} \leq 1$ for an appropriate finite collection of seminorms $\|\cdot\|_{\alpha,\beta}$.

Since we are dealing with the problem of L^2 boundedness of the operator T, it is natural to expect that whatever conditions are imposed on T should also be required of its adjoint T^*. Here T^* is given by

$$\langle Tf, g\rangle = \langle f, T^*g\rangle, \qquad \text{whenever } f, g \in \mathcal{S}.$$

Then $T^* : \mathcal{S} \to \mathcal{S}'$ is a continuous mapping (since T is). Moreover, T^* is associated to the kernel

$$K^*(x,y) = \overline{K}(y,x),$$

in the same sense as T is associated to K; it is immediate that K^* satisfies the estimates (48). We are also interested in the condition that T^* be restrictedly bounded, i.e., the existence of an $A > 0$ so that

$$\|T^*(\phi^{R,x_0})\|_{L^2} \leq A \cdot R^{n/2}, \tag{52}$$

for all normalized bump functions ϕ, all $x_0 \in \mathbf{R}^n$, and all $R > 0$.

THEOREM 3. *Suppose T is a continuous linear mapping from $\mathcal{S}$ to $\mathcal{S}'$ associated to a kernel K satisfying* (48) *and* (49). *Then T extends to a bounded linear operator from $L^2(\mathbf{R}^n)$ to itself if and only if both T and T^* are restrictedly bounded, in the sense of* (51) *and* (52).

3.2.1 Proof of the theorem: first part. To begin with, we note that it is immediate that conditions (51) and (52) are necessary. Indeed, if T is bounded on L^2, then

$$\|T(\phi^{R,x_0})\|_{L^2} \leq A'\|\phi^{R,x_0}\|_{L^2} \leq A \cdot R^{n/2},$$

merely because ϕ^{R,x_0} is supported in $B(x_0, R)$ and its absolute value is at most 1; a similar observation holds for T^*.

The thrust of the theorem is of course the converse, to which we now turn. As a preliminary matter, we point out that if f is a smooth function of compact support and $\int_{\mathbf{R}^n} f(x)\,dx = 0$, then $Tf \in L^1(\mathbf{R}^n)$. In fact, Tf is in $L^2(\mathbf{R}^n)$ because f is a multiple of a bump function, hence the restriction of Tf to any bounded set is in L^1. At a positive distance from the support of f, we can use the realization (49) and write

$$Tf(x) = \int K(x,y)\, f(y)\, dy = \int [K(x,y) - K(x,y')] f(y)\, dy,$$

where y' is some fixed point in the support of f. Then by (48)

$$|Tf(x)| \leq A|x|^{-n-\gamma}$$

for large x, which shows that $Tf \in L^1(\mathbf{R}^n)$.

Because of this, and motivated by the translation-invariant case (see §3.1.3 above), we shall first consider the special case of those T that satisfy the *special cancellation conditions*:

$$\int_{\mathbf{R}^n} Tf(x)\, dx = \int_{\mathbf{R}^n} T^*f(x)\, dx = 0, \tag{53}$$

whenever f is a smooth function of compact support with

$$\int_{\mathbf{R}^n} f(x)\, dx = 0.$$

3.2.2 In this section, we decompose a restrictedly bounded operator into a dyadic sum. The almost-orthogonality of the summands will then follow (in §3.2.5) from the special cancellation conditions (53).

We consider variants of the operators S_j and $\Delta_j = S_j - S_{j-1}$ that arose in the Littlewood-Paley decomposition treated in §4.1 of the previous chapter. Here

$$S_j f = f * \Phi_{2^{-j}},$$

and we restrict the support of Φ itself (instead of restricting the support of $\widehat{\Phi}$, as was done previously).

To be precise, we choose Φ to be a smooth function supported in the unit ball $|x| < 1$, with $\int \Phi\, dx = 1$. Taking

$$\Psi(x) = \Phi(x) - 2^{-n}\Phi(x/2),$$

we note that $\int \Psi\, dx = 0$. We let

$$\Phi_{2^{-j}}(x) = 2^{nj}\Phi(2^j x), \qquad \Psi_{2^{-j}}(x) = 2^{nj}\Psi(2^j x),$$

and define S_j and Δ_j by

$$S_j(f) = f * \Phi_{2^{-j}}, \qquad \Delta_j = S_j - S_{j-1},$$

so that $\Delta_j(f) = f * \Psi_{2^{-j}}$.

We observe that if $f \in \mathcal{S}$ then $S_j f \to f$ as $j \to +\infty$, where the convergence is in the topology of $\mathcal{S}$. We also note that if $f \in \mathcal{S}'$, then $\{S_j f\}$ is a bounded collection in $\mathcal{S}'$, and $S_j f \to f$ in $\mathcal{S}'$ as $j \to \infty$. From this it follows that, whenever $f \in \mathcal{S}$,

$$Tf = \lim_{j\to+\infty} (S_j T S_j) f.$$

We also claim that $\lim_{j\to-\infty} S_j T S_j(f) = 0$ in $\mathcal{S}'$, whenever $f \in \mathcal{S}$. To see this, observe first that when $f \in \mathcal{S}$, then

$$(S_j f)\widehat{\ }(\xi) = \widehat{f}(\xi)\widehat{\Phi}(2^{-j}\xi).$$

So $(S_j f)(x) = f^{(j)}(2^j x)\cdot 2^{nj}$, with

$$(f^{(j)})\widehat{\ }(\xi) = \widehat{f}(2^j\xi)\widehat{\Phi}(\xi).$$

Thus $\{f^{(j)}\}$ is a bounded sequence in $\mathcal{S}$, as $j \to -\infty$. Therefore by (51) (and the remarks following it), we have that

$$\|TS_j f\|_{L^2} \le A' 2^{nj/2},$$

from which our assertion follows.

Hence

$$T = \sum (S_j T S_j - S_{j-1} T S_{j-1}), \tag{54}$$

with the sum converging in $\mathcal{S}'$, when applied to an element of $\mathcal{S}$. It therefore suffices to prove the boundedness in the L^2 norm of

$$\sum_{j=m_1}^{m_2} (S_j T S_j - S_{j-1} T S_{j-1}) = \sum_{j=m_1}^{m_2} \Delta_j T S_j + \sum_{j=m_1}^{m_2} S_{j-1} T \Delta_j,$$

with bounds independent of m_1 and m_2.

We shall consider separately each of the last two sums and write first $\sum \Delta_j T S_j = \sum T_j$, where

$$T_j = \Delta_j T S_j.$$

3.2.3 The kernels of the T_j. Our first claim is that each T_j is an operator with a smooth kernel, that is

$$T_j f(x) = \int K_j(x,y)\, f(y)\, dy,$$

where K_j satisfies

$$|K_j(x,y)| \le A \cdot 2^{nj} \min\{1, (2^j|x-y|)^{-n-\gamma}\} \tag{55}$$

and, more generally,

$$|\partial_x^\alpha \partial_y^\beta K_j(x,y)| \le A_{\alpha,\beta} \cdot 2^{(n+|\alpha|+|\beta|)j} \min\{1, (2^j|x-y|)^{-n-\gamma})\}. \tag{56}$$

The K_j also satisfy the crucial cancellation conditions:

$$\begin{aligned} &\int K_j(x,y)\, dy = 0, && \text{for each fixed } x \in \mathbf{R}^n, \text{ and} \\ &\int K_j(x,y)\, dx = 0, && \text{for each fixed } y \in \mathbf{R}^n. \end{aligned} \tag{57}$$

3.2.4 We first prove (55) and (56). Like the previous construction, these assertions apply to any operator that satisfies the hypotheses of our theorem (and do not require the special conditions (53)).

We note that whenever $f \in \mathcal{S}$, and Φ, Ψ are C^∞ functions of compact support, then

$$[T(f * \Phi) * \Psi](x) = \int_{\mathbf{R}^n} \langle T(\Phi^y), \tilde{\Psi}^x\rangle\, f(y)\, dy, \tag{58}$$

where $\Phi^y(u) = \Phi(u-y)$, and $\tilde{\Psi}^x(u) = \bar{\Psi}(x-u)$. To verify (58), we observe that

$$f * \Phi = \int_{\mathbf{R}^n} \Phi^y\, f(y)\, dy,$$

while $(\alpha * \Psi)(x) = \langle \alpha, \tilde{\Psi}^x\rangle$ for any tempered distribution α.

Now (58) implies that

$$K_j(x,y) = \langle T(\Phi^y_{2^{-j}}), \tilde{\Psi}^x_{2^{-j+1}}\rangle, \tag{59}$$

where $\Phi^y_{2^{-j}} = (\Phi_{2^{-j}})^y$ and $\tilde{\Psi}^x_{2^{-j+1}} = (\tilde{\Psi}_{2^{-j+1}})^x$. However, $\Phi^y_{2^{-j}}$ is clearly a bump function associated to a ball of radius 2^j, while by the same token the L^2 norm of $\tilde{\Psi}^x_{2^{-j+1}}$ is majorized by a constant multiple of $2^{nj/2}$. Thus if we apply the restricted boundedness condition (51) to $\phi^{R,x_0} = \Phi^y_{2^{-j}}$, we get

$$|K_j(x,y)| \le A \cdot 2^{nj}.$$

For the second inequality in (55), we note that it suffices to prove it when $|x-y| \ge 2\cdot 2^{-j}$. This condition ensures that the supports of $\phi = \Phi^y_{2^{-j}}$ and $\psi = \tilde{\Psi}^x_{2^{-j+1}}$ are disjoint, and so we can rewrite (59), using the kernel representation (49). The result is

$$K_j(x,y) = \int\int \phi(v)\, K(u,v)\, \psi(u)\, du\, dv.$$

Here we exploit the fact that $\int \psi(u)\,du = 0$, and replace the above integral by

$$\int\!\!\int \phi(v)\,[K(u,v) - K(x,v)]\,\psi(u)\,du\,dv.$$

Recalling that ψ is supported in the ball $|u-x| \le 2^{-j-1}$ while ϕ is supported in the ball $|v-y| \le 2^{-j}$, and using the estimates (48), we get that

$$|K(u,v) - K(x,v)| \le A \cdot 2^{j\gamma} \cdot |x-y|^{-n-\gamma},$$

from which the second inequality in (55) follows directly. The proof of the differentiated version (56) is in reality a consequence of (55), once we observe that, for instance, replacing $K_j(x,y)$ by $\partial_y^\beta K_j(x,y)$ has the effect of substituting for Φ a similar function, multiplied by the factor $2^{j|\beta|}$.

3.2.5 The special cancellation conditions. It is easy to see that, given the estimates (56), the cancellation conditions (57) are equivalent to the statements

$$\int T_j^*(f)\,dx = \int T_j(f)\,dx = 0, \tag{60}$$

for all smooth functions f having compact support.

In proving (60), it is useful to make the following simple observations. First, if $f \in L^1$ then, by Fubini's theorem, $\Delta_j(F)$ has integral zero, because $\Delta_j(F) = F * \Psi_{2^{-j}}$ and $\int \Psi = 0$. However, $\Delta_j(F)$ may not have integral zero if (for instance) $F \in L^2$, because $\Delta_j(F)$ may not be integrable on $\mathbf{R}^n$.

Nevertheless, if F satisfies the special conditions

$$|F(x)| \le A(1+|x|)^{-n},$$

and $|F(x-y)-F(x)| \le A|y|^\gamma(1+|x|)^{-n-\gamma}$ for $|x| \ge 2|y|$, with some $\gamma > 0$, we then can conclude that $\int \Delta_j(F)\,dx = 0$. Indeed, we need only write

$$(\Delta_j F)(x) = \int_{\mathbf{R}^n} [F(x-y) - F(x)]\Psi_{2^{-j}}(y)\,dy,$$

and note that Fubini's theorem is applicable: $\int_{\mathbf{R}^n} \Delta_j(F) = 0$ because the indicated double integral (on $\mathbf{R}^n \times \mathbf{R}^n$) converges.

Turning now to the proof of (60), we first look at $T_j(f) = \Delta_j T S_j(f)$, and note that since f is a C^∞ function with compact support, $S_j(f)$ is a multiple of a "bump function", and so $TS_j(f) \in L^2(\mathbf{R}^n)$. However, looking outside the support of $S_j(f)$ (via the integral representation (49)), we see that we can write

$$TS_j(f) = F_1 + F_2,$$

where $F_1 \in L^2$ with compact support (and hence $F_1 \in L^1$), while F_2 satisfies the special conditions given above. By the remarks above, we see that $\int \Delta_j T S_j(f) = 0$, and the second condition in (60) is verified.

To show that $\int S_j^* T^* \Delta_j^*(f) = 0$, it suffices to see that

$$T^*\Delta_j^*(f)$$

is integrable and has integral zero. But this is a consequence of the second‡ special cancellation condition in (53), which we assumed, since $\int \Delta_j^*(f) = 0$.

3.2.6 Almost orthogonality. We next observe that if T_j are operators with kernels $K_j(x,y)$ satisfying (56) and (57), then we automatically have that

$$\begin{cases} \|T_i^*T_j\| & \le 2^{-\gamma'|i-j|}, \quad \text{and} \\ \|T_iT_j^*\| & \le 2^{-\gamma'|i-j|}, \end{cases} \tag{61}$$

for every fixed exponent γ', with $0 \le \gamma' < \gamma$.

Indeed, let $k_{ij}(x,y)$ denote the kernel of $T_i^*T_j$, and consider first the case when $i \ge j$. Then

$$\begin{aligned} k_{ij}(x,y) &= \int \bar{K}_i(z,x)K_j(z,y)\,dz \\ &= \int \bar{K}_i(z,x)[K_j(z,y) - K_j(x,y)]\,dz, \end{aligned}$$

in view of the cancellation condition (57).

Because of the estimates (56), we observe that

$$\begin{aligned} |K_j(z,y) - K_j(x,y)| \le A \cdot 2^{(n+\gamma')j}&|z-x|^{\gamma'} \\ &\times \min\{1, (2^j|x-y|)^{-n-\gamma} + (2^j|z-y|)^{-n-\gamma}\}. \end{aligned}$$

Using this, together with the estimate

$$|K_i(z,x)| \le A \cdot 2^{ni}(1+2^i|z-x|)^{-n-\gamma},$$

shows easily that $k_j(x,y)$ is majorized by

$$A \cdot 2^{\gamma'(j-i)} \cdot 2^{nj}[1+2^j|x-y|]^{-n-\gamma},$$

if $\gamma' < \gamma$. We now need only recall the lemma in §2.4.1 to see that as a result

$$\|T_i^*T_j\| \le A' \cdot 2^{\gamma'(j-i)},$$

‡ The first cancellation condition will be used in the parallel argument mentioned in §3.2.7 below.

which is the first inequality in (61) when $i \geq j$. The proof for $i \leq j$ is similar except here we replace

$$\int \bar{K}_i(z,x)\, K_j(z,y)\, dz$$

by

$$\int [\bar{K}_i(z,x) - \bar{K}_i(y,x)] K_j(z,y)\, dz.$$

The situation for $T_i T_j^*$ is also similar, but here the kernels K_i and K_j are replaced by their adjoints. Altogether then, (61) is proved, and the almost-orthogonality theorem in §2.1 therefore insures that

$$\sum_j \Delta_j T S_j = \sum_j T_j$$

is bounded in L^2.

3.2.7 A completely parallel argument (starting with §3.2.3) also works for

$$\sum_j S_{j-1} T \Delta_j,$$

and so our theorem is now proved for those T that satisfy the special cancellation assumption (53).

3.3 Proof of the theorem: second part. To deal with operators that do not satisfy the assumptions (53) requires different ideas, and here the space BMO plays a decisive role. Indeed, we shall first see that, if T is an operator satisfying the hypotheses of our theorem, then:

1. $T(c)$ can be defined for constant functions c.
2. Under this extension, we have $T(1) \in$ BMO.
3. Condition (53) for T^* holds exactly when $T(1) = 0$.

Analogous statements hold for T^*.

3.3.1 The role of BMO. We begin by noting the existence of a bound A so that whenever ϕ^{R,x_0} is a normalized bump function for the ball $B(x_0,R)$, then $T(\phi^{R,x_0}) \in$ BMO with

$$\|T(\phi^{R,x_0})\|_{\mathrm{BMO}} \leq A. \tag{62}$$

The proof of (62) is a reprise of the simple argument given in Chapter 4, §4.1 Let $\tilde{B} = B(x_1,R_1)$ be any other ball, and let $\tilde{B}_2$, $\tilde{B}_3$ be (respectively) its double and treble, i.e., $\tilde{B}_2 = B(x_1,2R_1)$, $\tilde{B}_3 = B(x_1,3R_1)$. Fix a C^∞ function θ, with $\theta(x) = 1$ for $|x| \leq 2$, and $\theta(x) = 0$ for $|x| \geq 3$; write

$$\phi^{R,x_0} = f_1 + f_2,$$

where $f_1(x) = \phi^{R,x_0}(x) \cdot \theta([x - x_1]/R)$, $f_2(x) = \phi^{R,x_0}(x) \cdot [1 - \theta([x - x_1]/R)]$.

Observe that f_1 is, up to a bounded multiplicative constant, a normalized bump function for either the ball $B(x_0,R)$, or the ball $\tilde{B}_3$, whichever has the smaller radius. Thus by assumption (51)

$$\int_{\tilde{B}} |Tf_1|^2\, dx \leq \|Tf_1\|_{L^2}^2 \leq A \cdot \min\{R^n, (3R_1)^n\} \leq A' \cdot |\tilde{B}|.$$

Next, since f_2 is supported in ${}^c\tilde{B}_2$, then whenever $x \in \tilde{B}$, the kernel realization (49) applies to $T(f_2)$, giving

$$Tf_2(x) = \int K(x,y)\, f_2(y)\, dy.$$

We define the constant

$$c_{\tilde{B}} = \int K(x_1,y)\, f_2(y)\, dy;$$

then

$$|Tf_2(x) - c_{\tilde{B}}| \leq \int_{|y-x_1| \geq 2R_1} |K(x,y) - K(x_1,y)|\, dy,$$

and this integral is in turn majorized by a constant, if we apply (48), because $|x - x_1| \leq R_1$. Therefore $|Tf_2(x) - c_{\tilde{B}}| \leq A'$ for $x \in \tilde{B}$, and then altogether

$$\int_{\tilde{B}} |T(\phi^{R,x_0}) - c_{\tilde{B}}|^2\, dx \leq A|\tilde{B}|,$$

for all balls $\tilde{B}$, establishing (62).

Now if η is any fixed smooth function with compact support so that $\eta(0) = 1$, then we set $\eta_\varepsilon(x) = \eta(\varepsilon x)$, and note that the family $\{T(\eta_\varepsilon)\}_{\varepsilon>0}$ is uniformly in BMO. Now BMO is the dual of H^1 (see Chapter 4, §1.2) and hence, as is well-known, is weakly compact in the dual topology. Thus every sequence $\{\eta_{\varepsilon_k}\}$ has a subsequence whose image under T converges weakly. We test this convergence on functions f that are smooth, have compact support, and have integral zero, using the identity

$$\langle f, T(\eta_\varepsilon)\rangle = \langle T^*(f), \eta_\varepsilon\rangle.$$

From this it is clear that whatever limit (call it a) we extract from a subsequence of the $T(\eta_\varepsilon)$, then[†]

$$\langle f, a\rangle = \langle T^*(f), 1\rangle.$$

This shows that the limit a is independent of the subsequence, and we are justified in setting

$$a = T(1) = \lim_{\varepsilon\to 0} T(\eta_\varepsilon).$$

We have thus established that

$$\langle f, T(1)\rangle = \int_{\mathbf{R}^n} T^*(f)\, dx, \tag{63}$$

with a similar statement when the roles of T and T^* are reversed.

[†] Recall that $T^*f \in L^1$, since $\int f\, dx = 0$.

3.3.2 The operators S_a. From these considerations, it follows that the general case of the theorem will be reduced to the special case previously established, once we find suitable operators S and S', so that

$$S(1) = T(1) \qquad \text{and} \qquad (S')^*(1) = T^*(1),$$

while

$$(S^*)(1) = 0 \qquad \text{and} \qquad S'(1) = 0.$$

Thus our goal can be restated as follows. Let $a \in \mathrm{BMO}$; we seek an operator S_a so that:

(i) $S_a : L^2 \to L^2$ is a bounded operator,

(ii) S_a has a kernel realization that satisfies (48) and (49), and

(iii) $S_a(1) = a$, while $(S_a)^*(1) = 0$.

Once the S_a have been found (for every $a \in \mathrm{BMO}$), we can take $S = S_a$ and $S' = (S_b)^*$, where $a = T(1)$ and $b = T^*(1)$.

It turns out that the operators that we need have already occurred in §1, when we examined symbols in the forbidden class $S_{1,1}$. More particularly, for a given $a \in \mathrm{BMO}$, we take S_a to be the paraproduct operator given by (13), namely

$$S_a(f) = \sum_{j=-\infty}^{\infty} \Delta_{j+\ell}(a) \cdot S_j(f). \tag{64}$$

Here $S_j(f) = f * \Phi_{2^{-j}}$ and $\Delta_j = S_j - S_{j-1}$. The Φ that will be used now is as in §4.1 of the previous chapter—its Fourier transform has compact support; we use it rather than the Φ that occurred in §3.2 of this chapter (which had compact x-support).

More precisely, $(S_j(f))\widehat{\ }$ is supported in $|\xi| \le 2^{j+1}$, while $\Delta_{j+\ell}(a)\widehat{\ }$ is supported in $2^{j+\ell-1} \le |\xi| \le 2^{j+\ell+1}$. Hence if $f \in L^2$ and $a \in \mathrm{BMO}$, the L^2 function $\Delta_{j+\ell}(a) \cdot S_j(f)$ will have its Fourier transform supported in the convolution of the sets

$$\{\xi : |\xi| \le 2^{j+1}\} \quad \text{and} \quad \{\xi : 2^{j+\ell-1} \le |\xi| \le 2^{j+\ell+1}\};$$

this set is contained in $\{\xi : 2^{j+\ell-2} \le |\xi| \le 2^{j+\ell+2}\}$, provided that ℓ is sufficiently large ($\ell \ge 3$ will do); we now fix such an ℓ.

As a result, the sum above is essentially an orthogonal decomposition: we have that each ξ is contained in the support of the Fourier transform of at most four terms in the sum (64).

3.3.3 We shall now establish properties (i)–(iii) of the operator S_a; we assume that $a \in \mathrm{BMO}$ and $f \in L^2$. Observe first that each term in the series (64) is in L^2, because $\Delta_{j+\ell}(a)$ is bounded (see Chapter 4, §4.3.3) and because $S_j(f) \in L^2$. Next, we claim that the series (64) converges in the L^2 norm, and that the resulting operator

$$f \mapsto S_a(f)$$

is bounded on L^2. In view of the fact that the supports of (the Fourier transforms of) the terms in the sum (64) have a bounded number of overlaps, it suffices by Plancherel's theorem to show that

$$\sum_j \|\Delta_{j+\ell}(a) S_j(f)\|_{L^2}^2 \le A^2 \|f\|_{L^2}^2. \tag{65}$$

If we define the measure $d\mu$ on $\mathbf{R}_+^{n+1} = \{(x,t) : x \in \mathbf{R}^n, t > 0\}$ by

$$d\mu = \sum_j |\Delta_{j+\ell}(a)(x)|^2 \, dx \cdot \delta_{2^{-j}}(t),$$

where $\delta_{2^{-j}}(t)$ is the unit Dirac mass at $t = 2^{-j}$, then (65) can be rewritten as

$$\int_{\mathbf{R}_+^{n+1}} |(f * \Phi_t)(x)|^2 \, d\mu(x,t) \le A^2 \int_{\mathbf{R}^n} |f(x)|^2 \, dx. \tag{66}$$

However

$$d\mu = \sum_j |a * \Psi_{2^{-j-\ell}}(x)|^2 \cdot \delta_{2^{-j}}(t) = \sum_j |a * \Phi'_{2^{-j}}(x)|^2 \cdot \delta_{2^{-j}}(t),$$

with $\Phi' = \Psi_{2^{-\ell}}$, and therefore by Chapter 4, §4.3.2, $d\mu$ is a Carleson measure. As a result, the inequality (66) is a consequence of the principal majorization satisfied by such measures, namely (24) in Chapter 2, §2.4. Thus (65) is proved, and with it the boundedness of $f \mapsto S_a(f)$ on L^2. If we set

$$S_a^{(m)}(f) = \sum_{|j| \le m} \Delta_{j+\ell}(a) S_j(f),$$

we see also that the $S_a^{(m)}$ converge strongly in L^2 to S_a as $m \to \infty$.

Next, observe that the kernel $k_j(x,y)$ of the operator

$$f \mapsto \Delta_{j+\ell}(a) S_j(f)$$

is given by $\Delta_{j+\ell}(a)\Phi_{2^{-j}}(x-y)$. Since

$$\|\Delta_{j+\ell}(a)\|_{L^\infty} = \|a * \Psi_{2^{-j-\ell}}\|_{L^\infty} \le A'$$

(by Chapter 4, §4.3.3), we get that

$$|k_j(x,y)| \le A\cdot 2^{nj}(1+2^j|x-y|)^{-n-1}.$$

Thus the kernel $K_m(x,y)$ of $S_a^{(m)}$ is majorized by

$$\sum_j 2^{nj}(1+2^j|x-y|)^{-n-1} \le A|x-y|^{-n};$$

letting $m \to \infty$ shows that the operator S_a is represented by a kernel K that has the bound

$$|K(x,y)| \le A|x-y|^{-n}.$$

A similar argument gives the estimates

$$|\partial_x^\alpha \partial_y^\beta K(x,y)| \le A|x-y|^{-n-|\alpha|-|\beta|}$$

for all α and β, proving (48) and (49).

We now turn to the fact that $S_a(1) = a$. This is obvious on a formal level because $S_j(1) \equiv 1$, while $\sum_j \Delta_{j+\ell} = I$, by (25a) of the previous chapter.

For the actual proof, we require the following formula for $S_a(1)$. Let B be any fixed ball containing the origin. Suppose ϕ is a smooth function of compact support with $\phi(x) = 1$ for all $x \in B$. Then, up to an additive constant,

$$S_a(1)(x) = S_a(\phi)(x) + \int [K(x,y) - K(0,y)]\,(1-\phi(y))\,dy \tag{67}$$

for $x \in B$. Notice that the introduction of the "infinite constant"

$$\int K(0,y)\,(1-\phi(y))\,dy$$

makes the integral converge, by virtue of the estimate (49) for the kernel K.

To verify (67), take η_ε as in §3.3.1 and observe that, up to an additive constant,

$$S_a(\eta_\varepsilon)(x) = S_a(\phi\eta_\varepsilon)(x) + \int [K(x,y) - K(0,y)]\,(1-\phi(y))\,\eta_\varepsilon(y)\,dy$$

for $x \in B$. Now the right side converges to the right side of (67), as $\varepsilon \to 0$, while, as we have observed, $S_a(\eta_\varepsilon)$ converges weakly to what was defined to be $S_a(1)$; of course, additive constants are zero elements in BMO. Therefore (67) is established.

This also shows that the parallel assertion for $S_a^{(m)}(1)$, with K replaced by K_m. Then we let $m \to \infty$, recalling that $K_m(x,y)$ converges dominatedly to K, while $S_a^{(m)}(\phi)$ converges to $S_a(\phi)$ in the L^2 norm. However

$$S_a^{(m)}(1) = \sum_{|j|\le m} \Delta_{j+\ell}(a) \to a$$

in the weak sense for BMO (see Chapter 4, §6.20). Altogether then $S_a(1) = a$ (modulo an additive constant) on each fixed ball, and so the identity $S_a(1) = a$ is proved.

Finally, to show that $(S_a)^*(1) = 0$, we observe first that the analogues of (67) hold also for $(S_a)^*$ and $(S_a^{(m)})^*$. This reduces the proof that $(S_a)^*(1) = 0$ in the sense of BMO, to that of the fact $(S_a^{(m)})^*(1) = 0$. However

$$(S_a^{(m)})^*(1) = \sum_{|j|\le m} S_j^*(\Delta_{j+\ell}^* a),$$

and S_j^*, $\Delta_{j+\ell}^*$ are both multiplier operators with multipliers $\bar\Phi(2^{-j}\xi)$ and $\bar\Psi(2^{-(j+\ell)}\xi)$ respectively. These have disjoint supports, from which the assertion (iii) in §3.3.2 follows.

3.3.4 We conclude the proof of the theorem by writing

$$T = T' + S_a + (S_b)^*,$$

where $a = T(1)$ and $b = T^*(1)$. Note, therefore, that $T'(1) = 0$ and also $(T')^*(1) = 0$, so the special cancellation conditions (57) hold for T'. Thus the case previously treated in §3.2 guarantees the boundedness of T', and hence the boundedness of T.

3.3.5 It is useful to observe that we have proved that the norm of the operator T is controlled entirely by the bounds that occur in (48), (51), and (52).

3.4 Cancellation conditions for Calderón-Zygmund kernels. We now turn to the second question posed at the beginning of §3. We are given a function $K(x,y)$, defined for $x \ne y$, that is assumed to satisfy the basic inequalities (48). We seek the cancellation conditions to be imposed on K so that there exists a bounded operator

$$T : L^2(\mathbf{R}^n) \to L^2(\mathbf{R}^n)$$

for which K is its kernel, in the sense of (49).

Our observation in the translation-invariant case (see §3.1.2) leads us to consider the quantities

$$I_{\varepsilon,N}(x) = \int_{\varepsilon<|x-y|<N} K(x,y)\,dy, \tag{68}$$

and the condition that is suggested is that these are bounded in some average sense; similar conditions would also be required for the adjoint kernel $\bar{K}(y,x)$:

$$I^*_{\varepsilon,N}(x) = \int_{\varepsilon<|x-y|<N} \bar{K}(y,x)\,dy. \tag{68*}$$

That this is indeed the case is guaranteed by the following theorem.

THEOREM 4. *Suppose $K(x,y)$ is given for $x \neq y$, and satisfies (48). Then there exists a bounded operator $T : L^2 \to L^2$ so that (49) holds if and only if there is an $A > 0$ so that*

$$\int_{|x-x_0|<N} |I_{\varepsilon,N}(x)|^2\,dx \le A \cdot N^n \qquad \textit{for all } \varepsilon,\ N,\ \textit{and } x_0, \tag{69}$$

with a similar condition for $I^*_{\varepsilon,N}(x)$.

We shall see that this result is actually a rather direct consequence of Theorem 3. Before we come to the proof we make two additional remarks:

(i) In the translation-invariant case, $I_{\varepsilon,N}(x)$ is obviously independent of x, so that condition (69) is then the same as (44).

(ii) The condition (69) can be replaced by an L^q average, for $1 \le q < \infty$; however, in general, the L^∞ analogue, namely

$$|I_{\varepsilon,N}(x)| \le A \qquad \text{for all } \varepsilon,\ N, \text{ and } x,$$

is not a necessary condition; see §5.17.

3.4.1 We first prove the necessity of the condition (69). We note that it suffices to show that

$$\int_{|x-x_0|\le N/2} |I_{\varepsilon,N}(x)|^2\,dx \le A \cdot N^n \qquad \text{for all } \varepsilon,\ N, \text{ and } x_0, \tag{70}$$

because every ball of radius N can be covered by a bounded number of balls of radius $N/2$.

Define the truncated kernel K_ε by $K_\varepsilon(x,y) = K(x,y)$ if $|x-y| > \varepsilon$, and $K_\varepsilon(x,y) = 0$ if $|x-y| \le \varepsilon$. Let

$$T_\varepsilon(f)(x) = \int K_\varepsilon(x,y)\,f(y)\,dy,$$

which of course is well-defined for each $f \in L^2(\mathbf{R}^n)$. According to the proposition in Chapter 1, §7.1, the operators T_ε are then uniformly bounded in the L^2 norm. We shall exploit this fact by comparing $I_{\varepsilon,N}(x)$ with $T_\varepsilon(\chi^{N,x_0})(x)$, when $|x-x_0| < N/2$. Here χ^{N,x_0} is the characteristic function of the ball $\{y : |y-x_0| < N\}$.

Now a moment's reflection shows that the symmetric difference

$$\{y : |x-y| < N\}\Delta\{y : |y-x_0| < N\}$$

is contained in $\{y : N/2 < |x-y| < 3N/2\}$, provided $|x-x_0| < N/2$. However,

$$I_{\varepsilon,N}(x) - T_\varepsilon(\chi^{N,x_0})(x) = \int_{\varepsilon<|x-y|<N} K(x,y)\,dy - \int_{\substack{\varepsilon<|x-y| \\ |y-x_0|<N}} K(x,y)\,dy.$$

As a result,

$$|I_{\varepsilon,N}(x) - T_\varepsilon(\chi^{N,x_0})(x)| \le \int_{N/2<|x-y|<3N/2} |K(x,y)|\,dy,$$

which means that $|I_{\varepsilon,N}(x) - T_\varepsilon(\chi^{N,x_0})(x)| \le A'$, when $|x-x_0| \le N/2$. Applying the uniform L^2 boundedness of the T_ε to the function χ^{N,x_0} then establishes (70), and thus (69). A similar argument proves the condition for $I^*_{\varepsilon,N}$.

3.4.2 To prove the converse, let $K^\varepsilon(x,y)$ be a smooth truncation of K, given by

$$K^\varepsilon(x,y) = \eta\left(\frac{x-y}{\varepsilon}\right) K(x,y),$$

where $\eta(x)$ is C^∞, $\eta(x) = 0$ if $|x| \le 1/2$, and $\eta(x) = 1$ if $|x| \ge 1$. Let

$$T^\varepsilon(f)(x) = \int K^\varepsilon(x,y)\,f(y)\,dy.$$

We observe first that the kernels $K^\varepsilon(x,y)$ satisfy the basic inequalities (48) uniformly in ε.

We prove next that the operators T^ε satisfy the restricted boundedness conditions (51) uniformly in ε, as well as the similar conditions for the $(T^\varepsilon)^*$.

In fact, if ϕ^{R,x_0} is a normalized bump function for the ball $B(x_0,R)$, and $|x-x_0| \le 2R$, then

$$\begin{aligned}(T^\varepsilon\phi^{R,x_0})(x) &= \int K^\varepsilon(x,y)\,\phi^{R,x_0}(y)\,dy \\ &= \int K^\varepsilon(x,y)[\phi^{R,x_0}(y) - \phi^{R,x_0}(x)]\chi^{3R,x}(y)\,dy \\ &\quad + \phi^{R,x_0}(x)\int_{|y-x|<3R} K^\varepsilon(x,y)\,dy.\end{aligned}$$

The first integral is easily estimated by

$$A\int_{|y-x|\le 3R} |x-y|^{-n}\cdot R^{-1}\cdot |x-y|\,dy \le A',$$

while the quantity $\int_{|y-x|<3R} K^{\varepsilon}(x,y)\,dy$ differs from

$$I_{\varepsilon,R}(x) = \int_{\varepsilon<|y-x|<3R} K(x,y)\,dy$$

by a bounded amount. Thus (70) implies that

$$\int_{|x-x_0|\le 2R} |T^{\varepsilon}(\phi^{R,x_0})(x)|^2\,dx \le A\cdot R^n.$$

For $|x-x_0|\ge 2R$, we use the obvious estimate

$$|T^{\varepsilon}(\phi^{R,x_0}(x))| \le \frac{A\cdot R^n}{|x-x_0|^n}.$$

Altogether then

$$\|T^{\varepsilon}(\phi^{R,x_0})\|_{L^2} \le A\cdot R^{n/2},$$

and (51) holds uniformly in ε. A similar estimate for $(T^{\varepsilon})^*$ then shows, using Theorem 3, that the norms of the operators T^{ε} are uniformly bounded. Thus selecting an appropriate sequence $\varepsilon_k \to 0$, the operators T^{ε_k} converge weakly (in L^2) to a bounded operator T on L^2. Since the $K^{\varepsilon_k}(x,y)$ converge dominatedly (and pointwise) to $K(x,y)$, it follows that the representation (49) holds, and the proof of the theorem is complete.

3.5 Examples and remarks.

3.5.1 The theory behind the results contained in Theorems 3 and 4 had as its impetus the "commutators" of Calderón, as well as the extensions of these ideas needed to achieve the ultimate goal of proving the L^2 boundedness of the Cauchy operator on Lipschitz domains. We illustrate this by presenting here the first step in this development.

Let T by a standard pseudo-differential operator of order 1, i.e., an operator as given in the previous chapter, whose symbol belongs to S^1. Suppose M_a denotes the multiplication operator

$$(M_a f)(x) = a(x)\cdot f(x),$$

where $a(x)$ is some function. If a is smooth and all of its derivatives are bounded, then a corresponds to a symbol in S^0, and thus, by the composition theorem of the previous chapter, the commutator

$$[T, M_a] = TM_a - M_aT$$

is an operator of order 0, and hence is bounded on $L^2(\mathbf{R}^n)$.

The question that arises is: what are the minimal smoothness conditions that can be imposed on a so that $[T, M_a]$ is still bounded on $L^2(\mathbf{R}^n)$? An obvious necessary condition on a occurs if we take $T = \partial/\partial x_j$, $j=1,\dots,n$. Then since

$$[\partial/\partial x_j, M_a] = M_{\partial a/\partial x_j},$$

it is clear that we need to require

$$\frac{\partial a}{\partial x_j} \in L^{\infty}(\mathbf{R}^n), \qquad j=1,\dots,n, \tag{71}$$

or, equivalently, the existence of an A so that

$$|a(x)-a(y)| \le A|x-y|, \qquad \text{for all } x,y\in\mathbf{R}^n. \tag{71'}$$

The remarkable fact is that this condition is also sufficient.

COROLLARY. *Suppose T is a pseudo-differential operator of order 1 and a satisfies* (71). *Then the operator $[T, M_a]$ falls under the scope of Theorem 3, and hence is bounded from $L^2(\mathbf{R}^n)$ to itself.*

According to the argument in §5.3.2 of the previous chapter, the operator T can be written as

$$T = T_0 + T_1\frac{\partial}{\partial x_1} + \cdots + T_n\frac{\partial}{\partial x_n},$$

where the T_j are pseudo-differential operators of order zero. The statement of the corollary is then reduced to the special case $T = T_j(\partial/\partial x_j)$, (or $T = T_0$). Now by (23) of the previous chapter, the operator T has a kernel K that satisfies

$$|\partial_x^{\alpha}\partial_y^{\beta} K(x,y)| \le A\cdot |x-y|^{-n-1-|\alpha|-|\beta|};$$

moreover, the kernel of $[T, M_a]$ is easily seen to be $K(x,y)[a(y)-a(x)]$. It is then an easy matter to verify that, as a consequence of (71), the estimate (48) holds for the kernel of $[T, M_a]$.

Next, it is easy to check that

$$\begin{aligned}[T_j(\partial/\partial x_j), M_a] &= [T_j, M_a](\partial/\partial x_j) + T_j[\partial/\partial x_j, M_a] \\ &= [T_j, M_a](\partial/\partial x_j) + T_j M_{\partial a/\partial x_j}.\end{aligned}$$

Again, by (23) of the previous chapter, since T_j is of order 0, the kernel of $[T_j, M_a]$ is bounded by $A|x-y|^{1-n}$. We can now test $[T, M_a]$ on bump functions ϕ^{R,x_0}. First, as is easily seen

$$\int_{|x-x_0|\le 2R} |([T_j, M_a](\partial/\partial x_j)\phi^{R,x_0})(x)|^2\,dx \le A\cdot R^n.$$

A similar inequality holds for the term $T_j M_{\partial a/\partial x_j}\phi^{R,x_0}$, because T_j is bounded on L^2. For $|x-x_0| \geq 2R$, using the size of the kernel of $[T, M_a]$, we easily get

$$|([T, M_a]\phi^{R,x_0})(x)| \leq \frac{A \cdot R^n}{|x-x_0|^n}.$$

Altogether then

$$\|[T, M_a]\phi^{R,x_0}\|_{L^2} \leq A \cdot R^{n/2},$$

which is (51) for the operator in question. A similar argument gives the corresponding inequality for the adjoint, and the corollary is proved.

There is also an analogous result for "higher" commutators; see §4.2.

3.5.2 Suppose T is a linear mapping from $\mathcal{S}$ to $\mathcal{S}'$, with an associated kernel satisfying (48) and (49). Then, whenever f is a bounded function on $\mathbf{R}^n$ that is also smooth, we can define Tf as an element of $\mathcal{S}'$ modulo an additive constant. In fact, in analogy with (67), if B is any fixed ball, with (say) $0 \in B$, and ϕ is a smooth function with compact support, so that $\phi(x) = 1$ on B, then the identity

$$Tf = T(\phi f) + \int [K(x,y) - K(0,y)](1-\phi(y))\, f(y)\, dy$$

defines Tf (as a distribution) on B, and thereby (up to an additive constant) on all of $\mathbf{R}^n$.

Now let $f(x) = e^{2\pi i x\cdot\xi}$. Then the condition that $Tf = T(e^{2\pi i x\cdot\xi})$ is in BMO for all ξ and

$$\sup_{\xi\in\mathbf{R}^n} \|T(e^{2\pi i x\cdot\xi})\|_{\mathrm{BMO}} \leq A < \infty \tag{72}$$

can be shown to imply the restricted boundedness conditions (51) for T; see §5.19.

3.5.3 For some purposes it is desirable to relax further the hypotheses in Theorem 3 that require that both T and T^* be restrictedly bounded. These can be combined in the statement that T is "weakly bounded"; however, then we must add as further assumptions the condition that both $T(1)$ and $T^*(1)$ be in BMO. See §5.18.

4. Appendix: The Cauchy integral

4.1 Statement of the theorem. The L^2 theory of singular integrals presented in §3 had at its source the problem of the Cauchy integral. One way of stating the question that was considered is as follows. Suppose γ is (say) a

rectifiable curve in the complex plane. We let C_γ denote the Cauchy integral over γ, i.e., the operator

$$(C_\gamma f)(z) = \frac{1}{2\pi i}\int_\gamma \frac{f(\zeta)\, d\zeta}{\zeta - z}, \qquad z \notin \gamma.$$

The problem, then, is that of understanding the L^2 (and other) boundedness properties of the induced mapping

$$f \mapsto C_\gamma(f)|_\gamma,$$

when this is suitably defined.

We recall two facts from the classical theory. First, the above problem reduces largely to that of the behavior of the corresponding Hilbert transform H_γ, defined by

$$(H_\gamma f)(x) = \text{p.v.}\int_{-\infty}^{\infty} \frac{f(y)\,\gamma'(y)\, dy}{\gamma(x)-\gamma(y)}, \tag{73}$$

once we have chosen a suitable parametrization $x \mapsto \gamma(x)$ of the curve γ.‡ Second, when γ is sufficiently smooth (e.g., $\gamma \in \Lambda_{1+\epsilon}$), then the boundedness properties of H_γ are easily reduced to those of the usual Hilbert transform H (for which $\gamma(x) \equiv x$).

Turning to the situation with less smoothness, the key case to consider is when γ is merely Lipschitz, and more particularly when γ is a Lipschitz graph; that is

$$\gamma(x) = x + iA(x), \qquad \text{where} \quad A' = a \in L^\infty(\mathbf{R}); \tag{74}$$

thus $\gamma'(x) = 1 + ia(x)$.

THEOREM. *The operator*

$$(H_\gamma)(f) = p.v.\int_{-\infty}^{\infty} \frac{f(y)[1+ia(y)]\, dy}{x-y+i[A(x)-A(y)]} \tag{75}$$

is bounded from $L^2(\mathbf{R}^1)$ *to itself, under the assumption* (74).

Note that, as far as the *conclusion* of the L^2 boundedness is concerned, the factor $1 + ia(y)$ is irrelevant; it will, however, be used in the proof. Hence, omitting this factor, we get an operator for which the theory of Chapter 1 applies (in particular, it satisfies (10) and (29) of that chapter); thus, H_γ is also bounded on L^p, $1 < p < \infty$.

4.2 To get an insight into this operator, we consider the expansion

$$\frac{1}{x-y+i[A(x)-A(y)]} = \frac{1}{x-y}\cdot\sum_{k=0}^{\infty}(-i)^k\left[\frac{A(x)-A(y)}{x-y}\right]^k. \tag{76}$$

‡ We omit the usual factor of π in our definition of H_γ.

Now $|A(x) - A(y)| \le M|x-y|$, if $M = \|A'\|_{L^\infty} = \|a\|_{L^\infty}$; thus the above series converges if $M < 1$. This suggests that our theorem should have a close connection with the boundedness of the operators C_k, defined by

$$(C_k f)(x) = \text{p.v.} \int_{-\infty}^{\infty} \frac{[A(x)-A(y)]^k}{(x-y)^{k+1}} f(y)\, dy.$$

PROPOSITION. *Each operator C_k is bounded on $L^2(\mathbf{R}^1)$. Moreover, there exist constants c and L so that*

$$\|C_k\|_{2,2} \le cL^k M^k. \tag{77}$$

The proof of this is a direct consequence of Theorem 3 in §3.2. Note that C_0 is the Hilbert transform, and C_1 is a variant of the commutator in §3.5.1. To show that C_k, $k \ge 1$, satisfy the hypotheses of Theorem 3, one first notes that the kernel verifies the differential inequalities (48) (with $\gamma = 1$). Next, if $f = \phi^{R,x_0}$ is a bump function supported in the ball $B = B(x_0, R)$, then the L^2 control of $C_k(f)$ outside the double of B can be obtained entirely by the size of the kernel. Now for x inside the double of B, we write

$$\begin{aligned}(C_k f)(x) &= \frac{1}{k}\int_{-\infty}^{\infty} [A(x)-A(y)]^k f(y) \frac{d}{dy}(x-y)^{-k}\, dy \\ &= \frac{-1}{k}\int_{-\infty}^{\infty} \left[\frac{A(x)-A(y)}{x-y}\right]^k \frac{df}{dy}\, dy \\ &\quad - \frac{1}{k}\int_{-\infty}^{\infty} \frac{d}{dy}[A(x)-A(y)]^k \frac{f(y)\, dy}{(x-y)^k}.\end{aligned}$$

The first integral can be estimated crudely, since f is a bump function. An L^2 estimate for the second integral reduces essentially to the case of C_{k-1}, since A' is a bounded function. An inductive argument then establishes our assertion and the bound (77).

The expansion (76) then shows that if $M = \|A'\|_{L^\infty}$ is sufficiently small ($\|A'\|_{L^\infty} < \min(L^{-1}, 1)$), then the assertion of the theorem is valid for such A. The proof of the full theorem can be deduced from the case just proved, or by a refinement of the formulation of Theorem 3. We shall not pursue these possibilities here, but instead describe an independent approach that is quite simple in its conception.

4.3 A pseudo-orthogonal basis. The proof we have in mind is a generalization of an argument for the usual Hilbert transform H that is based on the observation that H is "almost" diagonal when expressed as a matrix in the usual Haar basis of L^2; this basis is described in Chapter 4, §4.5.4.†

The situation for H_γ leads us to the following generalization of the Haar basis and its orthogonality. We fix a bounded function $b : \mathbf{R} \to \mathbf{C}$ with $\operatorname{Re} b(x) \ge c > 0$ for all x.‡ Now define an "inner product" $\langle \cdot, \cdot \rangle_b$ by

$$\langle f, g\rangle_b = \int_{\mathbf{R}} f(x)\, g(x)\, b(x)\, dx, \qquad f, g, \in L^2(\mathbf{R}).$$

† Alternatively, one can follow the construction below, taking $b(x) \equiv 1$.

‡ In our application, b will be $1 + ia$, $a = A'$.

For each dyadic interval I, we define a modified Haar function $\mathfrak{h}_I$ as follows. Writing I_ℓ and I_r for the left and right halves of the interval I, we want $\mathfrak{h}_I$ to be supported on I, constant on each of I_ℓ, I_r, and to satisfy

$$\langle \mathfrak{h}_I, 1\rangle_b = 0, \qquad \langle \mathfrak{h}_I, \mathfrak{h}_I\rangle_b = 1. \tag{78}$$

This can be achieved by taking

$$\mathfrak{h}_I = b(I)^{-1/2}\left\{\left[\frac{b(I_\ell)}{b(I_r)}\right]^{1/2}\chi_{I_r} - \left[\frac{b(I_r)}{b(I_\ell)}\right]^{1/2}\chi_{I_\ell}\right\},$$

with a fixed choice of the square roots; here $b(E) = \int_E b\, dx$.

Note that the first part of (78) guarantees that

$$\langle \mathfrak{h}_I, \mathfrak{h}_J\rangle_b = 0, \qquad I \ne J.$$

Thus $\{\mathfrak{h}_I\}$ is "orthonormal" with respect to $\langle\cdot,\cdot\rangle_b$. Moreover, if

$$(A_I f)(x) = \chi_I(x)\cdot b(I)^{-1}\int_I f(x)\, b(x)\, dx$$

then

$$[A_{I_\ell} + A_{I_r} - A_I]f = \langle f, \mathfrak{h}_I\rangle_b \mathfrak{h}_I. \tag{79}$$

The key assertion is the following.

PROPOSITION. *If $f \in L^2(\mathbf{R}^1)$, then $f = \sum_I a_i \mathfrak{h}_I$ with $a_I = \langle f, \mathfrak{h}_I\rangle_b$, where this sum converges in L^2. Moreover,*

$$c^{-1}\|f\|_{L^2}^2 \le \sum_I |a_I|^2 \le c\|f\|_{L^2}^2 \tag{80}$$

for some absolute constant c.

To prove this proposition, we define the "conditional expectation" operator $\mathfrak{E}_k$ by

$$\mathfrak{E}_k(f) = \sum_{|I| = 2^{-k}} A_I(f).$$

We also set $\mathfrak{D}_k = \mathfrak{E}_{k+1} - \mathfrak{E}_k$; by (79), we have

$$\mathfrak{D}_k(f) = \sum_{|I|=2^{-k}} a_I \mathfrak{h}_I. \tag{81}$$

Now $f = \lim_{k\to\infty} \mathfrak{E}_k(f)$ almost everywhere,† while $\mathfrak{E}_k(f) \to 0$ as $k \to -\infty$, thus $f = \sum_k \mathfrak{D}_k(f)$ pointwise, which is our expansion. On the other hand, because

† This follows from the classical result on differentiation of the integral, together with our hypotheses on b.

they have disjoint supports, the terms in (81) are pairwise orthogonal (in the usual sense).

Thus, to prove the right side of the inequality (80), it suffices to show that

$$\sum_{k=-\infty}^{\infty} \|\mathfrak{D}_k(f)\|_{L^2}^2 \le c'\|f\|_{L^2}^2. \tag{82}$$

Now $\mathfrak{E}_k(f) = E_k(fb)/E_k(b)$, where E_k is the usual dyadic conditional expectation. Writing $\Delta_k = E_{k+1} - E_k$, we get

$$\mathfrak{D}_k(f) = \frac{E_{k+1}(fb)}{E_{k+1}(b)} - \frac{E_k(fb)}{E_k(b)} = \frac{\Delta_k(fb)}{E_k(b)} - \frac{\Delta_k(b)}{E_k(b)E_{k+1}(b)}E_k(fb).$$

Because $\operatorname{Re} E_k(b) \ge c > 0$, the estimate in (82) is reduced to a corresponding estimate for

$$\sum_k \|\Delta_k(fb)\|_{L^2}^2 + \sum_k \int |\Delta_k(b)|^2 \, E_k(|f|)^2 \, dx.$$

The orthogonality properties of the usual Haar basis and the Δ_k show that the first term is bounded by a multiple of $\|f\|_{L^2}^2$.

For the second term, we write

$$u(x,t) = (|f| * \Phi_t)(x),$$

where Φ is the characteristic function of the unit ball, and let $d\mu(x,t)$ be the measure on

$$\mathbf{R}_+^2 = \{(x,t) \in \mathbf{R}^2 : x \in \mathbf{R}, t > 0\}$$

given by

$$d\mu(x,t) = 2^k \cdot |(\Delta_k b)(x)|^2 \, dx \, dt, \quad 2^{-k} \le t < 2^{1-k}.$$

Then the second term is majorized by

$$\int_{\mathbf{R}_+^2} |u(x,t)|^2 \, d\mu(x,t),$$

which is turn is dominated by $\|f\|_{L^2}^2$, once we show that $d\mu$ is a Carleson measure.‡

If I is a dyadic interval with $|I| = 2^{-k}$ then

$$d\mu(T(I)) \le d\mu(I \times I) = \sum_{\ell \ge k} \int_I |(\Delta_\ell b)(x)|^2 \, dx \le \|b\chi_I\|_{L^2}^2 \le c2^{-k},$$

which is the Carleson condition on I. If J is an arbitrary interval, then J is contained in the union of two dyadic intervals of comparable length, so

‡ For this last implication, see (24) in Chapter 2, §2.4.

$d\mu(T(J)) \le c|J|$. Thus we have shown that $d\mu$ is a Carleson measure and, with it, the inequality

$$\sum_I |a_I|^2 \le c\|f\|_{L^2}^2.$$

The converse inequality follows directly from the direct inequality: let $g = \overline{fb}$, $a_I' = \langle g, \mathfrak{h}_I \rangle_b$, and note that

$$\sum_I a_I a_I' = \langle f, g \rangle_b = \|fb\|_{L^2}^2.$$

4.4 Proof of the theorem. Having set down the facts concerning the basis $\{\mathfrak{h}_I\}$, we now observe that the theorem stated in §4.1 is an immediate consequence of the following two lemmas.

LEMMA 1. *Suppose T is a linear operator, defined on finite linear combinations of the $\{\mathfrak{h}_I\}$. Let $\mu_{IJ} = \langle T\mathfrak{h}_I, \mathfrak{h}_J \rangle_b$. Assume that*

$$\sup_I \sum_J |\mu_{IJ}| < \infty, \qquad \textit{and} \qquad \sup_J \sum_J |\mu_{IJ}| < \infty.$$

Then T extends to a bounded linear operator on $L^2(\mathbf{R}^1)$.

LEMMA 2. *If $T = H_\gamma$ is as in* (75) *and $b = 1 + ia$, then the hypotheses of Lemma 1 are satisfied.*

To prove the first lemma, one notes that (by the Proposition in §4.3)

$$\|Tf\|_{L^2}^2 \le c \sum_J |\langle Tf, \mathfrak{h}_J \rangle_b|^2 = \sum_J \left| \sum_I \mu_{IJ} a_I \right|^2,$$

where $a_I = \langle f, \mathfrak{h}_I \rangle_b$. Now by Schwarz's inequality

$$\sum_J \left| \sum_I \mu_{IJ} a_I \right|^2 \le \sum_J \left\{ \left(\sum_I |\mu_{IJ}| \right) \left(\sum_I |\mu_{IJ}| \cdot |a_I|^2 \right) \right\}.$$

The right side of this inequality can be estimated by

$$\left(\sup_J \sum_I |\mu_{IJ}| \right) \left(\sup_I \sum_J |\mu_{IJ}| \right) \sum_I |a_I|^2,$$

and another application of the proposition proves Lemma 1.

To prove the second lemma, one remarks first that in our case

$$\langle H_\gamma f, g \rangle_b = -\langle f, H_\gamma g \rangle_b;$$

hence $\mu_{IJ} = -\mu_{JI}$, and it suffices to estimate μ_{IJ} when $|I| \le |J|$. We restrict ourselves to this case. Write J^* for the double of J and δ_{IJ} for the minimum of the distances of I to the center or one of the endpoints of the interval J. Then we have

$$|\mu_{IJ}| \le c|J|^{-1/2}|I|^{3/2}(\delta_{IJ} + |I|)^{-1}, \qquad \text{if} \quad I \cap J^* \ne \emptyset$$

and

$$|\mu_{IJ}| \le c|J|^{3/2}|I|^{1/2}(\delta_{IJ} + |I|)^{-2}, \qquad \text{if} \quad I \cap J^* = \emptyset.$$

These come about by straightforward computation from the following two easily obtained estimates for $H_\gamma(\mathfrak{h}_I)$:

$$|(H_\gamma \mathfrak{h}_I)(x)| \le c|x - x_I|^{-2}|I|^{3/2}, \qquad x \notin I^*,$$

and

$$|(H_\gamma \mathfrak{h}_I)(x)| \le c|I|^{-1/2} \log\left(\frac{c|I|}{|x - x_I^*|}\right), \qquad x \in I^*,$$

where x_I is the center of I, and x_I^* is the point nearest to x among the center and two end-points of I.

The last estimate is a consequence of the fact that

$$(H_\gamma \chi_I)(x) = \text{p.v.} \int_I \frac{\gamma'(y)\,dy}{\gamma(x) - \gamma(y)}, \qquad \gamma(x) = x + iA(x);$$

choosing an appropriate branch of the logarithm function, we have

$$\frac{d}{dy} \log[\gamma(x) - \gamma(y)] = \frac{-\gamma'(y)}{\gamma(x) - \gamma(y)}.$$

The estimates then allow one to verify the conditions required of the μ_{IJ}, and thus the applicability of Lemma 1.

For a more detailed version of this argument, see Coifman, P. Jones, and Semmes [1989].

4.5 Some further references. The theorem for small M, proved by a different method than that of §4.2, goes back to Calderón [1977a]; the theorem for the first commutator C_1 originates in Calderón [1965]. The full result (for any M) was obtained by Coifman, McIntosh, and Y. Meyer [1982]. They showed that if A is normalized by assuming $M = 1$, then the norm of C_k grows at most polynomially in k. David [1984] showed how the result for large M can be deduced from that for small M by a localization method.

Another approach is to prove an analogue of Theorem 3 (more precisely, its variant in §5.18) where, instead of a "$T(1)$" theorem, we have a "$T(b)$" theorem. For this, see David, Journé, and Semmes [1985]; also David [1991].

A substantial further literature exists concerning the Cauchy integral and related topics. For this, the reader may consult the accounts in Murai [1988], Y. Meyer and Coifman [1991], David [1991].

5. Further results

A. Method of orthogonality

5.1 Among the simplest examples of the method of orthogonality described in §2.1 is the proof of the L^2 boundedness of the Hilbert transform H, and more generally the Riesz transforms R_j; this can be based on the identities $H^* = -H$, $H^2 = -I$, and $R_j^* = -R_j$, $\sum_{j=1}^n R_j^2 = -I$ (here I denotes the identity operator). To implement this idea, it is useful to proceed directly from the definition of the R_j in terms of the generalized Cauchy-Riemann equations ((56) in Chapter 3).

Let f be a real-valued function on $\mathbf{R}^n$; we may argue formally as follows. Let $u_0(x,t), u_1(x,t), \dots, u_n(x,t)$ be the Poisson integrals of

$$f, R_1(f), \dots, R_n(f),$$

respectively, and set $F_j(t) = \int_{\mathbf{R}^n} [u_j(x,t)]^2\,dx$. Then for $j \ge 1$

$$\frac{1}{2}\frac{\partial F_j}{\partial t} = \int_{\mathbf{R}^n} u_j \frac{\partial u_j}{\partial t}\,dx = \int_{\mathbf{R}^n} u_j \frac{\partial u_0}{\partial x_j}\,dx = -\int_{\mathbf{R}^n} \frac{\partial u_j}{\partial x_j} u_0\,dx.$$

Hence

$$\frac{1}{2}\sum_{j=1}^n \frac{\partial F_j}{\partial t} = \int_{\mathbf{R}^n} \frac{\partial u_0}{\partial t} u_0\,dx = \frac{1}{2}\frac{\partial F_0}{\partial t}.$$

Integrating this in t gives $\sum_{j=1}^n F_j(0) = F_0(0)$; i.e.,

$$\sum_{j=1}^n \int_{\mathbf{R}^n} [(R_j f)(x)]^2\,dx = \int_{\mathbf{R}^n} f(x)^2\,dx.$$

This argument can be made rigorous by applying it to $f \in C_0^\infty$. For such f, the $u_j(x,t)$ and all their partial derivatives are $O(1 + |x| + t)^{-n}$.

5.2 The method of TT^* can also be used to prove various maximal inequalities for L^2. We illustrate this by considering

$$(M_\Phi f)(x) = \sup_{t>0} |(f * \Phi_t)(x)|, \qquad \Phi(x) = (1 + |x|)^{-N},$$

for some fixed $N > n$. We show that $\|M_\Phi f\|_{L^2} \le A\|f\|_{L^2}$, which of course implies the corresponding result for the standard maximal operator M.

Pick a strictly positive, finite-valued measurable function $x \mapsto t(x)$ on $\mathbf{R}^n$; the following estimates will be independent of the choice of stopping time $t(x)$. Let $(Tf)(x) = (f * \Phi_{t(x)})(x)$. Then $(TT^*f)(x) = \int K(x,y)\,f(y)\,dy$, with

$$K(x,y) = \tilde{\Phi}_{t(x),t(y)}(x - y), \qquad \tilde{\Phi}_{t_1,t_2} = \Phi_{t_1} * \Phi_{t_2}.$$

Note that $\tilde{\Phi}_{t_1,t_2} \le c(\Phi_{t_1} + \Phi_{t_2})$; this can be proved by rescaling to the case $t_1 = 1$, $t_2 \le 1$. The result is that, for nonnegative f, we have $TT^*f \le c(Tf + T^*f)$. It follows that

$$\|T^*\|^2 = \|T\|^2 \le 2c\|T\|;$$

hence $\|T\| \le 2c$. Taking the supremum over all stopping times $t(x)$ then gives the desired result.

The idea of this argument goes back to Kolmogorov and Seliverstov [1925]; it was elaborated in Paley [1930], Stein [1970a], Nagel, Stein, and Wainger [1979]. It can sometimes also be used to prove L^p inequalities for $p < 2$. For this see, e.g., Nagel and Stein [1984].

Note that the above example shows graphically that, while TT^* and T^*T have the same norm, dealing with the former is relatively straightforward, even though the latter appears to be intractable.

5.3 Theorem 1 also applies when there are *infinitely* many operators T_j that satisfy the bounds (22). The conclusion is then that $\sum T_j$ converges strongly with (of course) $\|\sum T_j\| \leq A$. The strong convergence of the sum is a consequence of Theorem 1 and the following Hilbert space lemma.

Let f_j be a sequence in a Hilbert space H so that, for all choices of scalars ε_j, $|\varepsilon_j| \leq 1$, with $\varepsilon_j \neq 0$ for only finitely many j, we have $\|\sum \varepsilon_j f_j\| \leq A$, for some fixed A. Then $\sum f_j$ converges in H.

To prove the lemma, assume the contrary. Thus there exists a $\delta > 0$ and sequences N_k, N'_k, $N_k < N'_k < N_{k+1}$, so that if $\Delta_k = \sum_{N_k}^{N'_k} f_j$, then $\|\Delta_k\| \geq \delta$. Now our assumptions also ensure that, for all sequences ε_k with $|\varepsilon_k| \leq 1$, we have $\|\sum \varepsilon_k \Delta_k\| \leq A$, which implies that $\sum \|\Delta_k\|^2 < \infty$, yielding a contradiction.

5.4 The special case of Theorem 1 in which the operators T_j are self-adjoint and mutually commutative was considered first in Cotlar [1955]. It turns out that in that case there is a simple proof. Using a common spectral resolution reduces matters to the case when the underlying Hilbert space is one-dimensional. One then needs the following observation: If t_j are (real) scalars with $|t_i t_j| \leq \gamma^2(i-j)$, then $\sum_i |t_i| \leq A = \sum_i \gamma(i)$.

To see this, note that $|t_i|^{1/2} \leq \gamma(i-j)|t_j|^{-1/2}$. Choosing j so that $|t_j| = \max_i |t_i|$, we have $\sum_i |t_i|^{1/2} \leq A|t_j|^{-1/2}$. Thus $\sum_i |t_i| \leq |t_j|^{1/2} \sum_i |t_i|^{1/2} \leq A$.

5.5 (i) The hypothesis (22) of Theorem 1 can be generalized to $\|T_i^* T_j\| \leq \gamma^2(i,j)$ and $\|T_i T_j^*\| \leq \gamma^2(i,j)$ where γ is a nonnegative function on $\mathbf{Z} \times \mathbf{Z}$ with $\sup_i \sum_j \gamma(i,j) \leq A$. Again, the conclusion is that $\|\sum T_j\| \leq A$. The proof given in §2.2 remains essentially unchanged.

(ii) An analogue for integrals is as follows. Suppose that $t \mapsto T_t$ is a (measurable) function from $\mathbf{R}$ to bounded operators on L^2, with $\|T_t^* T_s\| \leq \gamma^2(t,s)$ and $\|T_t T_s^*\| \leq \gamma^2(t,s)$ for some nonnegative function γ on $\mathbf{R} \times \mathbf{R}$ with $\sup_t \int_R \gamma(t,s)\, ds \leq A$. The conclusion is that, with $T = \int T_t\, dt$ suitably defined, we have $\|T\| \leq A$.

See Calderón and Vaillancourt [1972].

B. Exotic and forbidden symbols

5.6 (a) The boundedness of T_a on Λ_γ, $\gamma > 0$, for $a \in S^0_{1,1}$, which was proved in §1.3, extends to the class of Lipschitz spaces $\Lambda^{p,q}_\gamma$ appearing in §7.8 of the previous chapter. Indeed, if $a \in S^m_{1,1}$, then T_a maps $\Lambda^{p,q}_\gamma$ to $\Lambda^{p,q}_{\gamma-m}$, whenever $\gamma - m > 0$.

(b) A similar result holds for the Sobolev spaces L^p_k when $k > 0$: If $a \in S^m_{1,1}$ then T_a maps L^p_k to L^p_{k-m}, provided $k - m > 0$. This can be generalized to the analogous assertion for the spaces $F^{p,q}_\gamma$ discussed in Chapter 6, §7.9.

See Y. Meyer [1980], Runst [1985], Bourdaud [1988].

5.7 The L^2 boundedness of T_a for $a \in S^0_{\rho,\rho}$, $0 \leq \rho < 1$, proved in §2.5, is sharp, in that such a conclusion fails for $a \in S^0_{\rho,\delta}$ whenever $\delta > \rho$. The requisite counterexamples can be constructed by a method similar to that of §1.2.

We illustrate this in the case $n = 1$, $\rho = 1/2$, $\delta > 1/2$. Choose a $\psi \in C_0^\infty$ with $\operatorname{supp}\psi \subset [1/2, 3/2]$ so that $\psi(\xi) = 1$ for $\psi \in [3/4, 5/4]$. Let

$$\psi^k(\xi) = \psi\left(\frac{\xi - k^2}{2k+1}\right), \quad k = 1, 2, \ldots.$$

Note that the ψ^k have disjoint supports and are uniformly in $S^0_{1/2,0} \subset S^0_{1/2,\delta}$. Then if $|\alpha_k| \leq ck^{2\delta}$, we have that

$$a(x,\xi) = \sum_{k=1}^{\infty} e^{-2\pi i \alpha_k x} \psi^k(\xi) \in S^0_{1/2,\delta}.$$

Now let $f \in L^2$, $\|f\|_{L^2} = 1$, with $\operatorname{supp}\widehat{f} \subset [3/4, 5/4]$ and set

$$f_{m_1,m_2}(x) = \left(\sum_{k=m_1}^{m_2-1} e^{2\pi i k^2 x}\right) f(x).$$

Since the spectra of $e^{2\pi i k^2 x} f(x)$ are disjoint, we have

$$\|f_{m_1,m_2}\|_{L^2} = (m_2 - m_1)^{1/2}.$$

However, if α_k are such that $k^2 - \alpha_k = m_1^2$, then

$$\|T_a(f_{m_1,m_2})\|_{L^2} = \|(m_2 - m_1) e^{2\pi i m_1^2 x} f(x)\|_{L^2(dx)} = m_2 - m_1.$$

To see that such a choice of α_k is permissible, notice that

$$\alpha_k = k^2 - m_1^2 \leq 2k(k - m_1) < 2k(m_2 - m_1) \leq 2k^{2\delta}, \quad m_1 \leq k < m_2,$$

if we take $m_2 \leq \lfloor m_1^{2\delta - 1} \rfloor$. Letting $m_1 \to \infty$ (and hence $(m_2 - m_1) \to \infty$) gives a contradiction.

5.8 The symbolic calculus treated in §3 of the previous chapter extends, without essential change, to the exotic symbols. We assume that $\delta \le \rho$ and $0 \le \delta < 1$. The statement of the composition rule is then: If $a \in S^{m_1}_{\rho,\delta}$, $b \in S^{m_2}_{\rho,\delta}$, then there exists a $c \in S^{m_1+m_2}_{\rho,\delta}$ with $T_c = T_a \circ T_b$. Moreover, we have that

$$c - \sum_{|\alpha|<N} \frac{(2\pi i)^{-|\alpha|}}{\alpha!} \partial_\xi^\alpha a \cdot \partial_x^\alpha b \in S^{m_1+m+2-N(\rho-\delta)}_{\rho,\delta}$$

for each $N > 0$.

Note that when $\rho = \delta$, each symbol $\partial_\xi^\alpha a \cdot \partial_x^\alpha b$ is in $S^{m_1+m_2}_{\rho,\delta}$, and the thrust of the assertion is limited to the conclusion that $c \in S^{m_1+m_2}_{\rho,\delta}$. A generalization of these results follows in §5.9 below.

5.9 Suppose we are given a fixed pair of functions Φ and ϕ on $\mathbf{R}^n \times \mathbf{R}^n$ that satisfy:

$$1 \le \Phi(x,\xi) \le C(1+|\xi),$$

$$c(1+|\xi|)^{\varepsilon-1} \le \phi(x,\xi) \le 1, \quad \text{for some } \varepsilon > 0,$$

and

$$\Phi\phi \ge 1,$$

as well as the regularity conditions

$$|\nabla_x \phi| \le C, \quad |\nabla_\xi \phi| \le C\frac{\phi}{\Phi}, \qquad |\nabla_x \Phi| \le C\frac{\Phi}{\phi}, \quad |\nabla_\xi \Phi| \le C.$$

In addition, we assume that

$$\frac{\Phi(x,\xi)}{\phi(x,\xi)} \approx \frac{\Phi(x,\eta)}{\phi(x,\eta)} \qquad \text{whenever } |\xi-\eta| < |\xi|/2.$$

We then define the symbol class $S^{M,m}_{\Phi,\phi}$ to consist of those $a \in C^\infty(\mathbf{R}^n \times \mathbf{R}^n)$ for which

$$|\partial_x^\beta \partial_\xi^\alpha a(x,\xi)| \le A_{\alpha,\beta} \Phi^{M-|\alpha|} \phi^{m-|\beta|}$$

for all multi-indices α and β. On the basis of these assumptions, we can assert the following results.

(i) If $a \in S^{0,0}_{\Phi,\phi}$, then T_a is bounded on $L^2(\mathbf{R}^n)$.

(ii) If $a \in S^{M_1,m_1}_{\Phi,\phi}$, $b \in S^{M_2,m_2}_{\Phi,\phi}$, then there is a $c \in S^{M_1+M_2,m_1+m_2}_{\Phi,\phi}$ with $T_c = T_a \circ T_b$. Moreover,

$$c - \sum_{|\alpha|<N} \frac{(2\pi i)^{-|\alpha|}}{\alpha!} \partial_\xi^\alpha a \cdot \partial_x^\alpha b \in S^{M_1+M_2-N,m_1+m_2-N}_{\Phi,\phi},$$

for all $N > 0$.

(iii) If $a \in S^{M,m}_{\Phi,\phi}$, then there is a symbol $a^* \in S^{M,m}_{\Phi,\phi}$ with $(T_a)^* = T_{a^*}$. In fact, we have that $a^* - \overline{a} \in S^{M-1,m-1}_{\Phi,\phi}$.

Note that the assertions above include the results for the $S_{\rho,\delta}$ symbols when $\delta \le \rho$, $0 \le \rho \le 1$, and $0 \le \delta < 1$; we take $\Phi = (1+|\xi|)^\rho$, $\phi = (1+|\xi|)^{-\delta}$. Details may be found in R. Beals and C. Fefferman [1974]. Some important applications are described below.

5.10 The "classical" Gårding inequality, which is an assertion about a pseudo-differential operator with positive symbol, is a direct consequence of the calculus set forth in the previous chapter. It can be stated as follows: Suppose a is a standard symbol of order m (i.e., $a \in S^m_{1,0}$), and assume that $a(x,\xi) \ge 0$ everywhere. Then for every $\varepsilon > 0$, there is a c_ε so that

$$\mathrm{Re}\langle T_a f, f\rangle + \varepsilon \|f\|^2_{L^2_{m/2}} + c_\varepsilon \|f\|^2_{L^2_{(m-1)/2}} \ge 0.$$

The conclusion can be extended to its "sharp" version, which states that ε may be taken to be zero. In fact, one has the even stronger result that, under the above assumptions on the symbol a, there is a constant $c > 0$ so that

$$\mathrm{Re}\langle T_a f, f\rangle + c\|f\|^2_{L^2_{(m-\sigma)/2}} \ge 0$$

with $\sigma = 2$.

For $\sigma = 1$ the conclusion goes back to Hörmander [1966b]. In this case, a proof can be obtained directly from the calculus described in §5.9. Indeed, if we take $m = 1$, replace a by $a+1$, and set

$$\Phi(x,\xi) = (1+|\xi|^2)^{1/4} \cdot a(x,\xi)^{1/2}, \qquad \phi(x,\xi) = (1+|\xi|^2)^{-1/4} \cdot a(x,\xi)^{1/2},$$

we see that the hypotheses of §5.9 are (essentially) satisfied. We also note that $a^{1/2} \in S^{1/2,1/2}_{\Phi,\phi}$. Thus $T^*_{a^{1/2}} T_{a^{1/2}} = T_a + E$, where $E = T_b$ for some $b \in S^{0,0}_{\Phi,\phi}$, and the result follows. For details, see R. Beals and C. Fefferman [1974].

The stronger conclusion (with $\sigma = 2$) can be proved using the same calculus, but it requires a more refined argument involving induction on the dimension, as well as a stopping-time decomposition of the (x,ξ) space. See C. Fefferman and Phong [1978].

5.11 The question of solvability of a partial differential operator of principal type is reducible to the following pseudo-differential equation:

$$\frac{\partial u}{\partial t} - a_t(x,D)u = f, \quad x \in \mathbf{R}^n, \quad t \in \mathbf{R}.$$

Here, for each t, $a_t(x,\xi)$ is a standard symbol of order 1, i.e., $a \in S^1_{1,0}$, and it varies smoothly with t. We use the common notation $a(x,D)$ for the pseudo-differential operator T_a having symbol a. We make the crucial assumption that, for each (x,ξ), the real function $t \mapsto a_t(x,\xi)$ does not change sign as t varies. We can then conclude that the above equation is always locally solvable in $\mathbf{R}^{n+1}$.

The proof depends on the following factorization. One can always find a pair Φ,ϕ that satisfy the hypotheses of §5.9, and symbols $b_t \in S^{0,0}_{\Phi,\phi}$, $c \in S^{1,1}_{\Phi,\phi}$, and $d_t \in S^{0,0}_{\Phi,\phi}$, with b_t and d_t varying smoothly in t, with $b_t(x,\xi) \ge 0$ everywhere, and so that

$$a_t(x,\xi) = b_t(x,\xi)c(x,\xi) + d_t(x,\xi).$$

For the details and applications to local solvability, see R. Beals and C. Fefferman [1973] and [1974]; important relevant ideas are in Nirenberg and Treves [1971]. In particular, the latter shows that if a_t is real analytic, then it has a factorization of the above type in which the symbols b_t, c, and d may be taken to be of the standard kind.

5.12 We now describe the L^p theory for operators with symbols belonging to the exotic class $S^m_{\rho,\delta}$.

(a) To begin with, let $a(\xi)$ be a symbol that does not depend on x. Suppose $0 < \rho \le 1$, $m = (\rho-1)n/2$, and $a \in S^m_\rho$. If K is the distribution given by $\widehat{K}(\xi) = a(\xi)$ then, away from the origin, K agrees with a locally integrable function $K(x)$ that satisfies

$$\int_{|x|\ge 2|y|^\rho} |K(x-y) - K(x)|\,dx \le A, \qquad \text{if } |y| \le 1.$$

(b) An example arises when $n = 1$, $\rho = 1/2$, and $a(\xi) = e^{i|\xi|^{1/2}}/|\xi|^{1/4}$ for large ξ. Then $K(x) = c_1 e^{c_2 i/|x|}/|x| + O(|x|^{-1/2})$ as $x \to 0$. See §1.4.3 above and Chapter 8, §1.4.2.

(c) Under the hypotheses set forth in (a), the operator T_a is of weak-type (1,1) and is bounded from $L^p(\mathbf{R}^n)$ to itself, for $1 < p < \infty$.

(d) Under the same hypotheses,

$$T : H^1(\mathbf{R}^n) \to L^1(\mathbf{R}^n) \qquad \text{and} \qquad T : L^\infty(\mathbf{R}^n) \to \mathrm{BMO}(\mathbf{R}^n).$$

(e) Suppose we now assume that $(\rho-1)n/2 < m \le 0$, rather than $(\rho-1)n/2 = m$. Then $T_a : L^p(\mathbf{R}^n) \to L^p(\mathbf{R}^n)$ if $|p^{-1} - 2^{-1}| \le m/[n(\rho-1)]$.

(f) The above conclusion is sharp: if $0 < \rho < 1$ and $|p^{-1} - 2^{-1}| > m/[n(\rho-1)]$, then there exists an $a(\xi) \in S^m_\rho$ so that T_a is not a bounded operator on $L^p(\mathbf{R}^n)$.

(g) The results in (a), (c), (d), and (e) extend to operators having symbols in the class $S^m_{\rho,\delta}$ (which may depend on x) when $0 < \rho \le 1$, $\delta \le \rho$, and $\delta < 1$. In particular, when $a \in S^m_{\rho,\delta}$, $m = (\rho-1)n/2$, we have that T_a is of weak-type (1,1); also $T_a : H^1 \to L^1$; and $T_a : L^\infty \to \mathrm{BMO}$. Also, if $|p^{-1} - 2^{-1}| \le m/[n(\rho-1)]$, then $T_a : L^p \to L^p$.

(h) Some of these results extend to the limiting case $\rho = \delta = 0$: If $a \in S^{-n/2}_{0,0}$, then $T_a : H^1 \to L^1$ and $T_a : L^\infty \to \mathrm{BMO}$. Also, if $a \in S^m_{0,0}$, $1 < p < \infty$, and $|p^{-1} - 2^{-1}| \le -m/n$, then $T_a : L^p \to L^p$.

To prove (a), one uses Plancherel's theorem and argues as in §4.4.2 of the previous chapter. We let

$$K_j(x) = \int_{\mathbf{R}^n} e^{2\pi i x\cdot\xi}\,\delta(2^{-j}\xi)\,a(\xi)\,d\xi.$$

Then (using that $m = (\rho-1)n/2$) one has

$$\int |K_j(x)|\,dx \le A, \qquad \int |\nabla_x K_j(x)|\,dx \le A2^j, \qquad \int |x|^N |K_j(x)|\,dx \le A2^{-jN\rho},$$

for all $N \ge 0$. Since $K = \sum K_j$, the desired inequality follows easily.

Conclusion (c) is proved in C. Fefferman [1970]. The assertions (d) and (e) are in C. Fefferman and Stein [1972], where (e) is deduced from (d) and the complex interpolation theorem for BMO (or H^1) given in Chapter 5, §5.2. The sharpness of these conclusions, as stated in (f), can be seen when $\rho = 1/2$, $n = 1$, via examples like (b). Indeed, if $a(\xi) = e^{i|\xi|^{1/2}}|\xi|^m$ for large ξ, and $f(x) = |x|^{\alpha-1}$ for small x (with $0 < \alpha < 1$), then $|(T_a f)(x)| \approx |x|^\beta$ as $x \to 0$, where $\beta = 2\alpha - 2m - 3/2$; if $\beta < 0$, one obtains a counterexample from this. See also Hirschman [1956], Wainger [1965].

The case of pseudo-differential operators with variable symbols can be handled by adapting the arguments used in the case of symbols that are independent of x, if one makes use of the fact that

$$T_a = T_{a^0} J, \qquad \text{where} \quad a^0(x,\xi) = a(x,\xi)(1+|\xi|^2)^{-m/2} \in S^0_{\rho,\delta},$$

and $J = T_{(1+|\xi|^2)^{m/2}}$. By Theorem 2, $T_{a^0} : L^2 \to L^2$. See also C. Fefferman [1973a], Stein [1973b], Miyachi [1980a].

Note that the case $\rho = 1$, $\delta < 1$, and $m = 0$ falls under the scope of the usual theory: according to Theorem 2, such operators are bounded on L^2. Moreover, their kernels satisfy the Calderón-Zygmund estimates, as guaranteed by Proposition 1 in §1, and give bounded operators on $L^p(\mathbf{R}^n)$, $1 < p < \infty$.

The analysis of the limiting case $\rho = \delta = 0$ is different in nature from the situation when $0 < \rho$. In part, this difference is accounted for by the fact that, when $\rho > 0$, operators with symbols in $S^m_{\rho,\delta}$ are pseudo-local, in the sense that they are represented by kernels that are smooth away from the diagonal. But when $\rho = 0$, these operators are not pseudo-local. For the proof of the assertions made in (h), see Coifman and Y. Meyer [1978].

5.13 We consider the question as to how many derivatives of symbols in $S^0_{\rho,\rho}$ need to be controlled for the resulting operators to be bounded on L^2. Suppose $0 \le \rho < 1$ and a is a symbol for which

$$|\partial_x^\beta \partial_\xi^\alpha a(x,\xi)| \le A(1+|\xi|)^{\rho(|\beta|-|\alpha|)}, \qquad \text{for } |\alpha| + |\beta| \le N.$$

The proofs in §2.4 and §2.5 in actuality show that $T_a : L^2 \to L^2$, provided

$$N > \max\{4n, 2n/(1-\rho)\}.$$

By using more elaborate arguments, one can reduce this to $N > n$. For the proof and for further results of this kind see Cordes [1975], Coifman and Y. Meyer [1978].

C. Operators with Calderón-Zygmund kernels

5.14 We exhibit examples that show that the various assumptions in Theorems 3 and 4 are not redundant.

First, there is an operator $T : \mathcal{S} \to \mathcal{S}'$ so that both T and T^* are restrictedly bounded (in the sense of (51) and (52)) that does not extend to a bounded operator on L^2; the kernel of T, of course, does not satisfy the estimates (48).

Such an operator can be given by an unbounded multiplier $m(\xi) \in L^2(\mathbf{R}^n)$ that is, however, bounded near the origin; an example is $m(\xi) = (1-|\xi|^2)_+^\delta$ with $-1/2 < \delta < 0$. If we define T by $(\widehat{Tf})(\xi) = m(\xi)\widehat{f}(\xi)$, an easy calculation shows that T verifies the above assertions.

5.15 There exists an operator $T : \mathcal{S} \to \mathcal{S}'$ that is representable by a Calderón-Zygmund kernel satisfying (48) and (49) and is restrictedly bounded, but which does not extend to a bounded operator on L^2; in this case, T^* is not restrictedly bounded.

Indeed, let a be any symbol in the forbidden class $S^0_{1,1}$. Then $T = T_a$ has a Calderón-Zygmund kernel (see §1.1); in the next paragraph, we show that T is restrictedly bounded. The construction in §1.2 then gives the required example.

To see that $\|T_a(\psi)\|_{L^2} \le AR^{n/2}$, $\psi = \phi^{R,x_0}$, one estimates $|(T_a\psi)(x)|$ by

$$A\int_{\mathbf{R}^n} |\widehat{\psi}(\xi)|\, d\xi \le A' \qquad \text{when } |x - x_0| \le 2R,$$

and by

$$\frac{AR^n}{|x - x_0|^n} \qquad \text{when } |x - x_0| \ge 2R.$$

5.16 There is a function $K_0(x)$ on $\mathbf{R}^1$, defined for $x \ne 0$, with $|K_0(x)| \le A|x|^{-1}$, and $\int_{\varepsilon<|x|<N} K_0(x)\,dx = 0$ for all $0 < \varepsilon \le N < \infty$, so that if $K =$ p.v. K_0, the operator $Tf = f * K$ is *not* bounded on $L^2(\mathbf{R}^n)$.

In particular, a converse to the proposition in §3.12 can hold only if one requires some additional regularity of K. Moreover, using this example, one can show that Theorems 3 and 4 both fail if one drops the difference conditions in the assumptions (48).

To construct K_0, we let $\alpha(x) = \sum \alpha_k e^{ikx}$ be an even, bounded, periodic function with the property that $\sum_{|k|<R} \alpha_k$ is unbounded as $R \to \infty$; the construction of such α may be found in Zygmund [1959], Chapter 8. Let $K_0(x) = \alpha(x)/x$. Then K_0 is odd, so the cancellation condition is automatically satisfied. However, if $K = \text{p.v.}K_0$, one verifies that $\widehat{K}$ is unbounded.

One should note, though, that examples of this kind are by necessity limited to one dimension; see Chapter 8, §5.22.

5.17 The hypothesis (69) of Theorem 4 can be modified to read

$$\int_{|x-x_0|<N} |I_{\varepsilon,N}(x)|^q\,dx \le AN^n$$

for any q, $1 \le q < \infty$, with a similar condition on $I^*_{\varepsilon,N}(x)$. However, the corresponding conditions for $q = \infty$, i.e., $|I_{\varepsilon,N}(x)| + |I^*_{\varepsilon,N}(x)| \le A$ for all ε, N, and x, are sufficient but not necessary.

Indeed, if we let T be the Calderón commutator (see §4.2):

$$(Tf)(x) = \text{p.v.}\int_{-\infty}^{\infty} \frac{A(x) - A(y)}{(x-y)^2} f(y)\,dy, \quad A' = a \in L^\infty(\mathbf{R}^n),$$

then $\lim_{\substack{\varepsilon\to 0\\ N\to\infty}} I_{\varepsilon,N}(x) = I(x) = (T1)(x) = \pi(Ha)(x)$, where H is the Hilbert transform. Setting $a(x) = \text{sign}(x)$ gives $I(x) = c_1 \log|x| + c_2$ with $c_1 \ne 0$; this function is, of course, not uniformly bounded.

5.18 The original formulation of Theorem 3 was different, with apparently weaker sufficient conditions: An operator $T : \mathcal{S} \to \mathcal{S}'$, with a kernel satisfying the Calderón-Zygmund estimates (48), is bounded on L^2 exactly when the following two conditions hold:

(a) T is "weakly bounded", that is, there is a constant $A > 0$ so that, for all pairs of normalized bump functions ϕ, ψ, we have $|\langle T\phi^{R,x_0}, \psi^{R,x_0}\rangle| \le AR^n$ for all $R > 0$, $x_0 \in \mathbf{R}^n$.

(b) $T1 \in \text{BMO}$ and $T^*1 \in \text{BMO}$ (this is interpreted as in §3.3).

This theorem is due to David and Journé [1984]. Note that (a) holds whenever T or T^* is restrictedly bounded. We make a few remarks about the differences between this version and the one stated in §3.2.

(i) The proof given in §3.2 and §3.3 also proves the original theorem. Indeed, the argument for T satisfying the special cancellation conditions goes through unchanged, since in reality only the weak boundedness of T was used. Similarly, the argument in §3.3 is based only on the assertion that both $T1$ and T^*1 are in BMO.

(ii) Note that when T is "odd",[†] the weak boundedness condition is automatically satisfied. For such operators, it therefore suffices to verify that $T1 \in \text{BMO}$ (which in this case implies that $T^*1 \in \text{BMO}$).

(iii) However, checking (a) and (b) is, in practice, no easier than verifying the restricted boundedness conditions stated in §3.2. Moreover, the statement of Theorem 3 is conceptually simpler: It does not require of T two conditions of a different character, the second of which involves a notion not obviously related to L^2 boundedness.

5.19 Suppose $T : \mathcal{S} \to \mathcal{S}'$ is represented by a kernel satisfying (48) and (49). As we have remarked in §3.5.2, if f is bounded and smooth, then Tf is a well-defined element of $\mathcal{S}'$ modulo an additive constant. If $\xi \in \mathbf{R}^n$ then $f(x) = e^{2\pi i\xi\cdot x}$ is such a function, and it makes sense to say that $Tf \in \text{BMO}$ (or $T^*f \in \text{BMO}$). Keeping our assumptions on T, we can then say that T is bounded on L^2 exactly when

$$\sup_{\xi\in\mathbf{R}^n} \{\|T(e^{2\pi i\xi\cdot x})\|_{\text{BMO}} + \|T^*(e^{2\pi i\xi\cdot x})\|_{\text{BMO}}\} < \infty. \qquad (*)$$

Since Calderón-Zygmund operators[‡] map L^∞ to BMO, $(*)$ is necessary. To see that it is sufficient, let ϕ^{R,x_0} be a normalized bump function associated to the ball $B = B(x_0, R)$; let B^* be the double of B and let $\tilde{B}$ be some ball having radius R whose distance from B is also R. Write $F = T(\phi^{R,x_0})$. First, representing ϕ^{R,x_0} by its inverse Fourier transform, we see that $\|F\|_{\text{BMO}} \le A$. Hence

$$\int_{B^*} |F(x) - F_{B^*}|^2\,dx \le A|B^*|.$$

[†] This means that there is a kernel K, with $K(x,y) = -K(y,x)$, so that $\langle Tf, g\rangle = (1/2)\int\int K(x,y)\,[f(y)g(x) - f(x)g(y)]\,dy\,dx$ whenever $f, g \in \mathcal{S}$.

[‡] That is, those operators treated by Theorem 3 that are, in fact, bounded on L^2.

However $|F(x)| \leq A'R^n/|x-x_0|^n$ when $x \notin B^*$, so $|F_{\tilde{B}}| \leq A'$. Moreover, $|F_{B^*} - F_{\tilde{B}}| \leq A$, from which it follows that $\|F\|_{L^2} \leq cR^{n/2}$. This shows that T is restrictedly bounded; the same holds for T^*.

5.20 The following generalizes and amplifies the remarks in §1.2.2 and the results in §3.2.

(a) Suppose that Φ and Ψ are in $\mathcal{S}$, and that their spectra are compatible in the sense that there are positive numbers $a < b < c$ so that $\widehat{\Phi}(\xi)$ is supported in $|\xi| \leq a$, while $\widehat{\Psi}(\xi)$ is supported in $b \leq |\xi| \leq c$. Consider the bilinear operator

$$B(f_1, f_2) = \sum_j (f_1 * \Psi_{2^{-j}})(f_2 * \Phi_{2^{-j}}).$$

We can then assert: (i) If $a \in L^\infty$, then the mapping $f \mapsto B(f, a)$ is bounded from $L^2(\mathbf{R}^n)$ to itself; and (ii) If $a \in \mathrm{BMO}$, then the mapping $f \mapsto B(a, f)$ is also a bounded operator on L^2.

(b) With Φ, Ψ as above, similar results hold for the continuous analogue B' of B, given by

$$B'(f_1, f_2) = \int_0^\infty (f_1 * \Psi_t)(f_2 * \Phi_t)\frac{dt}{t}.$$

Notice that B' is a bilinear variant of the square function s_Φ discussed in Chapter 1, §6.3.1.

(c) One can drop the compatibility condition on the spectra of Φ and Ψ by more elaborate arguments (see also §5.21 below); one way to do this is to define a "corrected" version B_c of B' as follows. Assume that $\int \Psi\, dx = 0$ and let $\psi \in \mathcal{S}$ with $\int \psi\, dx = 0$. Set

$$B_c(f_1, f_2) = \int_0^\infty [(f_1 * \Psi_t)(f_2 * \Phi_t)] * \psi_t \frac{dt}{t}.$$

Then the mappings $f \mapsto B_c(f, a)$, $f \mapsto B_c(a, f)$ are bounded operators on $L^2(\mathbf{R}^n)$ whenever $a \in L^\infty$, or $a \in \mathrm{BMO}$, respectively.

The proof of (a) is based in part on the observation that the spectrum of $(f_1 * \Psi_{2^{-j}})(f_2 * \Phi_{2^{-j}})$ is contained in $(b-a)2^j \leq |\xi| \leq (c+a)2^j$; hence, there are a bounded number of overlaps as j varies.

To prove (c), one notes that

$$|\langle B_c(f_1, f_2), g\rangle| \leq \|s_{\psi^*}(g)\|_{L^2}\left(\int_{\mathbf{R}^{n+1}_+} |f_1 * \Psi_t|^2\, |f_2 * \Phi_t|^2 \frac{dx\, dx}{t}\right)^{1/2},$$

where $\psi^*(x) = \overline{\psi(-x)}$ and $(s_{\psi^*} g)(x) = (\int_0^\infty |(g * \psi_t^*)(x)|^2\, dt/t)^{1/2}$, and applies Chapter 1, §6.3.1 and Chapter 4, §4.3.

Further details, and variants of the above, may be found in Coifman and Y. Meyer [1978].

5.21 It turns out that it is useful to re-examine certain parts of the analysis in §3, as well as that in §5.20, from the point of view of multilinear transformations, and in particular those multilinear operators that generalize convolutions (i.e., "multiplier" operators).

For k fixed, we consider the k-linear operator

$$(f_1, \ldots, f_k) \mapsto M(f_1, \ldots, f_k),$$

initially defined for $f_k \in \mathcal{S}$, as follows. We write

$$\bar{\xi} = (\xi^{(1)}, \ldots, \xi^{(k)}) \in \mathbf{R}^{nk}, \qquad \xi^{(j)} \in \mathbf{R}^n,$$

and set $\sigma(\bar{\xi}) = \xi^{(1)} + \cdots + \xi^{(k)}$. If $\bar{\xi} \mapsto m(\bar{\xi})$ is a bounded function on $\mathbf{R}^{nk}$, we define M by

$$M(f_1, \ldots, f_k)(x) = \int_{\mathbf{R}^{nk}} e^{2\pi i x\cdot\sigma(\bar{\xi})}\, m(\bar{\xi})\, \hat{f}_1(\xi^{(1)}) \cdot \cdots \cdot \hat{f}_k(\xi^{(k)})\, d\bar{\xi}.$$

Here are some examples of M.

(a) $M(f_1, \ldots, f_k)(x) = f_1(x) \cdot \cdots \cdot f_k(x)$, which corresponds to $m(\bar{\xi}) \equiv 1$.

(b) $M(f_1, f_2) = B'(f_1, f_2)$, with B' as in §5.20, which corresponds to

$$m(\xi^{(1)}, \xi^{(2)}) = \int_0^\infty \Psi_t(\xi^{(1)})\, \Phi_t(\xi^{(2)})\, \frac{dt}{t}.$$

(c) $M(f_1, f_2) = f_1 H(f_2) - H(f_1 f_2)$, where H is the Hilbert transform, $n = 1$, and

$$m(\xi^{(1)}, \xi^{(2)}) = \frac{1}{i}[\mathrm{sign}(\xi^{(2)}) - \mathrm{sign}(\xi^{(1)} + \xi^{(2)})].$$

(d) The commutator $C_{k-1}(f)$ in §4.2 may also be written in this form.

We state two results concerning such operators.

(1) Suppose $m \in C^\infty(\mathbf{R}^{nk}\backslash 0)$ and $|\partial_{\bar{\xi}}^\alpha m(\bar{\xi})| \leq A_\alpha |\bar{\xi}|^{-|\alpha|}$ for all α. Then

$$\|M(f_1, \ldots, f_k)\|_{L^2(\mathbf{R}^n)} \leq A\left(\prod_{j=1}^{k-1} \|f_j\|_{L^\infty(\mathbf{R}^n)}\right)\|f_k\|_{L^2(\mathbf{R}^n)},$$

i.e., $M : L^\infty(\mathbf{R}^n) \times \cdots \times L^\infty(\mathbf{R}^n) \times L^2(\mathbf{R}^n) \to L^2(\mathbf{R}^n)$ is bounded.

(2) Suppose, in addition, that $m(\bar{\xi}) = 0$ for all $\bar{\xi}$ for which $\xi^{(j)} = 0$ for at least one j, $1 \leq j \leq k-1$. Then $M : \mathrm{BMO}(\mathbf{R}^n) \times \cdots \times \mathrm{BMO}(\mathbf{R}^n) \times L^2(\mathbf{R}^n) \to L^2(\mathbf{R}^n)$ with

$$\|M(f_1, \ldots, f_k)\|_{L^2} \leq A\left(\prod_{j=1}^{k-1} \|f_j\|_{\mathrm{BMO}}\right)\|f_k\|_{L^2}.$$

Note that these results apply to example (b) above (and to the B and B_c considered in §5.20). The result is that, with $\Phi, \Psi \in \mathcal{S}$ and $\int \Psi\, dx = 0$ (not assuming an additional compatibility condition on the spectra of Φ and Ψ), we have that

$$\|B(f,a)\|_{L^2} \leq A\|a\|_{L^\infty}\|f\|_{L^2}, \qquad \text{and} \quad \|B(a,f)\|_{L^2} \leq A\|a\|_{\mathrm{BMO}}\|f\|_{L^2},$$

with similar results for B' and B_c.

The results formulated in (1) and (2) also apply to examples (c) and (d), but in a more indirect fashion.

The theory just described was developed in Coifman and Y. Meyer [1978] and Y. Meyer and Coifman [1991], which should be consulted for further details and a variety of related ideas and applications.

Regarding the proofs of (1) and (2), we limit ourselves to the following brief indication. One proceeds by induction on k. Take, for example, the case $k = 2$. Fix f_1, and consider the operator $f \mapsto T(f) = M(f_1, f)$. One then verifies that T satisfies the assumptions of Theorem 3, in the form stated in §5.19.

Notes

The first form of the almost-orthogonality lemma is in Cotlar [1955]. It treated the case where the T_j are self-adjoint and mutually commutative, and was proved by an intricate combinatorial argument. The general version, found in 1967, is due independently to Cotlar and the author, but its proof did not appear until Knapp and Stein [1971]. That paper also gave the first application of the general lemma, namely to intertwining operators given by singular integrals on nilpotent groups (essentially the case $p = 2$ of Chapter 13, §5.3). The next application was that of Calderón and Vaillancourt [1972] who proved Theorem 2, but by a different arrangement of the argument. Some further types of almost-orthogonality devices (and their application) can be found in Coifman and Y. Meyer [1978]. The notion of paraproducts arising in §1.2.2 is exploited by Bony [1981].

The result in §1.3 was proved in Stein [1973a].

Theorem 3 is a reformulation of the "$T(1)$" theorem of David and Journé [1984]. Our version has the advantage over the original of not invoking BMO in its hypothesis, while still having roughly the same scope. The development leading to the $T(1)$ theorem was greatly influenced by the study of the Cauchy integral and the commutators of Calderón; for these, see the discussion and references in the appendix above.

The idea of Proposition 1 (in §3.1), highlighting the necessity of cancellation conditions, goes back to Knapp and Stein [1971]. The general formulation of such conditions (Theorem 4) is a consequence of Theorem 3, but has not appeared before.

CHAPTER VIII

Oscillatory Integrals of the First Kind

Oscillatory integrals in one form or another have been an essential part of harmonic analysis from the very beginnings of that subject. Besides the obvious fact that the Fourier transform is itself an oscillatory integral *par excellence*, one need only recall the occurrence of Bessel functions in the work of Fourier, the study of asymptotics related to these functions by Airy, Stokes, and Lipschitz, and Riemann's use of the method of "stationary phase" in finding the asymptotics of certain Fourier transforms, all of which took place well over 100 years ago. Another early impetus for the study of oscillatory integrals, initiated at the beginning of this century, came with their application to number theory, in particular to the distribution of lattice points and to their relation with exponential sums. In more recent times the subject has been transformed, with the principal emphasis now on operators fashioned from oscillatory integrals, of the kind arising in a variety of problems.

The purpose of this and the succeeding chapter is to present some of the main ideas of this theory. We will find it convenient to divide our discussion of oscillatory integrals by making a distinction between those of the *first kind* and those of the *second kind*. The main difference between the two is that for the former we will be dealing with a single function, which can typically be written in the form

$$I(\lambda) = \int e^{i\lambda\phi(x)}\psi(x)\, dx. \tag{1}$$

The problem then is the asymptotic behavior of $I(\lambda)$ as the parameter λ tends to infinity.

In the latter case we will be concerned with the boundedness properties of an operator that carries an oscillatory factor in its kernel and which can be given in the form

$$T_\lambda(f)(x) = \int e^{i\lambda\Phi(x,y)}K(x,y)\, f(y)\, dy. \tag{2}$$

Here, finding estimates for the norm of the operator T_λ as $\lambda \to \infty$ is our principal goal. Sometimes the parameter λ is incorporated into the phase Φ, and the objective is then to study the boundedness properties of the operator (2).

This chapter is devoted to oscillatory integrals of the first kind. We begin with the case of one dimension, for which there is an essentially complete theory. There are three simple principles that govern the behavior of $I(\lambda)$, as $\lambda \to \infty$, in that case. These are: the principal contributions to $I(\lambda)$ come from the critical points of ϕ; in addition, assuming the existence of a single critical point, there is a prescription for the complete asymptotics of $I(\lambda)$, which is qualitatively determined by the exact order of vanishing of ϕ' at this critical point; also, there is a universal estimate for the decay of $I(\lambda)$, consistent with these asymptotics, and given in terms of a lower bound for some derivative of ϕ.

The corresponding situation in $\mathbf{R}^n$ is far from being as clear-cut. This is connected with the multiplicity and complexity of critical points that can occur when the dimension exceeds 1. Nevertheless, one can establish the complete asymptotic development of $I(\lambda)$ if the critical point of ϕ is "nondegenerate", as well as the fact that $I(\lambda)$ always has a decay of order $O(\lambda^{-\varepsilon})$, for some $\varepsilon > 0$, as long as some partial derivative of ϕ is nonvanishing over the range of integration.

Our main use of these oscillatory integrals is to obtain decay estimates for Fourier transforms of measures carried on surfaces. Here the conditions required on the phase ϕ come about because of "curvature" conditions on the surfaces in question. In this way, oscillatory integrals provide the link between geometric properties of manifolds and harmonic analysis related to them.

We shall work out two important applications of this latter idea. One is to maximal averages associated with curved surfaces, which is the subject of Chapter 11. The other is to restriction theorems for the Fourier transform, a subject we only touch on here, but which will be taken up in greater depth in the next chapter.

1. Oscillatory integrals of the first kind, one variable

We are interested in the behavior for large positive λ of the integral

$$I(\lambda) = \int_a^b e^{i\lambda\phi(x)}\psi(x)\,dx, \tag{3}$$

where ϕ is a real-valued smooth function (the *phase*), and ψ is complex-valued and smooth; often, but not always, one assumes that ψ has compact support in (a, b).

The basic facts about $I(\lambda)$ can be presented in terms of three principles: *localization*, *scaling*, and *asymptotics*.

1.1 Localization. This is the observation that, assuming ϕ has compact support in (a, b), the asymptotic behavior of $I(\lambda)$ is determined by those points where $\phi'(x) = 0$. The proof of this fact is a consequence of the following:

PROPOSITION 1. *Let ϕ and ψ be smooth functions so that ψ has compact support in (a, b), and $\phi'(x) \neq 0$ for all $x \in [a, b]$. Then*

$$I(\lambda) = O(\lambda^{-N}) \quad \text{as } \lambda \to \infty$$

for all $N \geq 0$.

The proof is very simple. Let D denote the differential operator

$$Df(x) = (i\lambda\phi'(x))^{-1} \cdot \frac{df}{dx}$$

and let ${}^tD$ denote its transpose

$${}^tDf(x) = \frac{-d}{dx}\left(\frac{f}{i\lambda\phi'(x)}\right).$$

Then $D^N(e^{i\lambda\phi}) = e^{i\lambda\phi}$ for every N, and integration by parts shows that

$$\int_a^b e^{i\lambda\phi}\psi\,dx = \int_a^b D^N(e^{i\lambda\phi})\,\psi\,dx = \int_a^b e^{i\lambda\phi}\cdot({}^tD)^N(\psi)\,dx. \tag{4}$$

Thus clearly $|I(\lambda)| \leq A_N\lambda^{-N}$, and the proposition is proved.

We remark that, under the change of variable $x \mapsto \phi(x)$, our proposition becomes the familiar statement that a compactly supported smooth function has a rapidly decreasing Fourier transform.

Note that if we do not assume that ψ vanishes near the endpoints of the interval $[a, b]$, then the formula (4) must be modified by adding the extra integrated terms that appear. In that case the best conclusion that can be obtained for $I(\lambda)$ is the decrease $O(\lambda^{-1})$, as the example

$$\int_a^b e^{i\lambda x}\,dx = \frac{e^{i\lambda b} - e^{i\lambda a}}{i\lambda}$$

shows. Observe, however, that the integrated terms that would occur in (4) actually cancel, giving again the rapid decrease of $I(\lambda)$, in the "periodic" case, i.e., if we have

$$\phi^{(k)}(a) = \phi^{(k)}(b) \qquad \text{and} \qquad \psi^{(k)}(a) = \psi^{(k)}(b)$$

for all $k \geq 0$. In any case, the localization principle can be broadened to state that the asymptotic behavior of $I(\lambda)$ is determined by those points in (a, b) where $\phi'(x) = 0$, together with the contribution from the endpoints of the interval.

1.2 Scaling. Suppose we know only that

$$\left|\frac{d^k\phi(x)}{dx^k}\right| \geq 1$$

for some fixed k, and we wish to obtain an estimate for

$$\int_a^b e^{i\lambda\phi(x)}\,dx$$

that is *independent* of a and b. The change of variable $x \mapsto \lambda^{-1/k}x'$ shows that the only possible estimate for the integral is

$$O(\lambda^{-1/k}).$$

That this estimate indeed holds goes back to van der Corput, and is contained in the following:

PROPOSITION 2. *Suppose ϕ is real-valued and smooth in (a,b), and that $|\phi^{(k)}(x)| \geq 1$ for all $x \in (a,b)$. Then*

$$\left|\int_a^b e^{i\lambda\phi(x)}\,dx\right| \leq c_k\lambda^{-1/k} \tag{5}$$

holds when:

(i) *$k \geq 2$, or*

(ii) *$k = 1$ and $\phi'(x)$ is monotonic.*

The bound c_k is independent of ϕ and λ.

Note that, when $k = 1$, a simple lower bound on $|\phi'|$ is not sufficient. For example, taking $\phi' > 1$ and $\lambda = 1$, suppose that $\phi'(x)$ oscillates so that it is large when $\cos(\phi(x)) < 0$ and is (comparatively) small when $\cos(\phi(x)) > 0$. Thus the measure of the set where $\cos(\phi(x))$ is positive is much larger than the measure of the set where it is negative, and the real part of the above integral is unbounded (say as $b \to \infty$).

Proof. Let us show (ii) first. Taking D as in the proof of Proposition 1, we have

$$\int_a^b e^{i\lambda\phi}\,dx = \int_a^b D(e^{i\lambda\phi})\,dx = \int_a^b e^{i\lambda\phi}\cdot {}^tD(1)\,dx + (i\lambda\phi')^{-1}e^{i\lambda\phi}\Big|_a^b.$$

The boundary terms are majorized by $2/\lambda$, while

$$\begin{aligned}\left|\int_a^b e^{i\lambda\phi}\cdot {}^tD(1)\,dx\right| &= \left|\int_a^b e^{i\lambda\phi}(i\lambda)^{-1}\frac{d}{dx}\left(\frac{1}{\phi'}\right)dx\right| \\ &\leq \lambda^{-1}\int_a^b \left|\frac{d}{dx}\left(\frac{1}{\phi'}\right)\right|dx = \lambda^{-1}\left|\int_a^b \frac{d}{dx}\left(\frac{1}{\phi'}\right)dx\right|,\end{aligned}$$

by the monotonicity of ϕ'. The last expression equals

$$\lambda^{-1}\left|\frac{1}{\phi'(b)} - \frac{1}{\phi'(a)}\right|,$$

which is dominated by $1/\lambda$. This gives the desired conclusion with $c_1 = 3$.

We now prove (i) by induction on k. Let us suppose that the case k is known, and assume (replacing ϕ by $-\phi$ if necessary) that

$$\phi^{(k+1)}(x) \geq 1 \qquad \text{for all } x \in [a,b].$$

Let $x = c$ be the (unique) point in $[a,b]$ where $|\phi^{(k)}(x)|$ assumes its minimum value. If $\phi^{(k)}(c) = 0$ then, outside the interval $(c-\delta, c+\delta)$, we have that $|\phi^{(k)}(x)| \geq \delta$ (and, of course, $\phi'(x)$ is monotonic in the case $k=1$). Write

$$\int_a^b = \int_a^{c-\delta} + \int_{c-\delta}^{c+\delta} + \int_{c+\delta}^b.$$

By our inductive hypothesis,

$$\left|\int_a^{c-\delta} e^{i\lambda\phi}\,dx\right| \leq c_k(\lambda\delta)^{-1/k}.$$

Similarly,

$$\left|\int_{c+\delta}^b e^{i\lambda\phi}\,dx\right| \leq c_k(\lambda\delta)^{-1/k}.$$

Since $|\int_{c-\delta}^{c+\delta} e^{i\lambda\phi}\,dx| \leq 2\delta$, we have

$$\left|\int_a^b e^{i\lambda\phi}\,dx\right| \leq \frac{2c_k}{(\lambda\delta)^{1/k}} + 2\delta.$$

If $\phi^{(k)}(c) \neq 0$, and so c is one of the endpoints of $[a,b]$, a similar argument shows that $c_k(\lambda\delta)^{-1/k} + \delta$ is an upper bound for the integral. In either situation, the case $k+1$ follows by taking

$$\delta = \lambda^{-1/(k+1)},$$

which proves (5) with $c_{k+1} = 2c_k + 2$; since $c_1 = 3$, we have $c_k = 5\cdot 2^{k-1} - 2$.

This proposition leads to a similar estimate for integrals of the form (3). Here we do not assume that ψ vanishes near the endpoints of $[a,b]$.

COROLLARY. *Under the assumptions on ϕ in Proposition 2, we can conclude that*

$$\left|\int_a^b e^{i\lambda\phi(x)}\psi(x)\,dx\right| \le c_k\lambda^{-1/k}\left[|\psi(b)| + \int_a^b |\psi'(x)|\,dx\right]. \tag{6}$$

This is proved by writing (3) as $\int_a^b F'(x)\,\psi(x)\,dx$, with

$$F(x) = \int_a^x e^{i\lambda\phi(t)}\,dt,$$

integrating by parts, and using the estimate

$$|F(x)| \le c_k\lambda^{-1/k}, \qquad \text{for } x \in [a,b],$$

obtained previously.

1.3 Asymptotics. The third principle describes the full asymptotic development of $I(\lambda)$. We already know that, when the support ψ is a compact subset of (a,b), the behavior of

$$\int_a^b e^{i\lambda\phi}\psi\,dx$$

is determined by those points x_0 where $\phi'(x_0) = 0$ (the "critical" points of ϕ). Assuming that the support of ψ is so small that it contains only one critical point of ϕ, the character of the asymptotic expansion depends on the smallest $k \ge 2$ for which

$$\phi^{(k)}(x_0) \ne 0$$

and is given in terms of powers of λ, in a way that is consistent with Proposition 2. The following proposition encompasses this idea and is one of the expressions of the "method of stationary phase".

PROPOSITION 3. *Suppose $k \ge 2$, and*

$$\phi(x_0) = \phi'(x_0) = \cdots = \phi^{(k-1)}(x_0) = 0,$$

while $\phi^{(k)}(x_0) \ne 0$. If ψ is supported in a sufficiently small neighborhood of x_0, then

$$I(\lambda) = \int e^{i\lambda\phi(x)}\psi(x)\,dx \sim \lambda^{-1/k}\sum_{j=0}^{\infty} a_j\lambda^{-j/k}, \tag{7}$$

in the sense that, for all nonnegative integers N and r,

$$\left(\frac{d}{d\lambda}\right)^r\left[I(\lambda) - \lambda^{-1/k}\sum_{j=0}^{N} a_j\lambda^{-j/k}\right] = O(\lambda^{-r-(N+1)/k}) \quad \text{as } \lambda \to \infty. \tag{8}$$

1.3.1 We shall give the proof first for the case $k = 2$. There are three steps.

Step 1. This is the observation that

$$\int_{-\infty}^{\infty} e^{i\lambda x^2}x^\ell e^{-x^2}\,dx \sim \lambda^{-(\ell+1)/2}\sum_{j=0}^{\infty} c_j^{(\ell)}\lambda^{-j}, \tag{9}$$

for any nonnegative integer ℓ; when ℓ is odd, the integral vanishes. In fact, the left side of (9) is

$$\int_{-\infty}^{\infty} e^{-(1-i\lambda)x^2}x^\ell\,dx;$$

setting $z = (1-i\lambda)^{1/2}\cdot x$ and noting that the rapid decay of e^{-z^2} allows us to replace the contour $(1-i\lambda)^{1/2}\cdot\mathbf{R}$ by $\mathbf{R}$, we see that the above integral equals

$$(1-i\lambda)^{-1/2-\ell/2}\int_{-\infty}^{\infty} e^{-x^2}\,x^\ell\,dx.$$

Here we have fixed the principal branch of $z^{-(\ell+1)/2}$ in the complex plane slit along the negative half-axis. With this determination we have

$$(1-i\lambda)^{-(\ell+1)/2} = \lambda^{-(\ell+1)/2}\cdot(\lambda^{-1}-i)^{-(\ell+1)/2},$$

if $\lambda > 0$. Thus the power series expansion of $(w-i)^{-(\ell+1)/2}$ (on the disc $|w| < 1$) gives us the asymptotic expansion (9) with $w = \lambda^{-1} \to 0$.

Step 2. Observe next that, if $\eta \in C_0^\infty$ and ℓ is a nonnegative integer, then

$$\left|\int_{-\infty}^{\infty} e^{i\lambda x^2}x^\ell\,\eta(x)\,dx\right| \le A\lambda^{-1/2-\ell/2}. \tag{10}$$

To prove this, let α be a C^∞ function with the property that $\alpha(x) = 1$ for $|x| \le 1$, and $\alpha(x) = 0$ for $|x| \ge 2$, and write

$$\begin{aligned}\int e^{i\lambda x^2}x^\ell\,\eta(x)\,dx &= \int e^{i\lambda x^2}x^\ell\,\eta(x)\,\alpha(x/\varepsilon)\,dx \\ &\quad + \int e^{i\lambda x^2}x^\ell\,\eta(x)\,[1-\alpha(x/\varepsilon)]\,dx,\end{aligned}$$

where $\varepsilon > 0$ will be set momentarily.

The first integral is dominated by $C\varepsilon^{\ell+1}$. The second integral can be written as

$$\int e^{i\lambda x^2}\,({}^tD)^N\{x^\ell\eta(x)[1-\alpha(x/\varepsilon)]\}\,dx,$$

with ${}^tDf = -(i\lambda)^{-1}\cdot(d/dx)[f/2x]$. A simple computation then shows that this term is majorized by

$$\frac{C_N}{\lambda^N}\int_{|x|\ge\varepsilon} |x|^{\ell-2N}\,dx = C_N'\lambda^{-N}\varepsilon^{\ell-2N+1}$$

if $\ell - 2N < -1$. Altogether then the integral in (10) is bounded by

$$C_N[\varepsilon^{\ell+1} + \lambda^{-N}\varepsilon^{\ell-2N+1}],$$

and we need only take $\varepsilon = \lambda^{-1/2}$ (with $N > (\ell+1)/2$) to get the desired conclusion.

A similar (but simpler) argument of integration by parts also shows that

$$\int e^{i\lambda x^2} g(x)\, dx = O(\lambda^{-N}), \qquad \text{every } N \geq 0, \tag{11}$$

whenever $g \in \mathcal{S}$ and g vanishes near the origin.

Step 3. We now prove the proposition in the case $\phi(x) = x^2$. To do this, write

$$\int e^{i\lambda x^2} \psi(x)\, dx = \int e^{i\lambda x^2} e^{-x^2} [e^{x^2}\psi(x)]\tilde{\psi}(x)\, dx,$$

where $\tilde{\psi} \in C_0^\infty$ is 1 on the support of ψ. For each N, write the Taylor expansion

$$e^{x^2}\psi(x) = \sum_{j=0}^{N} b_j x^j + x^{N+1}R_N(x) = P(x) + x^{N+1}R_N(x).$$

Substituting this expansion in the above integral gives three terms:

$$\sum_{j=0}^{N} b_j \int_{-\infty}^{\infty} e^{i\lambda x^2} e^{-x^2} x^j\, dx, \tag{a}$$

$$\int_{-\infty}^{\infty} e^{i\lambda x^2} x^{N+1} R_N(x)\, e^{-x^2}\, \tilde{\psi}(x)\, dx, \tag{b}$$

and

$$\int_{-\infty}^{\infty} e^{i\lambda x^2} P(x)\, e^{-x^2}\, [\tilde{\psi}(x) - 1]\, dx. \tag{c}$$

For (a) we use (9), for (b) we use (10), and for (c) we use (11). It is then easy to see that their combination gives the desired asymptotic expansion for $\int e^{i\lambda x^2}\psi(x)\, dx$.

1.3.2 Let us now consider the general case when $k = 2$. We can then write

$$\phi(x) = c(x - x_0)^2 + O(|x - x_0|^3),$$

with $c \neq 0$ and set

$$\phi(x) = c(x - x_0)^2[1 + \varepsilon(x)],$$

where ε is a smooth function that is $O(|x - x_0|)$, and hence $|\varepsilon(x)| < 1$ when x is sufficiently close to x_0. Moreover, $\phi'(x) \neq 0$ when $x \neq x_0$ is sufficiently close to x_0. Let us fix a neighborhood U of x_0 so that both these conditions hold on U, and let

$$y = (x - x_0)[1 + \varepsilon(x)]^{1/2}.$$

Then the mapping $x \mapsto y$ is a diffeomorphism from U to a neighborhood of $y = 0$, and of course $cy^2 = \phi(x)$. Thus

$$\int e^{i\lambda\phi(x)}\psi(x)\, dx = \int e^{i\lambda c y^2}\tilde{\psi}(y)\, dy$$

with $\tilde{\psi} \in C_0^\infty$ if the support of ψ lies in U. The expansion (7) (for $k = 2$) is then proved as a consequence of the special case treated above.

1.3.3 The proof for higher k is similar and is based on the identity

$$\int_0^\infty e^{i\lambda x^k} e^{-x^k} x^\ell\, dx = c_{k,\ell}(1 - i\lambda)^{-(\ell+1)/k}.$$

We omit the details.

1.3.4 Remarks. 1. Each constant a_j that appears in the asymptotic expansion (7) depends on only finitely many derivatives of ϕ and ψ at x_0. Note that our proof shows that, say in the case $k = 2$, we have

$$a_0 = \left(\frac{2\pi}{-i\phi''(x_0)}\right)^{1/2} \psi(x_0).$$

Similarly, the bounds occurring in the error term of (8) depend on upper bounds of finitely many derivatives of ϕ and ψ in the support of ψ, the size of the support of ψ, and a lower bound for $|\phi^{(k)}(x_0)|$.

A graphic way of rephrasing the main point of (7) in the case $k = 2$ is to incorporate λ in ϕ, and to drop the restriction that $\phi(x_0) = 0$. Then the principal contribution to

$$\int e^{i\phi(x)}\,\psi(x)\, dx$$

is given by

$$e^{i\phi(x_0)}\left(\frac{2\pi}{-i\phi''(x_0)}\right)^{1/2}\psi(x_0). \tag{12}$$

2. Observe that our proof also shows that if k is even, then $a_j = 0$ for all odd j.

1.4 Examples. We describe two classical examples.

1.4.1 Bessel functions. The Bessel function $J_m(r)$ of integral order m is defined by

$$J_m(r) = \frac{1}{2\pi}\int_0^{2\pi} e^{ir\sin\theta}\, e^{-im\theta}\, d\theta. \tag{13}$$

This is of the form (3) with $\lambda = r$, $\phi(x) = \sin x$. Notice that ϕ' vanishes only at $\pi/2$ and $3\pi/2$ in the interval $[0, 2\pi]$, and there $\phi'' = \pm 1$. Now write $1 = \psi_1 + \psi_2 + \psi_3$, where ψ_1 has small support near $\pi/2$ and equals 1 near $\pi/2$, and ψ_2 has small support near $3\pi/2$ and equals 1 near $3\pi/2$. Inserting this partition in (13) allows us to write $J_m(r)$ as a sum of three terms. For the first two, the case $k = 2$ of the corollary (6) in §1.2 is applicable; for the third, we apply the case $k = 1$. The result is clearly

$$J_m(r) = O(r^{-1/2}), \qquad \text{as } r \to \infty.$$

If, instead, we use (the more precise) Proposition 3 in §1.3 for the first two terms, and the remarks following Proposition 1 (in §1.1) for the third term, we can obtain the complete asymptotic development of $J_m(r)$ as $r \to \infty$. We write down the main term,

$$J_m(r) = (2/\pi)^{1/2} \cdot r^{-1/2} \cdot \cos(r - \pi m/2 - \pi/4) + O(r^{-3/2}), \tag{14}$$

and more generally the complete asymptotic expansion,

$$J_m(r) \sim r^{-1/2} e^{ir} \sum_{j=0}^{\infty} a_j r^{-j} + r^{-1/2} e^{-ir} \sum_{j=0}^{\infty} b_j r^{-j}, \tag{15}$$

for suitable coefficients a_j and b_j.

The Bessel function can also be defined for real (nonintegral) values of m; when $m > -1/2$ it is given by

$$J_m(r) = \frac{(r/2)^m}{\Gamma(m+1/2)\pi^{1/2}} \cdot \int_{-1}^{1} e^{irt}(1-t^2)^{m-1/2}\, dt. \tag{16}$$

To check that this second definition agrees with the earlier one (13) (when m is a positive integer), note first that this identity is evident when $m = 0$. Next, a straightforward computation shows that both expressions satisfy the recursion relation

$$\frac{d}{dr}[r^{-m} J_m(r)] = -r^m J_{m+1}(r).$$

Besides the case when m is an integer, the situation when m is half-integral arises often in applications. For such m it is obvious from (16) that, after integrating by parts $m - 1/2$ times, we get an elementary function. Again the asymptotic expansion (15) is valid; in fact it is an exact formula, with the terms for $j > m$ vanishing. For more information about Bessel functions, see §5.2.

1.4.2 Riemann singularity. We will be concerned with a distribution supported in the half-line $x \geq 0$ that agrees with the highly oscillatory function

$$e^{i/x} \cdot x^{-\gamma}, \qquad \text{for small } x,$$

and is modified, strictly away from the origin, to have compact support and to be smooth.

So as not to be encumbered by irrelevant complications, we shall limit ourselves to the case $0 \leq \gamma < 2$, which is already typical. In that case, when studying the Fourier transform of this distribution, we are effectively dealing with the integral

$$\int_0^1 e^{ix\xi}\, e^{i/x}\, x^{-\gamma}\, dx, \tag{17}$$

which can be defined as

$$\lim_{\varepsilon\to 0}\int_\varepsilon^1 e^{ix\xi}\, e^{i/x}\, x^{-\gamma}\, dx = -\lim_{\varepsilon\to 0}\int_\varepsilon^1 e^{ix\xi}\, \frac{d}{dx}(e^{i/x})\, ix^2\, x^{-\gamma}\, dx\,;$$

an obvious integration by parts shows that the limit exists.

The principal result in this regard is the asymptotic formula

$$\int_0^1 e^{ix\xi}\, e^{i/x}\, x^{-\gamma}\, dx = \sqrt{\pi i} \cdot e^{2i\xi^{1/2}} \cdot \xi^{-3/4+\gamma/2} + O(\xi^{-1+\gamma/2}), \tag{18}$$

as $\xi \to +\infty$ through positive values.

To prove (18), observe first that if $\phi(x) = x^{-1} + x\xi$, then (17) is

$$\int_0^1 e^{i\phi(x)}\, \psi(x)\, dx, \qquad \text{with } \psi(x) = x^{-\gamma};$$

also $\phi'(x) = -x^{-2} + \xi$ vanishes only when $x = x_0 = \xi^{-1/2}$, while

$$\phi''(x_0) = 2x^{-3} = 2\xi^{3/2}.$$

Thus the main term of (18) is exactly as would have been predicted by (12), had ϕ and ψ been regular near $x = 0$. The actual proof of (18) requires a little care.

First we write the integral (17) as

$$\int_0^a + \int_a^b + \int_b^1 = I_1 + I_2 + I_3.$$

Here (a, b) is a small interval containing the critical point $x_0 = \xi^{-1/2}$; a reasonable choice is to take

$$a = \frac{1}{2}\cdot \xi^{-1/2} \qquad \text{and} \qquad b = \frac{3}{2}\cdot\xi^{-1/2}.$$

Now it is obvious that for $x \in [b,1]$, $\phi'(x) = -x^{-2} + \xi \geq (5/9)\xi$. Thus by applying the corollary in §1.2 (i.e., (6) with $\psi(x) = x^{-\gamma}$ and $k = 1$), we see that

$$I_3 = O(\xi^{-1} \cdot \xi^{\gamma/2}) = O(\xi^{\gamma/2-1}),$$

which can be subsumed in the error term of (18).

In I_1, we have $0 \leq x \leq a = (1/2)\xi^{-1/2}$. It is easy to see that for these x

$$\phi'(x) = -x^{-2} + \xi \leq (-3/4)x^{-2},$$

while $\phi''(x) = 2x^{-3}$. Thus

$$I_1 = \int_0^a \frac{d}{dx}(e^{i\phi(x)})[i\phi'(x)]^{-1}x^{-\gamma}\,dx = O(a^{2-\gamma}) = O(\xi^{\gamma/2-1}),$$

which again can be incorporated into the error term.

The main contribution comes from I_2. If we were to apply (6) again, this time with $k = 2$, and use the fact that

$$\phi''(x) = 2x^{-3} \geq c \cdot \xi^{3/2} \qquad \text{for } x \in (a,b),$$

then we would get $I_2 = O(\xi^{\gamma/2-3/4})$ and the result that the integral (17) is

$$O(\xi^{-3/4+\gamma/2}).$$

To get the more refined conclusion (18), it is necessary to make more precise calculations near the critical point x_0. This is best done by recasting the integral I_2 in a "normalized" form via a change of variables, so as to be able to utilize the asymptotics in §1.3.

This change of variables is given by

$$y = \xi^{1/2}(x - x_0), \qquad \text{with } x_0 = \xi^{-1/2},$$

which maps

$$[a,b] = \{x : |x - x_0| \leq (1/2)\xi^{-1/2}\} \to [-1/2, 1/2]$$

linearly. With λ to be determined, we set

$$\phi(x) - \phi(x_0) = \lambda\Phi(y), \quad \text{i.e.,} \quad \Phi(y) = \lambda^{-1}[\phi(y\xi^{-1/2} + x_0) - \phi(x_0)].$$

We now choose λ so that $\Phi''(0) = 1$, which gives

$$1 = \lambda^{-1}\xi^{-1}\phi''(x_0) = \lambda^{-1}\xi^{-1} \cdot 2\xi^{3/2},$$

implying that $\lambda = 2\xi^{1/2}$. It is easy to check that

$$|\Phi^{(k)}(y)| \leq c_k, \quad k = 0, \ldots, \qquad \text{for } y \in [-1/2, 1/2].$$

Also, $\Phi''(y) \geq c > 0$ for $y \in [-1/2, 1/2]$. Moreover,

$$\Phi(0) = \Phi''(0) = 0,$$

while $\Phi'(1/2) = 5/18$ and $\Phi'(-1/2) = -3/2$.

Next set $\Psi(y) = \xi^{-\gamma/2} \cdot x^{-\gamma}$. Then $\Psi(0) = 1$, while

$$|\Psi^{(k)}(y)| \leq c_k, \quad k = 0, \ldots, \qquad \text{for } y \in [-1/2, 1/2].$$

Also note that $dx = \xi^{-1/2}\,dy$. Therefore,

$$\int_a^b e^{i\phi(x)}\,x^{-\gamma}\,dx = e^{i\phi(x_0)} \cdot \xi^{\gamma/2-1/2} \cdot \int_{-1/2}^{1/2} e^{i\lambda\Phi(y)}\,\Psi(y)\,dy, \tag{19}$$

with $\lambda = 2\xi^{1/2}$, and $\phi(x_0) = x_0^{-1} + x_0 \cdot \xi = 2\xi^{1/2}$.

We are now in a position to use Proposition 3, and the remarks concerning it in §1.3.4 above. For this purpose, write

$$\Psi(y) = \Psi_0(y) + \Psi_1(y),$$

where Ψ_1 is smooth and vanishes near the endpoints of $[-1/2, 1/2]$, and Ψ_0 is supported in small neighborhoods of these endpoints. Then

$$\int_{-1/2}^{1/2} e^{i\lambda\Phi(y)}\,\Psi_1(y)\,dy = \sqrt{2\pi i} \cdot \lambda^{-1/2} + O(\lambda^{-1}),$$

while the proposition in §1.2 (for $k = 1$) shows that

$$\int_{-1/2}^{1/2} e^{i\lambda\Phi(y)}\,\Psi_0(y)\,dy = O(\lambda^{-1}).$$

Combining this with (19) gives the desired asymptotics for I_2 and finally proves (18).

Remark. The estimates for I_1 and I_3 above show that the integral in (18) is $O(|\xi|^{-1+\gamma/2})$ when $\xi \to -\infty$.

2. *Oscillatory integrals of the first kind, several variables*

Only some of the above results have analogues when $n \geq 2$, but the extension of Proposition 1 is simple. Continuing the terminology used above, we say that a phase function ϕ defined in a neighborhood of a point x_0 has x_0 as a *critical point* if

$$(\nabla\phi)(x_0) = \left(\frac{\partial\phi}{\partial x_1}, \ldots, \frac{\partial\phi}{\partial x_n}\right)\bigg|_{x=x_0} = 0.$$

2.1 PROPOSITION 4. *Suppose ψ is smooth, has compact support, and that ϕ is a smooth real-valued function that has no critical points in the support of ψ. Then*

$$I(\lambda) = \int_{\mathbf{R}^n} e^{i\lambda\phi(x)}\,\psi(x)\,dx = O(\lambda^{-N}),$$

as $\lambda \to \infty$, for every $N \geq 0$.

Proof. For each x_0 in the support of ψ, there is a unit vector ξ and a small ball $B(x_0)$, centered at x_0, so that

$$\xi \cdot (\nabla\phi)(x) \geq c > 0$$

for all $x \in B(x_0)$. Decompose the integral $I(\lambda)$ as a finite sum

$$\sum_k \int e^{i\lambda\phi(x)}\, \psi_k(x)\, dx,$$

where each ψ_k is smooth and has compact support in one of these balls. It then suffices to prove the corresponding estimate for each of these integrals. Now choose a coordinate system $x_1, \dots, x_n$ so that x_1 lies along ξ. Then

$$\int e^{i\lambda\phi(x)}\, \psi_k(x)\, dx = \int \left(\int e^{i\lambda\phi(x_1,\dots,x_n)}\, \psi_k(x_1,\dots,x_n)\, dx_1 \right) dx_2 \cdots dx_n.$$

By Proposition 1, the inner integral is rapidly decreasing, and our conclusion follows.

2.2 In several variables, we can only assert a weak analogue of Proposition 2 (the scaling principle). However, it will prove useful in what follows.

PROPOSITION 5. *Suppose ψ is smooth and is supported in the unit ball; also let ϕ be a real-valued function so that, for some multi-index α with $|\alpha| > 0$, we have*

$$|\partial_x^\alpha \phi| \geq 1$$

throughout the support of ψ. Then

$$\left| \int_{\mathbf{R}^n} e^{i\lambda\phi(x)}\, \psi(x)\, dx \right| \leq c_k(\phi) \cdot \lambda^{-1/k} \cdot (\|\psi\|_{L^\infty} + \|\nabla\psi\|_{L^1}) \tag{20}$$

where $k = |\alpha|$; the constant $c_k(\phi)$ is independent of λ and ψ, and remains bounded as long as the C^{k+1} norm of ϕ remains bounded.

Proof. Consider the real linear space of homogeneous polynomials of degree k in $\mathbf{R}^n$; let $d(k,n)$ denote its dimension. Of course, $\{x^\alpha : |\alpha| = k\}$ is a basis for this space. It is not difficult to see that there are unit vectors

$$\xi^1, \dots, \xi^{d(k,n)}$$

so that the homogeneous polynomials

$$(\xi^j \cdot x)^k, \qquad j = 1, \dots, d(k,n)$$

also give a basis.

Momentarily taking this fact for granted (we prove it in §2.2.1 below), we see that if

$$|\partial_x^\alpha \phi(x_0)| \geq 1 \qquad \text{for some } \alpha \text{ with } |\alpha| = k,$$

then there is a unit vector $\xi = \xi(x_0)$ so that

$$|(\xi \cdot \nabla)^k \phi(x_0)| \geq a_k > 0.$$

Moreover, since we can assume that the C^{k+1} norm of ϕ is bounded, we can also conclude that

$$|(\xi \cdot \nabla)^k \phi(x)| \geq a_k/2, \qquad x \in B(x_0).$$

Here $B(x_0)$ is a ball about x_0 of fixed radius $c \cdot \|\phi\|_{C^{k+1}}^{-1}$.

Next choose an appropriate covering of $\mathbf{R}^n$ by such balls of fixed radius and a corresponding partition of unity

$$1 = \sum \eta_j(x), \qquad 0 \leq \eta_j \leq 1, \qquad \sum |\nabla \eta_j| \leq b_k,$$

with each η_j supported in one of our balls. Write $\psi_j = \psi \cdot \eta_j$ and

$$\int e^{i\lambda\phi}\psi\, dx = \sum \int e^{i\lambda\phi}\psi_j\, dx.$$

To estimate $\int e^{i\lambda\phi}\psi_j\, dx$, with ξ determined as above, choose a coordinate system so that x_1 lies along ξ. Then

$$\int e^{i\lambda\phi}\psi_j\, dx = \int \left(\int e^{i\lambda\phi(x_1,\dots,x_n)}\psi_j(x_1,\dots,x_n)\, dx_1 \right) dx_2 \cdots dx_n.$$

For the inner integral, we invoke (6), giving us an estimate of the form

$$c_k(a_k\lambda)^{-1/k} \left\{ \|\psi\|_{L^\infty} + \int \left| \frac{\partial\psi}{\partial x_1}(x_1,\dots,x_n) \right| dx_1 \right\}.$$

A final integration in the other variables then gives (20).

2.2.1 We prove that the polynomials of the form

$$(\xi \cdot x)^k, \qquad \xi \in \mathbf{R}^n,$$

span the linear space of real homogeneous polynomials of degree k. On this linear space, consider the positive definite inner product

$$\langle P, Q \rangle = \sum_{|\alpha|=k} \alpha!\, a_\alpha\, b_\alpha,$$

where $P(x) = \sum a_\alpha x^\alpha$ and $Q(x) = \sum b_\alpha x^\alpha$. Note that

$$\langle P, Q \rangle = [Q(\partial/\partial x)](P).$$

So if P were orthogonal to all the $(\xi \cdot x)^k$, then

$$(\xi \cdot \nabla)^k(P) = 0, \qquad \text{for all } \xi \in \mathbf{R}^n;$$

in other words,

$$\left(\frac{\partial}{\partial t}\right)^k P(t\xi) = 0 \qquad \text{for all } \xi \in \mathbf{R}^n,$$

which can happen only if P is identically zero, proving our assertion.

2.2.2 Remarks. 1. In a variety of cases, the estimates given by (20) are not optimal. Besides the obvious improvement for $k = 1$ given by Proposition 4, note, for instance, that in $\mathbf{R}^2$, if $\phi(x) = x_1 \cdot x_2$, then (20) gives no better than a decrease of order $\lambda^{-1/2}$, while the proposition below shows that the true order is λ^{-1}. See also §5.6 below.

2. Let us return for the moment to the case of one dimension. If ϕ has a critical point at x_0, and ϕ' does not vanish to infinite order at x_0 then, after a smooth change of variables, ϕ can be transformed to a simple canonical form $\tilde{\phi}$, with $\tilde{\phi}(x) = \pm x^k$ for x near 0. There is no analogue of this in higher dimensions, except in a special case corresponding to $k = 2$. We now turn to the asymptotics of the latter situation.

2.3 The case of nondegenerate critical points. Suppose ϕ has a critical point at x_0. If the symmetric $n \times n$ matrix

$$\left[\frac{\partial^2 \phi}{\partial x_i \partial x_j}\right](x_0) \tag{$*$}$$

is invertible, then the critical point is said to be *nondegenerate.* Using a Taylor expansion, it is an easy matter to see that if x_0 is a nondegenerate critical point, then in fact it is an isolated critical point.

PROPOSITION 6. *Suppose $\phi(x_0) = 0$, and ϕ has a nondegenerate critical point at x_0. If ψ is supported in a sufficiently small neighborhood of x_0, then*

$$\int_{\mathbf{R}^n} e^{i\lambda\phi(x)}\, \psi(x)\, dx \sim \lambda^{-n/2} \sum_{j=0}^{\infty} a_j \lambda^{-j}, \quad \text{as } \lambda \to \infty, \tag{21}$$

where the asymptotics hold in the same sense as in (7) *and* (8).

Note. Again, each of the constants a_j appearing in the asymptotic expansion depends on the values of only finitely many derivatives of ϕ and ψ at x_0. Thus, for instance,

$$a_0 = \psi(x_0) \cdot (2\pi)^{n/2} \prod_{j=1}^{n} (-i\mu_j)^{-1/2},$$

where $\mu_1, \dots, \mu_n$ are the eigenvalues of the matrix $(*)$. Similarly, each of the bounds occurring in the error terms depends only on upper bounds for finitely many derivatives of ϕ and ψ in the support of ψ.

2.3.1 The proof of the proposition follows closely the same pattern as that of Proposition 3. First, let $Q(x)$ denote the unit quadratic form given by

$$Q(x) = \sum_{1}^{m} x_j^2 - \sum_{m+1}^{n} x_j^2,$$

where $0 \le m \le n$, and m is fixed. The analogue of (9) is

$$\int_{\mathbf{R}^n} e^{i\lambda Q(x)} e^{-|x|^2} x^\ell \, dx \sim \lambda^{-n/2-|\ell|/2} \sum_{j=0}^{\infty} c_j(m,\ell) \lambda^{-j}, \tag{22}$$

with $\ell = (\ell_1, \dots, \ell_n)$ a multi-index, $|\ell| = \ell_1 + \cdots + \ell_n$, and $x^\ell = x_1^{\ell_1} \cdots x_n^{\ell_n}$; also note that if one ℓ_j is odd, then (22) is identically zero. To prove (22), write it as a product

$$\prod_{j=1}^{n} \int_{-\infty}^{\infty} e^{\pm i\lambda x_j^2} e^{-x_j^2} x_j^{\ell_j}\, dx_j = \prod_{j=1}^{n} \int_{-\infty}^{\infty} e^{-x^2} x^{\ell_j}\, dx \cdot (1 \mp i\lambda)^{-1/2-\ell_j/2},$$

remove the factor $\prod_{j=1}^{n} \lambda^{-(\ell_j+1)/2} = \lambda^{-(n+|\ell|)/2}$, and expand the function

$$\prod_{j=1}^{n} (\lambda^{-1} \mp i)^{-1/2-\ell_j/2},$$

for large λ, in a power series in λ^{-1}.

The analogue of (10) is the statement that

$$\left| \int_{\mathbf{R}^n} e^{i\lambda Q(x)} x^\ell \eta(x)\, dx \right| \le A\lambda^{-n/2-|\ell|/2}; \tag{23}$$

if $\eta \in C_0^\infty(\mathbf{R}^n)$. To prove it we consider cones

$$\Gamma_j = \{x : |x_j|^2 \ge |x|^2/2n\},$$

and the smaller

$$\Gamma_j^0 = \{x : |x_j|^2 \ge |x|^2/n\}.$$

Then since

$$\bigcup_{j=1}^{n} \Gamma_j^0 = \mathbf{R}^n,$$

we can find functions $\Omega_1, \dots, \Omega_n$, each homogeneous of degree 0 and smooth away from the origin, so that

$$\sum_{j=1}^{n} \Omega_j(x) = 1 \qquad \text{for all } x \ne 0,$$

and each Ω_j is supported in Γ_j. Then we can write

$$\int_{\mathbf{R}^n} e^{i\lambda Q(x)} x^\ell \eta(x)\, dx = \sum_j \int_{\mathbf{R}^n} e^{i\lambda Q(x)} x^\ell \eta(x) \Omega_j(x)\, dx.$$

In the cone Γ_j, one uses integration by parts via

$$D_j e^{i\lambda Q(x)} = e^{i\lambda Q(x)} \qquad \text{with} \qquad D_j f(x) = (\pm 2i\lambda x_j)^{-1} \cdot \frac{\partial f}{\partial x_j}.$$

This, together with the fact that $|x_j| \geq (2n)^{-1/2}|x|$ in Γ_j, and

$$|({}^tD_j)^N \Omega_j(x)| \leq C_N \cdot \lambda^{-N} \cdot |x|^{-2N},$$

allows one to conclude the proof of (23) in analogy with that of (10).

A similar argument shows that whenever $\eta \in \mathcal{S}$ and η vanishes near the origin, then

$$\int e^{i\lambda Q(x)}\, \eta(x)\, dx = O(\lambda^{-N}), \qquad \text{for every } N \geq 0.$$

We then combine this with (22) and (23) as before to obtain the asymptotic formula (21) in the special case $\phi(x) = Q(x)$.

2.3.2 To pass to the general case, one can appeal to the change of variables guaranteed by Morse's lemma: Since $\phi(x_0) = \nabla\phi(x_0) = 0$, and the critical point x_0 is assumed to be nondegenerate, there exists a diffeomorphism from a small neighborhood of x_0 in the x-space, to a neighborhood of the origin in the y-space, under which ϕ is transformed into

$$\sum_{j=1}^{m} y_j^2 - \sum_{j=m+1}^{n} y_j^2,$$

for some $0 \leq m \leq n$. The index m is the same as that of the form corresponding to

$$\left[\frac{\partial^2\phi}{\partial x_i \partial x_j}\right](x_0).$$

To prove this version of Morse's lemma, we take $x_0 = 0$ and note that, in a sufficiently small neighborhood of x_0, we can write

$$\phi(x) = \sum_{i,j} x_i x_j \phi_{ij}(x),$$

where the ϕ_{ij} are smooth, and $\phi_{ij} = \phi_{ji}$.

Indeed, since $\phi(0) = \nabla\phi(0) = 0$, we have

$$\phi(x) = \int_0^1 \frac{d}{dt}[\phi(tx)]\, dt = \int_0^1 (1-t)\frac{d^2}{dt^2}[\phi(tx)]\, dt;$$

carrying out the indicated differentiation and integration in the second integral gives the asserted form for ϕ. Observe also that

$$\phi_{ij}(0) = \frac{1}{2}\, \frac{\partial^2 \phi}{\partial x_i \partial x_j}(0).$$

Next assume as an inductive hypothesis that we can introduce new variables $y_1, \ldots, y_n$, so that

$$\phi(y) = \pm y_1^2 \pm \cdots \pm y_{r-1}^2 + \sum_{i,j \geq r}^{n} y_i y_j \tilde{\phi}_{ij}(y). \tag{24}$$

After a further possible linear change in the variables $y_r, \ldots, y_n$, we can assume that

$$\phi_{rr}(0) \neq 0,$$

and hence that $\phi_{rr}(y) \neq 0$ for all sufficiently small y.

Now introduce new variables $y'_1, \ldots, y'_n$, with $y'_j = y_j$ for $j \neq r$, and

$$y'_r = [\pm\phi_{rr}(y)]^{1/2}\left[y_r + \sum_{j>r} \frac{y_j \phi_{jr}(y)}{\pm\phi_{rr}(y)}\right];$$

here the sign $\pm$ is the sign of $\phi_{rr}(0)$, so that $\pm\phi_{rr}(y) = |\phi_{rr}(y)|$ for all sufficiently small y.

With these new variables, ϕ is in the form (24), with $r-1$ replaced by r; continuing the induction, we can achieve $r = n$. A permutation of the coordinates in (24) proves our assertion.

This change of variables shows that the relation (21) for general ϕ reduces to that for the special case $\phi = Q$ treated in §2.3.1, and completes the proof of Proposition 6.

3. Fourier transforms of measures supported on surfaces

Let S be an open subset of a smooth m-dimensional submanifold of $\mathbf{R}^n$. We let $d\sigma$ denote the measure on S induced by Lebesgue measure on $\mathbf{R}^n$, and we fix a function $\psi \in C_0^\infty(\mathbf{R}^n)$ whose support intersects S in a compact subset of S.

Consider now the finite Borel measure $d\mu = \psi(x)\, d\sigma$ on $\mathbf{R}^n$, which is of course carried on S. We wish to discuss the behavior at infinity of the Fourier transform of $d\mu$. This problem has a long history, beginning with its appearance in number theory and, in particular, with the distribution of lattice points in regions of $\mathbf{R}^n$ (see §5.12 below).

Our first important observation in this connection concerns the case when S is the unit sphere $\mathbf{S}^{n-1} \subset \mathbf{R}^n$. It is the fact that, when $n > 1$, the Fourier transform

$$\widehat{d\sigma}(\xi) = \int_{\mathbf{S}^{n-1}} e^{-2\pi i x\cdot\xi}\, d\sigma(x)$$

has an unexpected decrease at infinity. More precisely, one has that

$$\widehat{d\sigma}(\xi) = 2\pi|\xi|^{(2-n)/2} J_{(n-2)/2}(2\pi|\xi|), \tag{25}$$

and from the asymptotic formula (15) it follows that

$$|\widehat{d\sigma}(\xi)| = O(|\xi|^{(1-n)/2}). \tag{26}$$

To prove (25), we may assume $\xi = (0, \ldots, 0, \xi_n)$; then

$$\widehat{d\sigma}(\xi) = \int_{\mathbf{S}^{n-1}} e^{-2\pi i|\xi|\cos\theta}\, d\sigma(x),$$

where θ denotes the angle that the unit vector x makes with the "north pole" $(0, \ldots, 0, 1)$. Passing to spherical coordinates on $\mathbf{S}^{n-1}$, we can write

$$d\sigma = d\sigma_{n-1} = (\sin\theta)^{n-2}\, d\sigma_{n-2}\, d\theta.$$

Thus

$$\widehat{d\sigma}(\xi) = |\sigma_{n-2}| \cdot \int_0^\pi e^{-2\pi i|\xi|\cos\theta}(\sin\theta)^{n-2}\, d\theta;$$

here $|\sigma_{n-2}| = \int_{S_{n-2}} d\sigma_{n-2}$ is the area of the $(n-2)$-dimensional sphere. If we set $r = 2\pi|\xi|$, $t = -\cos\theta$, then because of the identity (16), we obtain (25).

We shall see below that decay estimates of the kind (26) are of a much more general nature and are not limited to the fortuitous circumstances connecting rotational symmetry with Bessel functions. They are in fact deducible from "curvature" properties of S, to which we now turn.

3.1 Hypersurfaces of nonzero Gaussian curvature. We suppose first that the dimension of S is $n-1$, and that S has nonzero Gaussian curvature at each point; by this we mean the following: Let x_0 be any point of S, and consider a rotation and translation of the ambient $\mathbf{R}^n$, so that the point x_0 is moved to the origin, and the tangent plane to S at x_0 becomes the hyperplane $x_n = 0$. Then near the origin (i.e., near x_0), the surface S can be given as a graph

$$x_n = \phi(x_1, \ldots, x_{n-1}),$$

with $\phi \in C_0^\infty$, and $\phi(0) = \nabla\phi(0) = 0$. Now consider the $(n-1)\times(n-1)$ matrix

$$\left(\frac{\partial^2\phi}{\partial x_j \partial x_k}\right)(x_0).$$

Its eigenvalues $\nu_1, \ldots, \nu_{n-1}$ are called the *principal curvatures* of S at x_0, and their product (which equals the determinant of the above matrix) is the *Gaussian curvature* of S at x_0; it is sometimes called the *total curvature* of S at x_0.

THEOREM 1. *Suppose S is a smooth hypersurface in $\mathbf{R}^n$, whose Gaussian curvature is nonzero everywhere, and let $d\mu = \psi\, d\sigma$ be as above. Then*

$$|\widehat{d\mu}(\xi)| \leq A|\xi|^{(1-n)/2}. \tag{27}$$

3.1.1 For the purposes of the proof of the theorem it will be convenient (in applying Proposition 6) to change notation by replacing n with $n+1$. By compactness we may assume that S is given as a graph

$$x_{n+1} = \phi(x_1, \ldots, x_n),$$

so $d\sigma = (1+|\nabla\phi|^2)^{1/2}\, dx_1 \cdots dx_n$. Thus we can reduce matters to showing that, if $\tilde\psi$ is supported in a small neighborhood of the origin,

$$\left|\int_{\mathbf{R}^n} e^{i\lambda\Phi(x,\eta)}\, \tilde\psi(x)\, dx\right| \leq A\lambda^{-n/2} \tag{28}$$

where $\lambda = |\xi| > 0$, $\xi = \lambda\eta$; here $\eta = (\eta_1, \ldots, \eta_{n+1})$ is a unit vector, and

$$\Phi(x,\eta) = x \cdot \eta = \sum_1^n x_j\eta_j + \phi(x_1, \ldots, x_n)\eta_{n+1}.$$

Also, we have that $\phi(0) = \nabla\phi(0) = 0$, and

$$\det_{1\leq j,k\leq n}\left(\frac{\partial^2\phi}{\partial x_j \partial x_k}\right)(0) \neq 0.$$

We divide the proof into three cases, depending on the position of $\eta \in \mathbf{S}^n$:

1. η is sufficiently close to the "north pole" $\eta_N = (0, \ldots, 0, 1)$,
2. η is sufficiently close to the "south pole" $\eta_S = (0, \ldots, 0, -1)$, and
3. η lies in the complementary set on the unit sphere.

We begin with the first case. We have that $\nabla_x\Phi(x, \eta_N)|_{x=0} = 0$ and want to see that for each η sufficiently close to η_N there is a (unique) $x = x(\eta)$ so that

$$\nabla_x\Phi(x,\eta)|_{x=x(\eta)} = 0.$$

The latter is a series of n equations, and one can find the desired solution by the implicit function theorem, which requires that we check that the Jacobian determinant

$$\det\left[\frac{\partial^2\Phi}{\partial x_j \partial x_k}\right](0, \eta_N) \neq 0,$$

but this is of course our assumption of nonvanishing curvature. Notice that if the η-neighborhood of η_N is sufficiently small, then

$$\det\left[\frac{\partial^2\Phi}{\partial x_j \partial x_k}\right](x(\eta), \eta) \neq 0;$$

so we can invoke Proposition 6 (with $x_0 = x(\eta)$), as long as the support of $\tilde\psi$ is small enough. This proves (28) when η is in the first region. The proof for η in the second region is the same.

Thus we are left with the third class of η. By definition,

$$\nabla_x \Phi(x,\eta) = (\eta_1, \ldots, \eta_n) + \eta_{n+1} \nabla\phi(x).$$

However, $(\eta_1^2 + \cdots + \eta_n^2)^{1/2} \geq c > 0$, since we are in case 3, and

$$\nabla\phi(x) = O(x) \qquad \text{as } x \to 0;$$

so $|\nabla_x \Phi(x,\eta)| \geq c' > 0$, if the support of $\tilde{\psi}$ is a sufficiently small neighborhood of the origin. Hence for the η in region 3, we may use Proposition 4 to conclude that the left side of (28) is actually $O(\lambda^{-N})$, for every $N \geq 0$. The proof of Theorem 1 is therefore concluded.

Remark. We have used only a special consequence of the asymptotic formula (21), namely the "remainder estimate" analogous to (8) when $N = r = 0$. Had we used the full formula, we could have gotten an asymptotic expansion for $\widehat{d\mu}(\xi)$; its main term is explicitly expressible in terms of the Gaussian curvature of those points $x \in S$ for which the normal is in the direction ξ or $-\xi$. For this, see §5.7 below.

3.2 Submanifolds of finite type. We shall now consider the problem in a wider setting. Here S will be a smooth m-dimensional submanifold of $\mathbf{R}^n$, with $1 \leq m \leq n-1$, and our assumptions regarding curvature will be replaced by the more general assumption that, at each point, S has at most a finite order of contact with any affine hyperplane. We say that such submanifolds are of *finite type.*

The precise definitions we require are as follows. We consider S in a sufficiently small neighborhood of a given point and write S as the image of a smooth mapping $\phi : U \to \mathbf{R}^n$, where U is a neighborhood of the origin in $\mathbf{R}^m$. To ensure that S is smoothly embedded, we suppose that the vectors $\partial\phi/\partial x_1, \ldots, \partial\phi/\partial x_m$ are linearly independent for each $x \in U$.

Now fix a point $x_0 \in U$, and a unit vector $\eta \in \mathbf{R}^n$. We assume that the function

$$[\phi(x) - \phi(x_0)] \cdot \eta$$

does not vanish to infinite order as $x \to x_0$; that is, for each $x_0 \in U$ and each unit vector $\eta \in \mathbf{R}^n$, there is a multi-index α, with $|\alpha| \geq 1$, so that

$$\partial_x^\alpha [\phi(x) \cdot \eta]\big|_{x=x_0} \neq 0.$$

Notice that if (x', η') is sufficiently close to (x_0, η), then also

$$\partial_x^\alpha [\phi(x) \cdot \eta']\big|_{x=x'} \neq 0.$$

The smallest k so that, for each unit vector η, there exists an α with $|\alpha| \leq k$ for which

$$\partial_x^\alpha [\phi(x) \cdot \eta]\big|_{x=x_0} \neq 0,$$

is called the *type* of ϕ (and the type of S) at x_0. Also, if $U_1 \subset U$ is a compact set, the type of ϕ in U_1 is defined to be the maximum of the types of the $x_0 \in U_1$.

The following remarks may help to clarify the notion that a point $x_0 \in S$ is of finite type.[†]

(i) When S is a curve in $\mathbf{R}^2$, the condition is equivalent with the curvature of S not vanishing to infinite order at x_0. Similarly, when S is a curve in $\mathbf{R}^3$, the condition is that neither the curvature nor the torsion of S vanishes to infinite order at x_0.

(ii) If S is a hypersurface in $\mathbf{R}^n$, the condition is equivalent with the condition that at least one of the principal curvatures of S does not vanish to infinite order at x_0.

(iii) When S is real-analytic, the condition is equivalent to S not lying in any affine hyperplane.

THEOREM 2. *Suppose S is a smooth m-dimensional manifold in $\mathbf{R}^n$ of finite type. Let $d\mu = \psi\, d\sigma$ be as above. Then*

$$|\widehat{d\mu}(\xi)| \leq A|\xi|^{-1/k}, \tag{29}$$

where k is the type of S inside the support of ψ.

Proof. By a suitable partition of unity, we can reduce the problem to the case where

$$\int_S e^{-2\pi i x\cdot\xi}\, d\mu(x) = \int_{\mathbf{R}^m} e^{-2\pi i \phi(x)\cdot\xi}\, \tilde{\psi}(x)\, dx$$

and the support of $\tilde{\psi} \in C_0^\infty$ is as small as we like.

Now we can write $\xi = \lambda\eta$ with $\lambda > 0$ and $|\eta| = 1$. Then we know that there is an α, with $|\alpha| \leq k$, so that

$$\partial_x^\alpha [\phi(x) \cdot \eta] \neq 0$$

for all $x \in \operatorname{supp}(\tilde{\psi})$, provided we take $\operatorname{supp}(\tilde{\psi})$ to be sufficiently small. Thus the conclusion (29) follows from (20) of Proposition 5.

As a final comment, let us observe that when S is real-analytic, the finite type condition is also *necessary* for the kind of decay asserted in Theorem 2. For otherwise, S would lie in an affine hyperplane (see remark (iii) above), and then $\widehat{d\mu}(\xi)$ would be constant for those ξ that are orthogonal to that hyperplane.

[†] There are several related notions of type, for instance those which measure the maximal order of contact with k-dimensional hyperplanes. For our present purposes, the case $k = n-1$ is the most relevant.

4. Restriction of the Fourier transform

The Fourier transform of an $L^1(\mathbf{R}^n)$ function is continuous, and hence is defined everywhere on $\mathbf{R}^n$. On the other hand, the Fourier transform of an L^2 function is itself no better than an L^2 function, and so can be defined only almost everywhere, and is thus completely arbitrary on a set of measure zero. In addition, when $1 < p \leq 2$, the classical Hausdorff-Young theorem[‡] allows one to realize the Fourier transform of an L^p function as an element of $L^q(\mathbf{R}^n)$, $1/p + 1/q = 1$, and so, at first sight, may be thought to be definable only almost everywhere.

In view of this, it is a remarkable fact that when $n \geq 2$ and S is a submanifold of $\mathbf{R}^n$ that has appropriate curvature, there is a $p_0 = p_0(S)$, with $1 < p_0 < 2$, so that every function in $L^p(\mathbf{R}^n)$, $1 \leq p < p_0$, has a Fourier transform that restricts to S. That this phenomenon was observed so late in the development of the subject is an indication of the slowness of our progress in understanding the genuinely n-dimensional aspects of Fourier analysis.

4.1 We begin by making the notion of restriction of the Fourier transform precise.

Suppose that S is a given smooth submanifold of $\mathbf{R}^n$ and that $d\sigma$ is its induced Lebesgue measure. We say that the L^p *restriction property* is valid for S if there exists a $q = q(p)$ so that the inequality

$$\left(\int_{S_0} |\widehat{f}(\xi)|^q \, d\sigma(\xi)\right)^{1/q} \leq A_{p,q}(S_0) \cdot \|f\|_{L^p} \tag{30}$$

holds for each $f \in \mathcal{S}$ whenever S_0 is an open subset of S with compact closure in S. Because $\mathcal{S}$ is dense in L^p we can, when (30) holds, define $\widehat{f}$ on S (a.e. with respect to $d\sigma$), for each $f \in L^p$.

The characterization of S having the restriction property and the determination of optimal ranges for the exponents p and q are difficult problems which have not yet been completely solved. However, the following simple and general observation can be made.

THEOREM 3. *Suppose S is a smooth m-dimensional submanifold of $\mathbf{R}^n$ of type k. Then there exists a $p_0 = p_0(S)$, $1 < p_0$, so that S has the L^p restriction property* (30) *with $q = 2$, and $1 \leq p \leq p_0$.*

Note. The initial analysis below will give

$$p_0 = \frac{2nk}{2nk-1};$$

improvements are given in §5.14 below and in Chapter 9, §2. Also, interpolating between $p = 1$ (for which one may take $q = \infty$) and $p = p_0$ (for which $q = 2$) gives an improvement in q for the intermediate p.[†]

[‡] See *Fourier Analysis*, Chapter 5.

[†] This is done explicitly in Chapter 9, §2.1.

Proof. It will suffice to prove that, for appropriate $\psi \geq 0$ that are smooth and have compact support,

$$\left(\int_S |\widehat{f}(\xi)|^2 \psi(\xi) \, d\sigma(\xi)\right)^{1/2} \leq A \cdot \|f\|_{L^p(\mathbf{R}^n)}, \quad \text{for } f \in \mathcal{S}.$$

Set $d\mu = \psi \, d\sigma$. We are dealing with the operator R, where $Rf(\xi)$ is defined for $\xi \in S$ by the Fourier transform

$$Rf(\xi) = \int_{\mathbf{R}^n} e^{-2\pi i x\cdot\xi} f(x) \, dx.$$

The question then is whether R is a bounded mapping from $L^p(\mathbf{R}^n)$ to $L^2(S, d\mu)$, and, in studying this, we consider also its formal adjoint R^*, given by

$$R^* f(x) = \int_S e^{2\pi i x\cdot\xi} f(\xi) \, d\mu(\xi)$$

for $x \in \mathbf{R}^n$.

We now invoke the orthogonality principle[‡] as follows: We have

$$\langle Rf, Rf \rangle_{L^2(S,d\mu)} = \langle R^* Rf, f \rangle_{L^2(\mathbf{R}^n)}.$$

So to prove that

$$R : L^p(\mathbf{R}^n) \to L^2(S, d\mu)$$

is bounded, it suffices (by Hölder's inequality) to see that

$$R^* R : L^p(\mathbf{R}^n) \to L^{p'}(\mathbf{R}^n)$$

is bounded, where p' is the exponent conjugate to p. However

$$(R^* Rf)(x) = \int_{\mathbf{R}^n} \int_S e^{2\pi i\xi\cdot(x-y)} \, d\mu(\xi) \, f(y) \, dy,$$

so $(R^* Rf)(x) = (f * K)(x)$ with

$$K(x) = \widehat{d\mu}(-x).$$

By Theorem 2 in §3.2, we have

$$|K(x)| \leq A|x|^{-1/k},$$

and clearly K is bounded, so

$$|K(x)| \leq A|x|^{-\gamma}, \quad \text{whenever} \quad 0 \leq \gamma \leq 1/k.$$

We now use the theorem of fractional integration: The operator $f \mapsto f * (|x|^{-\gamma})$ is bounded from $L^p(\mathbf{R}^n)$ to $L^q(\mathbf{R}^n)$, whenever $1 < p < q < \infty$ and $1/q = 1/p - 1 + \gamma/n$.

Let us momentarily take this assertion for granted. Then if $q = p'$, we have $1/q = 1 - 1/p$, so the relation among the exponents becomes $2 - 2/p = \gamma/n$, and the restriction $0 \leq \gamma \leq 1/k$ becomes $1 \leq p \leq (2nk)/(2nk-1)$, completing the proof of the theorem (since the case $p = 1$ is trivial).

[‡] As discussed in Chapter 7, §2.

4.2 Hardy-Littlewood-Sobolev Inequality. † There are many proofs of the inequality

$$\|f * (|y|^{-\gamma})\|_{L^q(\mathbf{R}^n)} \leq A_{p,q} \cdot \|f\|_{L^p(\mathbf{R}^n)} \tag{31}$$

for

$$0 < \gamma < n, \quad 1 < p < q < \infty, \quad \text{and} \quad \frac{1}{q} = \frac{1}{p} - \frac{n-\gamma}{n}. \tag{32}$$

For us, the most direct approach will be to write

$$[f * (|y|^{-\gamma})](x) = \int f(x-y)\,|y|^{-\gamma}\,dy = \int_{|y|<R} + \int_{|y|\geq R}.$$

The first integral is the convolution of f with the function

$$|y|^{-\gamma}\chi_{B(R)}(y),$$

where $\chi_{B(R)}$ is the characteristic function of the ball of radius R centered at the origin. Since this function is radial, decreasing, and integrable, we can use the majorization given by the (usual) maximal function M (see Chapter 2, §2.1) to see that this integral is bounded by

$$(Mf)(x) \cdot \int_{|y|\leq R} |y|^{-\gamma}\,dy = cR^{n-\gamma} \cdot (Mf)(x).$$

By Hölder's inequality, the second integral is dominated by

$$\|f\|_{L^p(\mathbf{R}^n)} \cdot \|\,|y|^{-\gamma} \cdot \chi_{{}^cB(R)}\|_{L^{p'}(\mathbf{R}^n)}.$$

Now $|y|^{-\gamma} \cdot \chi_{{}^cB(R)} \in L^{p'}(\mathbf{R}^n)$ when $-\gamma p' < -n$ and, in view of (32),

$$\gamma p' - n = \frac{np'}{q} > 0.$$

Thus

$$\|\,|y|^{-\gamma} \cdot \chi_{{}^cB(R)}\|_{L^{p'}(\mathbf{R}^n)} = cR^{-n/q}.$$

Summing the two integrals, we have

$$|(f * |y|^{-\gamma})(x)| \leq A[(Mf)(x) \cdot R^{n-\gamma} + \|f\|_{L^p} \cdot R^{-n/q}].$$

Finally, choose R so that both terms on the right side are equal, i.e.,

$$\frac{(Mf)(x)}{\|f\|_{L^p}} = R^{-n+\gamma-n/q} = R^{-n/p}.$$

Substituting this in the above gives

$$|(f * |y|^{-\gamma})(x)| \leq A \cdot [(Mf)(x)]^{p/q} \cdot \|f\|_{L^p}^{1-p/q}. \tag{33}$$

The inequality (31) then follows from the usual L^p inequality for the maximal operator M.

† A more detailed discussion of this result (and a somewhat different proof) may be found in *Singular Integrals*, Chapter 5, §1.

4.3 The initial restriction theorem just proved may be thought of as a result for oscillatory integrals of the second kind. In this case, the operator is fashioned out of the Fourier transform, with its oscillatory factor $e^{2\pi i x\cdot\xi}$. We turn to more general operators of this character in the next chapter.

5. *Further results*

A. Asymptotics

5.1 The following more exact formulas are useful in the study of the asymptotics of oscillatory integrals.

(a) In $\mathbf{R}^n$, suppose that the phase function $\phi(x)$ is $|x|^2$; then the explicit version of the formula (21) is

$$\int_{\mathbf{R}^n} e^{i\lambda|x|^2}\,\psi(x)\,dx \sim \lambda^{-n/2}\sum_{j=0}^{\infty} a_j\lambda^{-j}, \qquad a_j = (i\pi)^{n/2}\frac{i^j}{j!}(\Delta^j\psi)(0).$$

(b) There is a similar formula when $|x|^2$ is replaced by $\langle Ax, x\rangle$, with A a real, symmetric, invertible matrix. Then the above asymptotics hold with

$$a_j = \left(\frac{i\pi}{\det A}\right)^{n/2}\frac{i^j}{j!}(\Delta_A)^j\psi(0), \qquad \Delta_A = \sum_{j,k} a^{j,k}\frac{\partial^2}{\partial x_j\partial x_k},$$

and $(a^{j,k})$ is the inverse matrix to A.

(c) The "endpoint" contributions mentioned in §1.1 can be described as follows. Suppose $\psi \in C_0^\infty(\mathbf{R})$. Then one has the asymptotic formula

$$\int_0^\infty e^{i\lambda x}\,\psi(x)\,dx \sim \sum_{j=0}^{\infty} a_j\lambda^{-j-1}$$

with $a_j = i^{j+1}\psi^{(j)}(0)$.

(d) More generally,

$$\int_0^\infty e^{i\lambda x}\,\psi(x)\,x^\mu\,dx \sim \sum_{j=0}^{\infty} a_j\lambda^{-j-1-\mu}, \quad a_j = i^{j+\mu+1}\frac{j!}{\Gamma(j+\mu+1)}\psi^{(j)}(0),$$

provided that $\mathrm{Re}(\mu) > -1$.

To prove (a), one uses the identity

$$\int_{\mathbf{R}^n} e^{-\pi\delta|x|^2}\,\phi(x)\,dx = \delta^{-n/2}\int_{\mathbf{R}^n} e^{-\pi|\xi|^2/\delta}\,\widehat{\phi}(\xi)\,d\xi$$

for $\delta > 0$ and passes to $\delta = -i\lambda/\pi$. This gives

$$\left(\frac{i\pi}{\lambda}\right)^{n/2}\int_{\mathbf{R}^n} e^{-i\pi^2|\xi|^2/\lambda}\,\widehat{f}(\xi)\,d\xi$$

for the integral, and the result follows by the use of the expansion

$$e^{-i\pi^2|\xi|^2/\lambda} = \sum_{j=0}^{\infty}(-i\pi^2|\xi|^2/\lambda)^j/j!.$$

The proof of (b) is similar.

To prove (d) (and thus (c)) it suffices, using a device akin to that in §1.3, to obtain it in the special case where $\psi = \tilde{\psi}$ for $\tilde{\psi} \in C_0^\infty$, with $\tilde{\psi} \equiv 1$ near the origin. In this case, one has

$$\int_0^\infty e^{i\lambda x}\,\tilde{\psi}(x)\,x^\mu\,dx \sim \frac{i^{\mu+1}}{\Gamma(\mu+1)}\lambda^{-\mu-1}, \qquad \lambda \to \infty.$$

This last assertion is a consequence of the identity

$$\lim_{\varepsilon\to 0+}\int_0^\infty e^{i\lambda x}\,e^{-\varepsilon x}\,x^\mu\,dx = \frac{i^{\mu+1}}{\Gamma(\mu+1)}\lambda^{-\mu-1},$$

and the fact that

$$\sup_{\varepsilon>0}\left|\int_0^\infty e^{i\lambda x}\,e^{-\varepsilon x}\,x^\mu[1-\tilde{\psi}(x)]\,dx\right| = O(|\lambda|^{-N}),$$

as $\lambda \to \infty$, for all $N > 0$.

Standard texts dealing with the asymptotics of such integrals (and related topics) are Erdelyi [1956], Copson [1965]; for (b), see also Hörmander [1983].

5.2 The classical reference for Bessel functions is G. N. Watson [1922]. Here we state some results in that subject that can be obtained by the methods described above.

(a) When m is fixed, the complete asymptotic formula for $J_m(r)$, as $r \to \infty$, is

$$J_m(r) \sim \left(\frac{\pi r}{2}\right)^{-\frac{1}{2}}\cos\left(r - \frac{m\pi}{2} - \frac{\pi}{4}\right)\sum_{j=0}^{\infty} a_j r^{-2j}$$
$$+\left(\frac{\pi r}{2}\right)^{-\frac{1}{2}}\sin\left(r - \frac{m\pi}{2} - \frac{\pi}{4}\right)\sum_{j=0}^{\infty} b_j r^{-2j-1},$$

where $a_j = (-1)^j(m,2j)2^{-2j}$ and $b_j = (-1)^j(m,2j+1)2^{-2j-1}$, with

$$(m,k) = \frac{\Gamma(\frac{1}{2}+m+k)}{k!\cdot\Gamma(\frac{1}{2}+m-k)}.$$

(b) A simple consequence is the estimate $|J_m(r)| \le A_m r^{-1/2}$ as $r \to \infty$, for each fixed m. However, if one seeks a uniform bound for large r and m, then the best one can do is

$$|J_m(r)| \le Ar^{-1/3}, \qquad r \ge 1.$$

(c) That this estimate is best possible can be seen by the fact that

$$J_m(m) = cm^{-1/3} + O(m^{-2/3}) \qquad \text{as } m \to \infty,$$

with $c = \dfrac{\Gamma(1/3)}{\pi 2^{2/3}3^{1/6}}$.

When m is an integer, the asymptotic formula in (a) is a consequence of the formula (9) and the proof given in §1.3. For general m, one can use formula (16) and part (d) of §5.1; an alternate approach, using contour integration, is described in *Fourier Analysis*, Chapter 4, §3.

When m is integral, the inequality in (b) follows from the proposition in §1.2 and the fact that, if $\phi(\theta) = \sin(\theta) - m\theta/r$, then $|\phi''(\theta)| + |\phi'''(\theta)| \ge c > 0$. The asymptotic formula in (c) is a consequence of the proposition in §1.3 and the fact that, when $m = r$, $\phi(0) = \phi'(0) = \phi''(0) = 0$, but $\phi'''(0) \ne 0$.

5.3 The calculation of the Fourier transform of the Riemann singularity (carried out in §1.4.2) can be put into the setting of $\mathbf{R}^n$, and also generalized, as follows. Consider the function $e^{i|x|^{-\alpha}}|x|^{-\gamma}$, where $0 < \alpha < \infty$. Then, in a sense to be made precise below, its Fourier transform behaves like $c_1|\xi|^{-b}e^{ic_2|\xi|^a}$ for large ξ, with $0 < a < 1$.

(i) Let $K_\gamma(x) = e^{i|x|^{-\alpha}}|x|^{-\gamma}\phi(x)$, where $\phi \in C_0^\infty(\mathbf{R}^n)$ and $\phi(x) = 1$ for x near the origin. Then the distribution-valued function $\gamma \mapsto K_\gamma$, initially defined for $\mathrm{Re}(\gamma) \ge 0$ as above, continues analytically to the entire complex plane. The Fourier transform $\widehat{K_\gamma}$ is a function (since K_γ is compactly supported), and

$$\widehat{K_\gamma}(\xi) = c_1|\xi|^{-b}e^{ic_2|\xi|^a} + O(|\xi|^{-b-(a/2)}) \qquad \text{as } |\xi| \to \infty.$$

Here $a^{-1} - \alpha^{-1} = 1$, so that $a = \alpha/(\alpha+1)$, and $0 < a < 1$ because $0 < \alpha < \infty$; moreover $b = n - [\gamma + (n\alpha/2)]/(\alpha+1)$.

(ii) In the reverse direction, consider the function $m_b(\xi) = |\xi|^{-b}e^{i|\xi|^a}\eta(\xi)$, where $0 < a < 1$, and $\eta \in C^\infty$ vanishes near the origin and equals 1 for large $|\xi|$. Then $m_b = \widehat{K}$, where the distribution K agrees with a function $K(x)$ away from the origin, and

$$K(x) = c_3 e^{ic_4|x|^{-\alpha}}\cdot|x|^{-\gamma} + O(|x|^{-\gamma+(\alpha/2)}), \qquad \text{as } x \to 0.$$

Again $a^{-1} - \alpha^{-1} = 1$, so that $\alpha = a/(1-a)$; also $\gamma = [n - b + (n\alpha/2)]/(1-\alpha)$, and we assume that $\gamma > 0$.

For these, and other results of this kind, see Hardy [1913] for the case $n = 1$, Wainger [1965] for the case $n \ge 1$, and also the discussion immediately below.

5.4 An interesting heuristic principle underlying the asymptotics of Fourier transforms is that of *duality of phases*. Under suitable circumstances, it can be formulated broadly as follows: The Fourier transform of $e^{i\Phi(x)}a(x)$ is essentially of the form $e^{-i\Psi(\xi)}a^*(\xi)$, where the pair (Φ, Ψ) of phases are "dual" to each other. The amplitude a is assumed to be nonoscillatory but "regularly varying"; then a^* has similar properties. We elaborate three variants of this general principle in the setting of $\mathbf{R}^1$.

(i) The first variant arises when $\Phi(x)$ is defined for large x (say $x \geq 1$), is strictly *convex*, and $\lim_{x\to+\infty} \Phi'(x) = +\infty$. Then, for appropriate $a(x)$,

$$\int_1^\infty e^{i[\Phi(x)-x\xi]}\, a(x)\, dx \sim (-2\pi i)^{1/2} e^{-i\Psi(\xi)} \cdot \Phi''(\Psi'(\xi))^{-1/2} \cdot a(\Psi'(\xi)),$$

in the sense that the quotient of both sides tends to 1 as $\xi \to +\infty$.

Here Ψ is defined by $\Psi(\xi) = \sup_x(x\xi - \Phi(x))$. As a result, Φ' and Ψ' are inverses of each other. A typical example of this situation is the pair (Φ, Ψ) that arises in Hölder's inequality, namely $\Phi(x) = x^p/p$, $\Psi(\xi) = \xi^q/q$, with $1 < p < \infty$ and $p^{-1} + q^{-1} = 1$. More generally, we may take a pair of functions that are complementary in the sense of Young's inequality.‡

(ii) A second version occurs when $\Phi(x)$ is strictly *concave*, increasing in $(1, \infty)$, and with $\Phi'(x) \to 0$ as $x \to +\infty$. Then the relevant asymptotics are for small ξ, and the above formula holds for $\xi \to 0^+$. Here $\Psi(\xi) = \inf_x(x\xi - \Phi(x))$. Again Φ' and Ψ' are inverses of each other; examples are given by $\Phi(x) = x^p/p$, $\Psi(\xi) = \xi^q/q$, this time for $0 < p < 1$ and $-\infty < q < 0$, with $p^{-1} + q^{-1} = 1$. This example is treated in §5.3(ii) above.

(iii) The third variant arises when Φ is defined on $(0, 1)$, is strictly *convex*, and $\Phi(x) \to \infty$ as $x \to 0^+$. The relevant asymptotics are

$$\int_0^1 e^{i[\Phi(x)+x\xi]}\, a(x)\, dx \sim (2\pi i)^{1/2} e^{i\Psi(\xi)} \cdot \Phi''(\Psi'(\xi))^{-1/2} \cdot a(\Psi'(\xi)), \qquad \xi \to +\infty.$$

Here $\Psi(\xi) = \inf_x(\Phi(x) + x\xi)$; as a result, $-\Phi'$ and Ψ' are inverses of each other. Examples are $\Phi(x) = -x^p/p$, $\Psi(\xi) = \xi^q/q$, with $-\infty < p < 0$, $0 < q < 1$, and $p^{-1} + q^{-1} = 1$. These examples are treated in §5.3(i) above.

We have limited ourselves here to the statement of general principles. However, once these assertions are made more precise, they can be deduced by the method of stationary phase. Early instances had already been indicated (at least implicitly) by Riemann [1854]. More precise and more general conclusions (which can be viewed as further illustrations of these principles) were obtained by Hardy [1913], under the assumptions that Φ and a are of "logarithmic-exponential" type.

‡ See Zygmund [1959], Chapter 1, for a definition of complementary functions and a proof of Young's inequality.

5.5 A general qualitative result describing the asymptotics of integrals in $\mathbf{R}^n$ is as follows. Suppose the phase ϕ is real and analytic near the origin, with $\phi(0) = 0$ and $(\nabla\phi)(0) = 0$. If the support of ψ is sufficiently small, then

$$\int_{\mathbf{R}^n} e^{i\lambda\phi}\, \psi\, dx \sim \sum_{\alpha,k} a_{\alpha,k} \lambda^\alpha (\log \lambda)^k, \qquad \text{as } \lambda \to \infty.$$

Here α ranges over a sequence of negative rationals lying in an arithmetic progression; the k are nonnegative integers with $0 \leq k < n$.

In this connection, the following comments are in order.

(i) One has similar asymptotics for the integrals $I_\pm(\lambda) = \int_{\pm\phi>0} e^{i\lambda\phi}\psi\, dx$. Indeed, in the case I_+, the corresponding asymptotics are valid (for $\lambda \to \infty$) in the half-plane $\text{Im}(\lambda) \geq 0$; a similar result holds for I_- in the half-plane $\text{Im}(\lambda) \leq 0$.

(ii) The asymptotics of $I_\pm(\lambda)$ are closely connected with the meromorphic continuation of $J_\pm(s) = \int_{\pm\phi>0} (\pm\phi)^s\, \psi\, dx$ into the half-plane $\text{Re}(s) < 0$. In fact,

$$\int_0^\infty I_+(i\mu)\, \mu^{s-1}\, d\mu = \Gamma(s) J_+(-s),$$

and

$$I_+(\lambda) = \frac{1}{2\pi i} \int_\gamma \Gamma(s)\, (i\lambda)^{s-1}\, J_+(s-1)\, ds$$

when $\text{Im}(\lambda) > 0$ and γ is a suitable contour.

(iii) The proofs of the meromorphic continuation of $J_\pm(s)$ (and thus of the asymptotics of $I_\pm(\lambda)$) can be carried out by "resolving" the singularity of Φ: There are neighborhoods U, V of the origin in $\mathbf{R}^n$ and a smooth proper mapping $\pi : U \to V$, so that

$$\phi(\pi(y)) = y_1^{k_1} \cdot \cdots \cdot y_n^{k_n} \cdot \phi^*(y),$$

where ϕ^* is a smooth, $\phi^*(0) \neq 0$, and $k_1, \ldots, k_n$ are nonnegative integers. The above assertions are then reduced to the corresponding statements when $\phi(x) = x_1^{k_1} \cdot \cdots \cdot x_n^{k_n}$.

The meromorphic continuation of $J_\pm(s)$ using resolution of singularities goes back to Bernstein and Gelfand [1969], Atiyah [1970]. For the asymptotics of $I_\pm(\lambda)$, see Malgrange [1974], Varčenko [1976], where a variety of special cases are treated in greater detail.

5.6 In certain cases, which cover a broad class of examples, one can give a precise description of the order of decrease at infinity of the integral $\int_{\mathbf{R}^n} e^{i\lambda\phi}\, \psi\, dx$ in terms of the Newton diagram of ϕ. We are concerned with the largest α for which there is a ψ so that $a_{\alpha,k}$ is nonzero for some k. This is the *oscillatory index* of ϕ. One can find it as follows.

Write $\phi(x) = \sum_{|k|>0} c_k x^k$, and consider the (unbounded) polyhedron $P \subset \mathbf{R}^n$ that is the union of all "octants" with vertex k, where $c_k \neq 0$, that is,

$$P = \bigcup_{c_k \neq 0} \{x \in \mathbf{R}^n : x_j \geq k_j,\ j = 1, \ldots, n\}.$$

The *Newton diagram* of ϕ is then the union of all the compact faces of the convex hull of P. If γ is such a face, let $\phi_\gamma = \sum_{k\in\gamma} c_k x^k$. We assume that ϕ is *nondegenerate* in the sense that $\nabla\phi_\gamma$ does not vanish (except at the origin) for each γ. Let β be the smallest number so that $(\beta, \ldots, \beta)$ is in the Newton diagram of ϕ. If we also suppose that $\beta > 1$, then the oscillatory index equals $-1/\beta$. Varčenko [1976].

B. Fourier transforms of surface-carried measures

5.7 Suppose S has nonzero Gaussian curvature, and let $d\mu = \psi\, d\sigma$, where $\psi \in C_0^\infty$ and $d\sigma$ is the induced Lebesgue measure on S. Then the estimate given by Theorem 1 (see (27)) can be replaced by a more precise asymptotic statement, which can be formulated as follows.

We assume that the support of ψ is so small that, for each $\xi \in \mathbf{R}^n \backslash \{0\}$, there is at most one point $\bar{x} = \bar{x}(\xi) \in S$ with a normal $\nu_{\bar{x}}$ pointing in the direction of ξ. In that case, after a possible translation and rotation, we may take $\bar{x}$ to be the origin, $\nu_{\bar{x}} = (0, \ldots, 0, 1)$; in a neighborhood of $\bar{x}$, we can represent S as a graph $\{x : x_n = \phi(x_1, \ldots, x_{n-1})\}$. The positive (negative) eigenvalues of $\left(\dfrac{\partial^2\phi}{\partial x_j \partial x_k}\right)(0)$ are then the positive (negative) principal curvatures of S, at $\bar{x}$, in the direction of ξ.

With these definitions we can state the result:

$$\widehat{d\mu}(\xi) = a(\xi) e^{-2\pi i \phi(\xi)} + e(\xi), \qquad |\xi| \geq 1.$$

Here $a(\xi)$ is nonzero only for those ξ for which $\bar{x}(\xi)$ is defined and lies in the support of ψ. For these ξ, the phase $\phi(\xi)$ is real and homogeneous of degree 1, and is given by $\phi(\xi) = \bar{x}(\xi) \cdot \xi$. Moreover (in the terminology of Chapter 6, §1.3), $a(\xi)$ is a symbol belonging to the class $S^{(1-n)/2}$, and, more precisely, for large ξ we have

$$a(\xi) = |\xi|^{(1-n)/2} \psi(\bar{x}(\xi)) \cdot |K(\bar{x}(\xi))|^{-1/2} e^{-i\pi d/4} \qquad \operatorname{mod} S^{-(n+1)/2},$$

where $|K(x)|$ is the absolute value of the Gaussian curvature at $\bar{x}$ (the product of the principal curvatures), and d is the excess of the number of positive curvatures (over the number of negative curvatures) in the direction ξ. The error term $e(\xi)$ is a symbol in the class $S^{-\infty}$. For these computations, see Hlawka [1950] and Herz [1962a].

Anticipating matters that will be dealt with more fully in the next chapter, we point out that $\Phi(x,\xi) = x \cdot \xi - \phi(\xi)$ obviously satisfies the nondegeneracy condition required of the phase of a Fourier integral operator (as will be set down in §3 of Chapter 9). As a result, $Tf = f * d\mu$ is, after we interpose a smooth cut-off function in x, a Fourier integral operator of order $(1-n)/2$. One can also verify that in this case

$$\operatorname{rank}(\Phi_{\xi\xi}) = \operatorname{rank}(\phi_{\xi\xi}) = n - 1,$$

because of the curvature condition imposed on S.

5.8 The next series of results deal with the situation that arises if the Gaussian curvature of our hypersurface is no longer required to be nonvanishing at every point.

Given a surface S, let $d\mu = \psi\, d\sigma$ be as above and assume that, for all $x \in \operatorname{supp}(\psi) \cap S$, at least k of the principal curvatures are not zero. Then a modification of the proof of Theorem 1 shows that

$$|\widehat{d\mu}(\xi)| \leq A|\xi|^{-k/2};$$

see Littman [1963].

5.9 Suppose that S is the boundary of a bounded smooth convex domain $U \subset \mathbf{R}^n$; this implies that, at each point of S, all the principal curvatures in the direction of the inner normal are nonnegative. If we also assume that S is real analytic, then with $d\mu = \psi\, d\sigma$ as above, we have

$$|\widehat{d\mu}(\xi)| \leq \Omega(\xi) \cdot |\xi|^{(1-n)/2}, \qquad |\xi| \geq 1,$$

where Ω is homogeneous of degree 0; we also have that $\Omega \in L^p(S^{n-1})$ for some $p > 2$.

A similar result holds if the assumption of analyticity is replaced by the requirement that any line in $\mathbf{R}^n$ has at most a finite order of contact with S.

The proof proceeds by showing that, if the support of ψ is sufficiently small, then as $|\xi| \to \infty$,

$$|\widehat{d\mu}(\xi)| \leq A\psi(\bar{x}(\xi)) \cdot |K(\bar{x}(\xi))|^{-1/2} |\xi|^{(1-n)/2} + O(|\xi|^{-N})$$

for each $N > 0$; here $\bar{x}(\xi)$ is the (unique) point in S whose normal is in the direction ξ. Further details are in Randol [1969a], [1969b], Svensson [1971].

5.10 As in §5.9, we assume that S is the boundary of a bounded smooth convex domain in $\mathbf{R}^n$ such that all lines in $\mathbf{R}^n$ have at most a finite order of contact with S. We indicate here a further relation between the geometry of S and the decay of Fourier transforms of smooth measures supported on S.

For each $x \in X$, let $B(x,\delta)$ denote the "ball" in S that is given as the "cap" of points in S whose distance from the tangent plane T_x is less than δ. One can then assert the following.

(i) The collection of balls $\{B(x,\delta)\}$ on the space S (equipped with the induced Lebesgue measure $d\sigma$) satisfies the basic engulfing and doubling postulates set forth in Chapter 1, §1.

(ii) If $d\mu = \psi\, d\sigma$, then the order of decrease of $\widehat{d\mu}(\xi)$ is controlled by these balls. In fact, if the support of ψ is sufficiently small, then

$$|\widehat{d\mu}(\xi)| \leq c\sigma(B(\bar{x},\delta)),$$

if $\bar{x} = \bar{x}(\xi) \in \operatorname{supp}(\psi)$ and $\delta = |\xi|^{-1}$. If there is no such $\bar{x} \in \operatorname{supp}(\psi)$, then $\widehat{du}(\xi)$ is rapidly decreasing in $|\xi|$. Bruna, Nagel, and Wainger [1988].

5.11 The results in §5.9 suggest that we could hope to restore the optimal decay

$$O(|\xi|^{(1-n)/2})$$

for the Fourier transform of a smooth measure supported on a smooth hypersurface S, if we introduce in the integration an appropriate mitigating factor that depends on the Gaussian curvature.

(a) In general, this may be done as follows. Let $S \subset \mathbf{R}^n$ be a smooth hypersurface and let K be its Gaussian curvature (on which no restrictions are placed). We then have

$$\int_S e^{-2\pi i x\cdot\xi}|K(x)|^N \psi(x)\,dx = O(|\xi|^{(1-n)/2})$$

whenever $N \geq 2n-2$.

(b) It is natural to try to find the "smallest" N for which the above optimal estimate holds. If we make the further assumption that S is of the special type treated in §5.10, we can obtain a sharper result: The estimate

$$\int_S e^{-2\pi i x\cdot\xi} a(x)\,\psi(x)\,dx = O(|\xi|^{(1-n)/2})$$

holds whenever a is smooth and $|a(x)| \leq c|K(x)|^{1/2}$.

For (a), see Sogge and Stein [1985]; for (b), see Cowling, Disney, Mauceri, and Müller [1990].

5.12 The study of the asymptotics of Fourier transforms of measures supported on surfaces had its initial impetus in number theory, in the context of questions relating to the distribution of lattice points. One formulation of the general problem dealt with is as follows.

Suppose $\Omega \subset \mathbf{R}^n$ is a bounded open domain with smooth boundary. Let $N(r)$ be the cardinality of $\mathbf{Z}^n \cap r\Omega$. An easy geometric argument shows that $N(r) \sim r^n|\Omega|$ as $r \to \infty$, and more accurately that

$$N(r) - r^n|\Omega| = O(r^{n-1}) \qquad \text{as } r \to \infty.$$

The determination of the true nature of the error term is a fascinating and largely unsolved problem which has a long history. This sought-after result has proved highly resistant, even in its two classical special cases:

(1) The situation when $\Omega \subset \mathbf{R}^2$ is the interior of the unit disc (where $N(r)/r^2$ gives the average number of representations of integers $< r^2$ as the sum of two squares), and

(2) The variant in which $\Omega \subset \mathbf{R}^2$ is the unbounded region between the coordinate axes and the hyperbola $xy = 1$ (which gives an analogous problem about the average number of divisors of integers).

The starting point for the analysis of these problems can be taken to be the Poisson summation formula, which in this case states, at least heuristically, that

$$\sum_{m\in\mathbf{Z}^n} \chi_r(m) = \sum_{m\in\mathbf{Z}^n} \widehat{\chi_r}(m);$$

here χ_r is the characteristic function of $r\Omega$ and $\widehat{\chi_r}$ is its Fourier transform. One observes that $\widehat{\chi_r}(\xi) = \widehat{\chi_1}(r\xi)\cdot r^n$ and that $\widehat{\chi_1}$ is closely related to $\widehat{d\mu}$, where $d\mu = \psi\,d\sigma$ is a smooth measure supported on the boundary S of Ω.

By the use of §5.7, we can then prove that if Ω is convex, bounded, and S has strictly positive Gaussian curvature at each point (i.e., Ω is strictly convex), then

$$N(r) - r^n|\Omega| = O(r^{n-[2n/(n+1)]}) \qquad \text{as } r \to \infty.$$

When $n = 2$, this error term is $O(r^{2/3})$. In the case of the disc, there has been a series of improvements to this result, replacing $O(r^{2/3})$ with $O(r^\alpha)$, for various α with $1/2 < \alpha < 2/3$; it has also been shown that $O(r^{1/2})$ is not possible.

The theory for the circle, and the parallel theory for the hyperbola, is due to Voronoi, Sierpinski, Hardy, and van der Corput (among others). It is set out in Landau [1927] (part 8), and Titchmarsh [1951] (Chapter 12). The general n-dimensional result stated above is in Hlawka [1950]; also see Herz [1962b].

5.13 The Fourier transforms of measures carried on surfaces defined by homogeneous functions can be treated with the aid of the following.

(a) Let ϕ be a real, homogeneous polynomial on $\mathbf{R}^n$ of degree $k \geq 2$ that is nondegenerate, in the sense that $\det\left(\dfrac{\partial^2\phi}{\partial x_i \partial x_j}\right) \neq 0$ whenever $x \neq 0$. Then if $\psi \in C_0^\infty$,

$$\int_{\mathbf{R}^n} e^{i[\lambda\phi(x)+\xi\cdot x]}\psi(x)\,dx = O((|\lambda| + |\xi|)^{-n/k}).$$

(b) More generally, suppose ϕ is a real function that is homogeneous of degree k and is smooth away from the origin. Assume ϕ has rank r; by this we mean that the matrix $\left(\dfrac{\partial^2\phi}{\partial x_i \partial x_j}\right)$ has rank at least r, whenever $x \neq 0$. Then the integral is $O((|\lambda| + |\xi|)^{-n/k})$, as long as $r > 2n/k$.

Notice that the rank r does not enter into the order of decay, but only in its relation to k and n. The situation when $r \leq 2n/k$ is different. The most interesting example of this arises when $k = 1$, to which we now turn.

(c) Suppose ϕ is real, homogeneous of degree 1, smooth away from the origin, and that ϕ has rank r (as in (b)); necessarily, $r \leq n-1$. Then

$$\int_{\mathbf{R}^n} e^{i[\lambda\phi(x)+\xi\cdot x]}\psi(x)\,dx = O((|\lambda| + |\xi|)^{-r/2}).$$

Integrals of this type are relevant to the study of Fourier integral operators (which are treated in the next chapter).

Note that (a), when $k = 2$, is essentially contained in §3.1.

To prove (b), let $I_0(\lambda,\xi) = \int_{\mathbf{R}^n} e^{i[\lambda\phi(x)+\xi\cdot x]}\psi_0(x)\,dx$, where ψ_0 is a suitable cut-off function that vanishes near the origin. Then the integral above is essentially

$$\sum_{j=0}^{\infty} 2^{-nj}\cdot I_0(2^{-jk}\lambda, 2^{-j}\xi).$$

Using the arguments that enter in §5.8, one shows that $|I_0(\mu,\xi)| \le A|\mu|^{-r/2}$, and of course $|I_0| \le A$. Hence the sum is dominated by

$$A|\lambda|^{-r/2}\sum_{2^{-jk}|\lambda|\ge 1}(2^{-jk})^{-r/2}\cdot 2^{-nj} + A\sum_{2^{-jk}|\lambda|\le 1} 2^{-nj} \le A'|\lambda|^{-n/k},$$

which proves the required assertion for $|\xi| \le c|\lambda|$. When $|\xi| \ge c|\lambda|$ (and c is sufficiently large), one uses the fact that $I_0(\lambda,\xi)$ is rapidly decreasing in $|\xi|$, which follows from §2.1.

The proof of (c) is similar, except here we write

$$\sum_{j\ge 0} I_0(2^{-j}\lambda, 2^{-j}\xi)\cdot 2^{-nj} = O(|\lambda|^{-r/2})\sum_{j\ge 0} 2^{rj/2}\cdot 2^{-nj} = O(|\lambda|^{-r/2}).$$

C. Restriction theorems

5.14 We consider two re-interpretations of the (L^p, L^2) restriction property for a submanifold S in $\mathbf{R}^n$.

With S given and $d\mu = \psi\,d\sigma$, where $\psi \in C_0^\infty$ is nonnegative, let R denote the "restriction" operator

$$(Rf)(\xi) = \int_{\mathbf{R}^n} e^{-2\pi i x\cdot\xi} f(x)\,dx, \qquad \xi \in S,$$

and let R^* be its dual, $(R^*f)(x) = \int_S e^{2\pi i x\cdot\xi} f(\xi)\,d\mu(\xi)$, $x \in \mathbf{R}^n$. The orthogonality argument in §4.1 led us to consider the operator $U = R^*R$, which is given by $Uf = f * K$, with $K(x) = \widehat{d\mu}(-x)$. The reasoning there shows that the following three assertions are equivalent; they deal with the boundedness of operators that are initially defined on the Schwartz class $\mathcal{S}$.

(i) $R : L^p(\mathbf{R}^n) \to L^2(S, d\mu)$ is bounded (the (L^p, L^2) restriction property).

(ii) $U : L^p(\mathbf{R}^n) \to L^{p'}(\mathbf{R}^n)$ is bounded; here p' is the dual exponent to p.

(iii) $f \mapsto \int_S \widehat{f}(\xi)e^{2\pi i x\cdot\xi}\,d\mu(\xi)$ gives a bounded operator from $L^p(\mathbf{R}^n)$ to $L^{p'}(\mathbf{R}^n)$.

5.15 The proof of Theorem 3 exploited only the size condition on the kernel K of the operator U. Anticipating the more refined analysis to be given in the next chapter, we can state the following better results:

(a) For the sphere $S^{n-1} \subset \mathbf{R}^n$, we have the (L^p, L^2) restriction property if and only if

$$1 \le p \le \frac{2n+2}{n+3}, \qquad \text{i.e.,} \quad \frac{1}{p} - \frac{1}{p'} \ge \frac{2}{n+1}.$$

A similar result holds for any smooth hypersurface S that has nonvanishing Gaussian curvature at every point.

(b) More generally, let S have dimension $m < n$, and suppose that we subsume its curvature properties in the inequality $|\widehat{d\mu}(x)| = O(|x|^{-r})$ where (as usual) $d\mu = \psi\,d\sigma$. Then the $(L^p, L^2(d\mu))$ restriction property holds whenever

$$\frac{1}{p} - \frac{1}{p'} \ge \frac{n-m}{r+n-m}.$$

Note that this gives an improvement for the exponent p_0 in Theorem 3.

(c) In particular, if S is a hypersurface (i.e., $m = n-1$), and at least k of the principal curvatures are nonzero at each point of S, then the restriction property holds for

$$1 \le p \le \frac{2k+4}{k+4},$$

which is the same range as occurs in (a) when the ambient space $\mathbf{R}^n$ has dimension $n = k+1$.

A sketch of the proof is as follows. Let U^s denote the analytic family of operators given by $\widehat{U_s f}(\xi) = m_s(\xi)\widehat{f}(\xi)$, where

$$m_s(\xi) = \gamma(s)\cdot \operatorname{dist}(\xi, S)^{-n+m-s}\tilde{\psi}(\xi);$$

here $\tilde{\psi} \in C_0^\infty$ is a suitable cut-off function, and $\gamma(s)$ is an appropriate analytic function that has a simple zero at $s = 0$. Then $U^0 = U$; moreover, when $\mathrm{Re}(s) = n-m$, the operators U^s map $L^2(\mathbf{R}^n)$ to itself, while if $\mathrm{Re}(s) = -r$, then the operators U^s map $L^1(\mathbf{R}^n)$ to $L^\infty(\mathbf{R}^n)$. The desired result for U (and hence for R) is then a consequence of the interpolation theorem in Chapter 9, §1.2.5.

For (a), see also Tomas [1975], [1979], Stein [1986]; (b) is in Greenleaf [1981].

5.16 The (L^p, L^2) restriction theorem described in §5.15 (a) and (c) above extends to unbounded quadric surfaces. We consider three cases.

(a) The paraboloid $S = \{\xi \in \mathbf{R}^n : \xi_n = Q'(\xi')\}$. Here we have written $\xi' = (\xi_1, \dots, \xi_{n-1}) \in \mathbf{R}^{n-1}$; Q' is a nondegenerate quadratic form on $\mathbf{R}^{n-1}$. We take $d\mu = d\xi'$. Then

$$\left(\int_S |\widehat{f}(\xi)|^2\,d\mu\right)^{1/2} = \left(\int_{\mathbf{R}^{n-1}} |\widehat{f}(\xi', Q'(\xi'))|^2\,d\xi'\right)^{1/2} \le A_p\|f\|_{L^p(\mathbf{R}^n)}$$

for $p = (2n+2)/(n+3)$, $n \ge 2$.

(b) The cone $S = \{\xi \in \mathbf{R}^n : Q(\xi) = 0\}$; Q is a nondegenerate indefinite quadratic form on $\mathbf{R}^n$. Near any point $\xi \neq 0$, the measure $d\mu$ on S is defined by $d\mu = \left|\frac{\partial Q}{\partial \xi_n}\right|^{-1} \cdot d\xi'$ where $\xi = (\xi', \xi_n)$ is a splitting of the coordinates chosen to give $\frac{\partial Q}{\partial \xi_n} \neq 0$.† The result is that

$$\left(\int_S |\widehat{f}(\xi)|^2 \, d\mu(\xi)\right)^{1/2} \leq A_p \|f\|_{L^p(\mathbf{R}^n)}$$

for $p = \frac{2n}{n+2}$, $n \geq 3$.

(c) The "hypersphere" $S = \{\xi \in \mathbf{R}^n : Q(\xi) = 1\}$; here Q is a nondegenerate quadratic form with at least one positive eigenvalue. Again $d\mu = \left|\frac{\partial Q}{\partial \xi_n}\right|^{-1} \cdot d\xi'$. The inequality is

$$\left(\int_S |\widehat{f}(\xi)|^2 \, d\mu(\xi)\right)^{1/2} \leq A_p \|f\|_{L^p(\mathbf{R}^n)}$$

for $\frac{2n}{n+2} \leq p \leq \frac{2n+2}{n+3}$, and $n \geq 3$.

Strichartz [1977]. The proof given there follows the pattern described in §5.15, but uses formulae for the Fourier transforms (of the relevant surface carried measures) that involve Bessel functions. One can bypass these explicit calculations; the argument that follows can also be used in the generalizations described in §5.17 below.

To prove (a), we note that by §5.15(a), the inequality

$$\left(\int_{|\xi'| \leq 1} |f(\xi', Q'(\xi'))|^2 \, d\xi'\right)^{1/2} \leq A_p \|f\|_{L^p(\mathbf{R}^n)}$$

holds when $1 \leq p \leq (2n+2)/(n+3)$. We now replace $f(x)$ with $f(x/R)$ in the above and observe that, when $p = (2n+2)/(n+3)$, the inequality re-scales to read

$$\left(\int_{|\xi'| \leq R} |f(\xi', Q'(\xi'))|^2 \, d\xi'\right)^{1/2} \leq A_p \|f\|_{L^p(\mathbf{R}^n)}.$$

Letting $R \to \infty$ gives the desired result.

To prove (b), we notice that the cone S has $n-2$ nonvanishing principal curvatures at each $\xi \in S \backslash \{0\}$. Then it follows by §5.15(c) that

$$\left(\int_{S \cap \{1/2 \leq |\xi'| \leq 2\}} |\widehat{f}(\xi)|^2 \, d\mu(\xi)\right)^{1/2} \leq A_p \|f\|_{L^p(\mathbf{R}^n)}$$

† A related definition may be found in Chapter 11, §3.1.2.

and a re-scaling argument shows that

$$\left(\int_{S \cap \{2^{k-1} \leq |\xi'| \leq 2^{k+1}\}} |\widehat{f}(\xi)|^2 \, d\mu(\xi)\right)^{1/2} \leq A_p \|f\|_{L^p(\mathbf{R}^n)},$$

when $p = 2n/(n+2)$. Now use the Littlewood-Paley decomposition (as in Chapter 6, §7.14) $f = \sum f_k = \sum \Delta_k(f)$, acting on the $\mathbf{R}^{n-1}$ variable x' (writing $x = (x', x_n)$). Since $\widehat{f}_k(\xi)$ is supported where $2^{k-1} \leq |\xi'| \leq 2^{k+1}$, we can replace f with f_k in the above. After summing, we get

$$\int_S |\widehat{f}(\xi)|^2 \, d\mu(\xi) \leq A \sum_k \|f_k\|^2_{L^p(\mathbf{R}^n)}.$$

Now

$$\sum_k \|f_k\|^2_{L^p(\mathbf{R}^n)} \leq \|(\sum_k |f_k|^2)^{1/2}\|_{L^p(\mathbf{R}^n)} \leq A \|f\|^2_{L^p(\mathbf{R}^n)}.$$

The first inequality follows from Minkowski's inequality (since $p \leq 2$), and the second from Chapter 6, §7.14

The inequality in (c) for $p = (2n+2)/(n+3)$ can be shown to be a consequence of the result (b), if we apply (b) to $\mathbf{R}^{n+1}$. Similarly, the inequality for $p = 2n/(n+2)$ can be deduced from the weighted version (as in §5.17(b)), again in $n+1$ dimensions.

For the case of the cone (b) when $n = 3$, a sharp (L^p, L^q) restriction theorem (analogous to the result for the circle in Chapter 9, §5.5) has been proved by Taberner [1985].

5.17 The (L^p, L^2) restriction theorems in §5.16 extend to hypersurfaces given by more general homogeneous functions.

(a) Suppose ϕ is a homogeneous function on $\mathbf{R}^{n-1}$ of degree k that is smooth away from the origin. We assume that ϕ has rank r, in the sense that the matrix $\left(\frac{\partial^2 \phi}{\partial \xi_i \partial \xi_j}\right)(\xi')$ has rank at least r whenever $\xi' = (\xi_1, \ldots, \xi_{n-1}) \neq 0$. Suppose also that $r > 2(n-1)/k$. Then

$$\left(\int_{\mathbf{R}^{n-1}} |\widehat{f}(\xi', \phi(\xi'))|^2 \frac{d\xi'}{|\xi'|^{2\alpha}}\right)^{1/2} \leq A \|f\|_{L^p(\mathbf{R}^n)},$$

where $p = \frac{2n-2+2k}{n-1+2k+2\alpha}$ and $0 \leq \alpha < \frac{n-1}{2}$.

(b) Next let ϕ be a homogeneous function of degree 1 on $\mathbf{R}^{n-1}$ that is smooth away from the origin. Assume that ϕ has rank r (as in (a)) and that $r \geq 1$. Consider the cone $S = \{\xi \in \mathbf{R}^n : \xi_n = \phi(\xi')\}$, and let $d\mu(\xi) = d\xi'/|\xi'|$ on S. Then

$$\left(\int_S |\widehat{f}(\xi)|^2 \frac{d\mu(\xi)}{|\xi'|^{2\alpha}}\right)^{1/2} = \left(\int_{\mathbf{R}^{n-1}} |\widehat{f}(\xi', \phi(\xi'))|^2 \frac{d\xi'}{|\xi'|^{1+2\alpha}}\right)^{1/2} \leq A_p \|f\|_{L^p(\mathbf{R}^n)},$$

for $p = 2n/(n+2+2\alpha)$ and $\alpha_0 \leq \alpha < r/2$, where $\alpha_0 = 2n/(2r+4) - 1$.

The proofs are similar to the arguments outlined in §5.16 above. For (a) we use the estimate §5.13(b) for the Fourier transform (with n replaced by $n-1$) and §5.15. For (b) one uses the estimate §5.8, together with §5.15(c), and the observation that if ϕ has rank r, then the surface S has r nonvanishing principal curvatures at every point with $\xi' \neq 0$. Note that the condition $\alpha \geq \alpha_0$ guarantees that $p \leq (2r+4)/(r+4)$.

D. Applications

5.18 The restriction theorems have consequences for various estimates on solutions of partial differential equations.

(a) For $(x,t) \in \mathbf{R}^n \times \mathbf{R}$, consider the wave equation

$$\frac{\partial^2 u}{\partial t^2} = \Delta_x u, \qquad u(x,0) \equiv 0, \qquad \frac{\partial u}{\partial t}(x,0) = f(x).$$

Then we have the *a priori* estimate

$$\|u(x,t)\|_{L^q(\mathbf{R}^n \times \mathbf{R})} \leq A\|f\|_{L^2(\mathbf{R}^n)},$$

where $q = (2n+2)/(n-2)$, if $n \geq 3$.

A generalization is the estimate

$$\|u(x,t)\|_{L^q(\mathbf{R}^n \times \mathbf{R})} \leq A\|\nabla^{\alpha-1/2} f\|_{L^2(\mathbf{R}^n)},$$

where $q = (2n+2)/(n-1-2\alpha)$, $0 \leq \alpha < (n-1)/2$, and $n \geq 2$; here $\nabla^{\alpha-1/2}$ is the pseudo-differential operator whose symbol is $(2\pi|\xi|)^{\alpha-1/2}$.

(b) Consider the generalized "wave" equation

$$\frac{\partial u}{\partial t} = i\phi(D)u, \qquad u(x,0) = f(x).$$

Here ϕ is a real homogeneous function of degree 1 that is smooth away from the origin and $\phi(D)$ is the pseudo-differential operator having symbol $\phi(\xi)$. We assume that $\operatorname{rank}\left(\dfrac{\partial^2\phi}{\partial\xi_i\partial\xi_j}\right) \geq r$ whenever $\xi \neq 0$, for some fixed $r \geq 1$. Then

$$\|u(x,t)\|_{L^q(\mathbf{R}^n \times \mathbf{R})} \leq A\|\nabla^{\alpha+1/2} f\|_{L^2(\mathbf{R}^n)},$$

where $q = \dfrac{2n+2}{n-1-2\alpha}$ and $\alpha_0 \leq \alpha < \dfrac{n-1}{2}$; here $\alpha_0 = \dfrac{2n+2}{2r+4} - 1$. Note that when $\phi(\xi) = |\xi|$, we have $r = n-1$ and $\alpha_0 = 0$.

To prove (b), observe that

$$u(x,t) = \int_{\mathbf{R}^n} e^{2\pi i(t\phi(\xi)+x\cdot\xi)}\,\widehat{f}(\xi)\,d\xi = \int_S e^{2\pi i\bar{x}\cdot\bar{\xi}}\,\widehat{f}(\bar{\xi})\cdot|\xi'|\,d\mu(\bar{\xi}),$$

where $\bar{\xi} = (\xi, \xi_{n+1}) \in \mathbf{R}^{n+1}$, $\bar{x} = (x, t/2\pi) \in \mathbf{R}^{n+1}$, and $S = \{\bar{\xi} \in \mathbf{R}^{n+1} : \xi_{n+1} = \phi(\xi)\}$. Now apply §5.17(b) (replacing n with $n+1$) and dualize. Part (a) is a consequence of (b).

For (a), see Strichartz [1977], where further related results are also given. Other estimates for the wave equation are given below in §5.20 and §5.21.

5.19 The following is an additional application of the restriction theorems, along the lines of §5.18.

(a) For $(x,t) \in \mathbf{R}^n \times \mathbf{R}$, consider the Schrödinger equation

$$\frac{\partial u}{\partial t} = i\Delta_x u, \qquad u(x,0) = f(x).$$

Then we have the *a priori* estimate

$$\|u(x,t)\|_{L^q(\mathbf{R}^n \times \mathbf{R})} \leq A\|f\|_{L^2(\mathbf{R}^n)},$$

when $q = (2n+4)/n$ and $n \geq 1$.

(b) More generally, consider the pseudo-differential equation

$$\frac{\partial u}{\partial t} = i\phi(D)u, \qquad u(x,0) = f(x),$$

where ϕ is a real homogeneous function of degree k on $\mathbf{R}^n$ that is smooth away from the origin. We assume that $\operatorname{rank}\left(\dfrac{\partial\phi(\xi)}{\partial\xi_i\partial\xi_j}\right) \geq r$ whenever $\xi \neq 0$, with $r \geq 2n/k$ (or $r = n$, if ϕ is a quadratic polynomial). We then have the *a priori* estimate

$$\|u(x,t)\|_{L^q(\mathbf{R}^n \times \mathbf{R})} \leq A\|\nabla^\alpha f\|_{L^2(\mathbf{R}^n)} \leq A\|f\|_{L^2_\alpha(\mathbf{R}^n)},$$

if $q = \dfrac{2n+2k}{n-2\alpha}$ and $0 \leq \alpha < \frac{n}{2}$.

Indeed, in the situation (b), we have

$$u(x,t) = \int_{\mathbf{R}^n} e^{it\phi(\xi)} e^{2\pi i x\cdot\xi}\,\widehat{f}(\xi)\,d\xi.$$

We now dualize the inequality §5.17(a) (replacing $n-1$ with n), giving the desired result.

For conclusion (a), see Strichartz [1977]. Note the following special case of (b). We consider $\dfrac{\partial u}{\partial t} = \dfrac{\partial^3 u}{\partial x^3}$; here $n = 1$ and $k = 3$. With $\alpha = 0$, we get $\|u(x,t)\|_{L^8(\mathbf{R}^1 \times \mathbf{R})} \leq A\|f\|_{L^2(\mathbf{R})}$. For this equation, certain variants, and applications to nonlinear problems, see Kenig, Ponce, and Vega [1991], also Ginibre and Velo [1985], Balabane [1989].

5.20 Let Q be a nondegenerate quadratic form on $\mathbf{R}^n$. We write $Q(D)$ for the second-order partial differential operator given by $Q\left(\dfrac{1}{2\pi i}\dfrac{\partial}{\partial x_j}\right)$. The following estimates are closely related to the restriction theorem.

(a) We have the *a priori* inequality, for test functions u,

$$\|u\|_{L^q(\mathbf{R}^n)} \le A_z \|(Q+zI)u\|_{L^p(\mathbf{R}^n)}$$

where $\frac{1}{p}+\frac{1}{q}=1$, $\frac{2n}{n+2}\le p\le\frac{2n+2}{n+3}$, and $n\ge 3$; z is a nonzero complex number. The norm A_z remains bounded as long as $|z|\ge 1$. By rescaling, one sees that the inequality holds uniformly in z, for all $z\in\mathbf{C}$, if $p=2n/(n+2)$ (then $q=2n/(n-2)$).

(b) A variant of this is the inequality

$$\|e^{a\cdot x}\cdot u(x)\|_{L^q(\mathbf{R}^n)} \le A\|e^{a\cdot x}P(D)u\|_{L^p(\mathbf{R}^n)}.$$

Here $p=2n/(n+2)$; again q is the conjugate exponent to p and $n\ge 3$. The vector a is in $\mathbf{R}^n$ and $P(D)=Q(D)+R(D)$, where R is any first-order differential operator having constant coefficients. The crucial fact is that the bound A can be chosen to be independent of a and of $R(D)$.

(c) One has (in the notation of §5.16(c))

$$\|\textstyle\int_S \widehat{f}(\xi)e^{2\pi i x\cdot\xi}\,d\mu(\xi)\|_{L^q(\mathbf{R}^n,dx)} \le A\|f\|_{L^p(\mathbf{R}^n)}$$

where $\frac{1}{p}+\frac{1}{q}=1$ and $\frac{2n}{n+2}\le p\le\frac{2n+2}{n+3}$.

We note that conclusion (c), by virtue of the discussion in §5.14, is merely a restatement of the restriction theorem formulated in §5.15(c). We also remark that conclusion (a) implies (c) directly. Indeed, if $z=-1+iy$, then $(Q+\bar{z}I)^{-1}-(Q+zI)^{-1}$ corresponds to the multiplier $\frac{2y}{[Q(\xi)-1]^2+y^2}$. This tends to $c\,d\mu$ as $y\to 0$.

Further details are in Kenig, Ruiz, and Sogge [1987], where the Carleman-type estimates (b) are deduced and applied to uniqueness theorems. For some related results also dealing with uniqueness, see Jerison and Kenig [1985], Sogge [1989].

5.21 We consider the "averaging operators" defined with respect to a smooth hypersurface that has curvature. These operators may be viewed as, in a sense, dual to the restriction operators discussed above. Moreover, they have an interest in their own right because they have some nonobvious smoothing properties, which can be described as follows.

Let $S\subset\mathbf{R}^n$ be a smooth hypersurface, and let $d\mu=\psi\,d\sigma$ where $\psi\in C_0^\infty(\mathbf{R}^n)$ and $d\sigma$ is the induced Lebesgue measure on S. We consider the operator A given by $Af=f*d\mu$. The curvature hypothesis on S is contained in the assumption that $\widehat{d\mu}(\xi)=O(|\xi|^{-r})$, for some fixed $r>0$. One can then conclude:

(a) $A: L^p(\mathbf{R}^n)\to L^p_\gamma(\mathbf{R}^n)$ is bounded whenever $1<p<\infty$ and $\left|\frac{1}{p}-\frac{1}{2}\right|\le\frac{1}{2}\left(1-\frac{\gamma}{r}\right)$.

(b) $A: L^p(\mathbf{R}^n)\to L^{p'}(\mathbf{R}^n)$ whenever $1\le p\le 2$, $\frac{1}{p}-\frac{1}{2}\le\frac{1}{2}\left(\frac{r}{1+r}\right)$, and p' is the exponent conjugate to p.

We observe that when $S\subset\mathbf{R}^n$ is the unit sphere, then $r=(n-1)/2$. In the special case $n=3$ (with $\psi\equiv 1$ on S), we have the identity $A(f)=u(x,1)$, where u is the solution of the wave equation

$$\frac{\partial^2 u}{\partial t^2}=\Delta_x u,\qquad u(x,0)\equiv 0,\qquad \frac{\partial u}{\partial t}(x,0)=f(x). \tag{$*$}$$

The above results can be generalized by embedding the operator A in an analytic family A_α as follows. Let $\delta(x)$ be a smooth function that is comparable to the (signed) distance from S. For a suitable cut-off function $\tilde{\psi}$, we write

$$d\mu_\alpha=\frac{1}{\Gamma(\alpha)}\tilde{\psi}(x)|\delta(x)|^{\alpha-1}\,dx,\qquad \text{when } \mathrm{Re}(\alpha)>0,$$

and set $A_\alpha(f)=f*d\mu_\alpha$. We note that the function $\alpha\mapsto d\mu_\alpha$ has an analytic continuation to a distribution-valued function of α, for all $\alpha\in\mathbf{C}$, and that, in this continuation, we have $A_0=A$. We can then assert:

(a^*) $A_\alpha: L^p(\mathbf{R}^n)\to L^p_{\gamma+\alpha}(\mathbf{R}^n)$ is bounded whenever $\alpha\in\mathbf{R}$, $1<p<\infty$, and $\left|\frac{1}{p}-\frac{1}{2}\right|\le\frac{1}{2}\left(1-\frac{\gamma}{r}\right)$.

(b^*) $A_\alpha: L^p(\mathbf{R}^n)\to L^{p'}(\mathbf{R}^n)$ whenever $1<p\le 2$, $\frac{1}{p}-\frac{1}{2}\le\frac{1}{2}\left(\frac{r+\alpha}{1+r}\right)$, and $-r\le\alpha$; again p' is the exponent conjugate to p.

The interpretation of these results is as follows. First, one should note that A_α behaves much like $(-\Delta)^{-\alpha/2}\circ A$. Second, in the special case when $S\subset\mathbf{R}^n$ is the unit sphere, $\delta(x)=1-|x|^2$, and $\alpha=(3-n)/2$, then $(A_\alpha f)(x)$ represents essentially the solution $u(x,t)$, at time $t=1$, of the wave equation ($*$). As a consequence, we can state the following estimates for this solution.

(c) $\|u(x,1)\|_{L^p_\beta(\mathbf{R}^n)}\le A\|f\|_{L^p(\mathbf{R}^n)}$ whenever $1\le p\le\infty$ and $\left|\frac{1}{p}-\frac{1}{2}\right|\le\frac{1-\beta}{n-1}$.

(d) $\|u(x,1)\|_{L^{p'}(\mathbf{R}^n)}\le A\|f\|_{L^p(\mathbf{R}^n)}$ with $p=\frac{2n+2}{n+3}$ and $p'=\frac{2n+2}{n-1}$.

To prove these results, one proceeds as in Chapter 4, §5 and Chapter 9, §1.2.3. To begin with, one shows that

$$\widehat{d\mu}(\xi)=O(|\xi|^{-r-\mathrm{Re}(\alpha)}),\qquad \text{provided } \mathrm{Re}(\alpha)\ge -r.$$

Then one verifies that

$$\begin{aligned}
A_\alpha &: L^\infty\to \mathrm{BMO} && \text{if } \mathrm{Re}(\alpha)=0,\\
A_\alpha &: L^2\to L^2 && \text{if } \mathrm{Re}(\alpha)=-r, \text{ and}\\
A_\alpha &: L^1\to L^\infty && \text{if } \mathrm{Re}(\alpha)=1.
\end{aligned}$$

An interpolation then gives the desired result.

For (a), (a^*), and (c), see Stein [1971], C. Fefferman and Stein [1972], Peral [1980], M. Beals [1982], and Miyachi [1980b]. For (b), (b^*), and (d), see Strichartz [1970a], Littman [1973].

We mention three related series of results.

(i) The Fourier integral operator formulations of these results, which are applicable to variable coefficient hyperbolic equations, are treated in Chapter 9, §3.14 and §6.16.

(ii) Maximal operators for these averages are treated in Chapter 11, §3.

(iii) Other estimates for the wave equation (and the closely related Klein-Gordon equation), having some connection with the above, may be found in, e.g., Marshall, Strauss, and Wainger [1980], Marshall [1981], Pecher [1985].

5.22 In §3.2, we saw that if $d\mu = \psi\, d\sigma$ is a measure with smooth density ψ, carried on a submanifold $S \subset \mathbf{R}^n$ of finite type, then $\widehat{d\mu}(\xi) = O(|\xi|^{-\varepsilon})$ as $|\xi| \to \infty$, for some $\varepsilon > 0$. It turns out that if the density ψ is merely in $L^p(d\sigma)$, for some $p > 1$, then there is still an average decrease of $\widehat{d\mu}$ at infinity along any ray emanating from the origin. More precisely, suppose $\psi \in L^p(d\sigma)$ has compact support. Then

$$R^{-1} \int_0^R |\widehat{d\mu}(\rho\xi)|^2\, d\rho \le A(R|\xi|)^{-\varepsilon}, \qquad (*)$$

where $\varepsilon < k^{-1}(1-p^{-1})$ and k is the type of S.

This result has the following consequences.

(a) Let $\Omega(x)\cdot|x|^{-n}$ be a homogeneous function of degree $-n$, with $\Omega \in L^p(S^{n-1})$, for some $p > 1$, and $\int_{S^{n-1}} \Omega(x)\, d\sigma(x) = 0$. Let $r \mapsto b(r)$ be a bounded function on $(0,\infty)$. We consider the distribution

$$K = \text{p.v. } b(|x|)\cdot\Omega(x)\cdot|x|^{-n}.$$

Under the restriction that $n > 1$, we can conclude that the mapping $f \mapsto f * K$ extends to a bounded operator on $L^2(\mathbf{R}^n)$.

(b) The operator also extends to a bounded mapping of $L^p(\mathbf{R}^n)$ to itself, for $1 < p < \infty$, even if b satisfies the weaker condition $R^{-1}\int_0^R |b(\rho)|^2\, d\rho \le A$, for all $R > 0$.

To prove (a), write

$$\begin{aligned}\widehat{K}(\xi) &= \int_0^\infty \int_{S^{n-1}} e^{-2\pi i \xi\cdot\rho x}\, b(r)\,\Omega(x)\, d\sigma(x) \frac{d\rho}{\rho} \\ &= \int_0^\infty \widehat{d\mu}(\rho\xi)\, b(\rho) \frac{d\rho}{\rho} = \int_0^{|\xi|^{-1}} + \int_{|\xi|^{-1}}^\infty,\end{aligned}$$

where $\widehat{d\mu}(\xi) = \int_{S^{n-1}} e^{-2\pi i\xi\cdot x}\,\Omega(x)\, d\sigma(x)$. To estimate $\int_0^{|\xi|^{-1}}$, use the fact that $|\widehat{d\mu}(\rho\xi)| \le A\rho|\xi|$, which is a consequence of the cancellation condition on Ω. To estimate $\int_{|\xi|^{-1}}^\infty$, we observe that $(*)$ implies

$$\int_R^\infty |\widehat{d\mu}(\rho\xi)|^2 \frac{d\rho}{\rho^{1-\delta}} \le AR^\delta (R|\xi|)^{-\varepsilon}, \qquad \text{if } 0 < \delta < \varepsilon,$$

which shows that $\widehat{K} \in L^\infty$, proving the L^2 assertion. This last inequality is proved as follows.

If $S = \{\phi(x)\}$ is the image of ϕ, we can write

$$\widehat{d\mu}(\xi) = \int_{\mathbf{R}^m} e^{-2\pi i\phi(x)\cdot\xi}\, \tilde\psi(x)\, dx,$$

where $\tilde\psi$ has compact support and is in $L^p(\mathbf{R}^n)$, for some $p > 1$. Then the left side of $(*)$ is majorized by

$$R^{-1} \int \eta(\rho/R) |\widehat{d\mu}(\rho\xi)|^2\, d\rho,$$

where $\eta \in C_0^\infty(\mathbf{R})$ is a suitable nonnegative function. Now the last integral equals

$$\int_{\mathbf{R}^m\times\mathbf{R}^m} \widehat{\eta}(R\xi\cdot[\phi(x)-\phi(y)])\, \tilde\psi(x)\, \tilde\psi(y)\, dx\, dy.$$

Since $\widehat{\eta}(u) = O(|u|^{-\varepsilon})$ as $u \to \infty$, and $\tilde\psi \in L^p(\mathbf{R}^n)$, the integral is dominated by

$$A(R|\xi|)^{-\varepsilon} \left(\int_{K\times K} |\bar\xi\cdot[\phi(x)-\phi(y)]|^{-\varepsilon p'}\, dx\, dy \right)^{1/p'}.$$

Here $K \subset \mathbf{R}^m$ is a compact set containing $\operatorname{supp}(\tilde\psi)$, and $\bar\xi = \xi/|\xi|$ is a unit vector. Finally, one observes that, since ϕ has type k,

$$\sup_{y\in K} \sup_{\bar\xi} \int_K |\bar\xi\cdot[\phi(x)-\phi(y)]|^{-\varepsilon p'}\, dx \le A < \infty,$$

whenever $\varepsilon p' < 1/k$.

Results of this kind are due to R. Fefferman [1979], Namazi [1984], Duoandikoetxea and Rubio de Francia [1986], Chen [1987].

5.23 Suppose $K(x)$ is a Calderón-Zygmund kernel on $\mathbf{R}^n$ of the following kind: K is homogeneous of degree $-n$, smooth on the unit sphere, and $\int_{|x|=1} K(x)\, d\sigma(x) = 0$. For each ε and N, define the truncated kernel $K_{\varepsilon,N}$ by $K_{\varepsilon,N} = K(x)$ if $\varepsilon < |x| < N$, with $K_{\varepsilon,N}(x) = 0$ otherwise. Let $P(x)$ denote any real-valued polynomial on $\mathbf{R}^n$ of degree d. Then:

(a) $\left|\int_{\mathbf{R}^n} K_{\varepsilon,N}(x)\, e^{iP(x)}\, dx\right| \le A_d$, where the bound A_d is independent of ε, N, and the coefficients of P.

(b) The discrete analogue also holds:

$$\left| \sum_{m\in\mathbf{Z}^n} K_{\varepsilon,N}(m)\, e^{iP(m)} \right| \le A_d',$$

again with A_d' independent of ε, N, and the coefficients of P.

For results related to (a), see §6.4 in the next chapter.

Part (a) can be proved using the method of stationary phase. For this, see Stein and Wainger [1970], Stein [1986]. Although in appearance quite similar, part (b) requires different methods, in particular the techniques of exponential sums which occur in number theory. See Arkhipov and Oskolkov [1987], Stein and Wainger [1990].

Notes

§1 and §2. For the one-dimensional theory of oscillatory integrals (in §1) the reader may consult Titchmarsh [1951], Erdelyi [1956], Copson [1965]. The asymptotics in §1.4.2 go back to Riemann [1854]; see also Hardy [1913], Wainger [1965]. An important technique used by some of these authors, which we have not discussed, is the method of "steepest descent".

§3. Theorem 1, in a more precise form, is due to Hlawka [1950]; see also Herz [1962a]. Theorem 2, in the case where S is real-analytic, goes back to Björk [1973].

§4. The idea of restriction theorems for the Fourier transform is due to the author (1967), but was not published at that time. The proof of Theorem 3 is an elaboration of the original argument; another variant is cited in C. Fefferman [1970]. The proof of the fractional integration theorem given in §4.2 is taken from Hedberg [1972].

CHAPTER IX

Oscillatory Integrals of the Second Kind

We shall next take up oscillatory integrals of the second kind. While these arise in a variety of forms and have many different uses, we will concentrate here on only two broad classes of such operators. The first group is, in a sense, more directly derived from the Fourier transform: it contains the restriction operators which already appeared in the previous chapter, as well as the closely connected operators of Bochner-Riesz summability. The other class of operators had its genesis both in hyperbolic equations and in the integration over certain curved submanifolds; these go by the name of "Fourier integral" operators.

Turning to the first group, our focus will be on the operator T_λ (and its adjoint), with T_λ mapping functions on $\mathbf{R}^{n-1}$ to functions on $\mathbf{R}^n$, given by

$$(T_\lambda f)(\xi) = \int_{\mathbf{R}^{n-1}} e^{i\lambda\Phi(x,\xi)}\,\psi(x,\xi)\,f(x)\,dx, \qquad \xi \in \mathbf{R}^n.$$

Here $\psi \in C_0^\infty(\mathbf{R}^{n-1}\times\mathbf{R}^n)$ is a cut-off function, and the phase Φ satisfies a suitable nondegeneracy condition.

One important special case arises when

$$\Phi(x,\xi) = x\cdot\xi' + \phi(x)\xi_n, \quad x \in \mathbf{R}^{n-1}, \quad \xi = (\xi', \xi_n),\ \xi' \in \mathbf{R}^{n-1};$$

here the adjoint T_λ^* is essentially the operator given by restriction of the Fourier transform to the surface $x_n = \phi(x)$ in $\mathbf{R}^n$. The assumption on Φ is in this case equivalent to the nonvanishing of the Gaussian curvature on this hypersurface. Obtaining appropriate bounds for T_λ as $\lambda \to \infty$ has the effect of removing the cut-off in the ξ-variable.

The key inequality that is proved is an (L^p, L^2) bound for T_λ^*. This generalizes the restriction theorem for the Fourier transform

$$\left(\int_{\mathbf{S}^{n-1}} |\widehat{f}(\xi)|^2\,d\sigma(\xi)\right)^{1/2} \le A_p\|f\|_{L^p},$$

and holds exactly in the range $1 \le p \le \dfrac{2n+2}{n+3}$. There is a corresponding result for Bochner-Riesz summability (now an (L^p, L^p) inequality) in that range of p, which comes about by expressing the relevant operator as a sum of such T_λ^*, with $\lambda = 2^k$, $k = 1, 2, \ldots$. We mention that the full range one expects for the summability operators is $1 \le p < \frac{2n}{n+1}$, together with its dual range. This is known only when $n = 2$ and is made possible by the special play of exponents valid in that case.

The second class of operators we will study are of the form

$$(Tf)(x) = \int_{\mathbf{R}^n} e^{2\pi i \Phi(x,\xi)}\, a(x,\xi)\, \widehat{f}(\xi)\, d\xi, \qquad x \in \mathbf{R}^n,$$

where the phase Φ is homogeneous of degree 1 in ξ and satisfies a nondegeneracy condition, and a is a symbol in the class S^m studied in Chapter 6. These are the Fourier integral operators and our concern will be with their L^p boundedness. Note that the factor λ is here subsumed in Φ.

It turns out that the treatment of such operators requires two complementary analyses; the proofs are consequently quite elaborate. The first is in terms of their kernels K, which are given by

$$K(x,y) = \int_{\mathbf{R}^n} e^{2\pi i[\Phi(x,\xi) - y\cdot\xi]}\, a(x,\xi)\, d\xi.$$

In this connection, we should observe that for fixed x, the kernel $K(x,y)$ is singular when the phase $\xi \mapsto \Phi(x,\xi) - y\cdot\xi$ has a critical point, that is, when

$$y \in \Sigma_x = \{y : y = \nabla_\xi \Phi(x,\xi) \text{ for some } \xi\},$$

and the dimension of this set can be as large as $n-1$. The second way of looking at the operator comes from decomposing the spectrum (the ξ-space). Recall the Littlewood-Paley decomposition (roughly into the shells $2^j \le |\xi| \le 2^{j+1}$), which was important in our treatment of pseudo-differential operators in Chapter 6. In the present situation this has to be refined and a further dyadic division is needed: each dyadic shell $2^j \le |\xi| \le 2^{j+1}$, $j \ge 0$, is (essentially) split into $2^{j(n-1)/2}$ thin sectors Γ_j^ν, corresponding to truncated cones of aperture $\approx 2^{-j/2}$ (see Figure 1). It is only when both points of view are exploited that we can prove our main result (Theorem 2).

The two kinds of oscillatory integrals we have discussed have a common link, which itself is a modification of the Fourier transform. We begin by discussing this variant.

1. *Oscillatory integrals related to the Fourier transform*

We shall be considering two versions of operators of this class. The first, which is a mapping of functions on $\mathbf{R}^n$ to functions on $\mathbf{R}^n$, may be thought of as a generalization of the Fourier transform. The second,

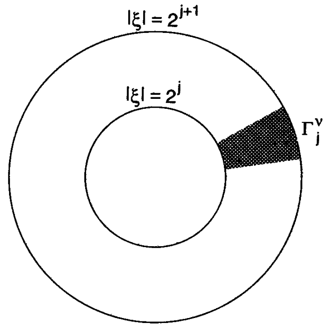

Figure 1. Second dyadic decomposition.

which is a mapping from functions on $\mathbf{R}^{n-1}$ to functions on $\mathbf{R}^n$, is a generalization of the operator R^* that arose in §4.1 of the previous chapter. Further results about restriction theorems, as well as some applications to Bochner-Riesz summability, will follow as a consequence.

1.1 The L^2 boundedness property of the Fourier transform will be subsumed in the properties of a family of operators T_λ, depending on a positive parameter λ, defined by

$$(T_\lambda f)(\xi) = \int_{\mathbf{R}^n} e^{i\lambda\Phi(x,\xi)} \psi(x,\xi)\, f(x)\, dx. \tag{1}$$

Here ψ is a fixed smooth function of compact support in x and ξ (a "cut-off" function); the phase Φ is real-valued and smooth. We assume that, on the support of ψ, the Hessian of Φ is nonvanishing, i.e.,

$$\det\left(\frac{\partial^2 \Phi(x,\xi)}{\partial x_i \partial \xi_j}\right) \ne 0. \tag{2}$$

PROPOSITION. *Under the above assumptions on Φ and ψ, we have that*

$$\|T_\lambda(f)\|_{L^2(\mathbf{R}^n)} \le A\lambda^{-n/2} \|f\|_{L^2(\mathbf{R}^n)}. \tag{3}$$

1.1.1 Before coming to the proof of (3), we make some remarks. First, the boundedness of T_λ for any fixed λ is trivial; what is of interest here is the indicated decrease in the norm as $\lambda \to \infty$. Second, the special case when $\Phi(x,\xi)$ is bilinear (and nondegenerate) corresponds essentially to the Fourier transform. Indeed, notice that with $\Phi(x,\xi) = -2\pi x \cdot \xi$, we get (after rescaling) that the family of operators

$$(\tilde{T}_\lambda f)(\xi) = \int_{\mathbf{R}^n} e^{-2\pi i x\cdot\xi}\, \psi(x/\lambda^{1/2}, \xi/\lambda^{1/2})\, f(x)\, dx$$

is uniformly bounded on $L^2(\mathbf{R}^n)$ as $\lambda \to \infty$; this reduces to the Fourier transform if we choose ψ so that $\psi(0,0) = 1$.

1.1.2 We begin the proof of the proposition by taking the precaution that the diameter of the support of the cut-off function ψ be sufficiently small; the exact requirement will become clear later in the argument. This restriction can be realized by using a partition of unity and writing T_λ as a finite sum of operators with small ψ-support.

To prove (3), it suffices to show that the operator norm of $T_\lambda T_\lambda^*$ is bounded by $A\lambda^{-n}$. We may write

$$(T_\lambda T_\lambda^* f)(\xi) = \int K_\lambda(\xi,\eta)\, f(\eta)\, d\eta,$$

where

$$K_\lambda(\xi,\eta) = \int_{\mathbf{R}^n} e^{i\lambda[\Phi(x,\xi)-\Phi(x,\eta)]}\psi(x,\xi)\, \bar{\psi}(x,\eta)\, dx; \tag{4}$$

an estimate for K_λ will give the required bound.

Let $M(x,\xi)$ be the matrix $\left(\dfrac{\partial^2\Phi}{\partial x_i \partial\xi_j}\right)$ and, for a vector $a \in \mathbf{R}^n$, let $\nabla_x^a = \langle a, \nabla_x\rangle$ denote differentiation in the direction a. Temporarily fix (ξ,η) and write

$$\Delta = \Delta(x,\xi,\eta) = \nabla_x^{a(x)}[\Phi(x,\xi) - \Phi(x,\eta)], \quad a(x) \in \mathbf{R}^n;$$

note that $\Delta = \langle M(x,\xi)[a(x)], \xi - \eta\rangle + O(|\xi-\eta|^2)$. Since M is invertible (by hypothesis), we may choose

$$a(x) = a(x,\xi,\eta) = M(x,\xi)^{-1}\left(\frac{\xi-\eta}{|\xi-\eta|}\right);$$

this gives $\langle M(x,\xi)[a(x)], \xi-\eta\rangle = |\xi-\eta|$. If we take $\operatorname{supp}\psi$ to be sufficiently small, we get that

$$|\Delta(x,\xi,\eta)| \geq c|\xi-\eta| \quad \text{for } (\xi,\eta) \in \operatorname{supp} K_\lambda;$$

note also that Δ is smooth (as are a and M).

We now set $D_x = [i\lambda\Delta(x,\xi,\eta)]^{-1}\nabla_x^{a(x)}$. Then since

$$(D_x)^N\left(e^{i\lambda[\Phi(x,\xi)-\Phi(x,\eta)]}\right) = e^{i\lambda[\Phi(x,\xi)-\Phi(x,\eta)]},$$

we can integrate (4) by parts N times,† and obtain

$$K_\lambda(\xi,\eta) = \int_{\mathbf{R}^n} e^{i\lambda[\Phi(x,\xi)-\Phi(x,\eta)]}\, ({}^tD_x)^N[\psi(x,\xi)\bar{\psi}(x,\eta)]\, dx.$$

Thus

$$|K_\lambda(\xi,\eta)| \leq A_N(1+\lambda|\xi-\eta|)^{-N}, \qquad N \geq 0.$$

It follows from this (with $N > n$) and the lemma in Chapter 7, §2.4.1 that the operator $T_\lambda T_\lambda^*$ (which has kernel K_λ) has norm bounded by

$$A' \int_{\mathbf{R}^n} \frac{d\xi}{(1+\lambda|\xi|)^N} = A\lambda^{-n},$$

and the proposition is proved.

1.2 Our next main concern is the oscillatory integral

$$(T_\lambda f)(\xi) = \int_{\mathbf{R}^{n-1}} e^{i\lambda\Phi(x,\xi)}\psi(x,\xi)\, f(x)\, dx, \tag{5}$$

mapping functions on $\mathbf{R}^{n-1}$ to functions on $\mathbf{R}^n$. We simultaneously consider the dual operator

$$(T_\lambda^* f)(x) = \int_{\mathbf{R}^n} e^{-i\lambda\Phi(x,\xi)}\, \bar{\psi}(x,\xi)\, f(\xi)\, d\xi, \tag{6}$$

which maps functions on $\mathbf{R}^n$ to functions on $\mathbf{R}^{n-1}$.

Here ψ is a smooth cut-off function with compact support in both variables. On the phase function, we impose the analogue of the nondegeneracy condition (2) dictated by experience, keeping in mind that the variable x ranges over a space of one lower dimension than does the variable ξ.

We require first that for each (x^0,ξ^0) in the support of ψ, the bilinear form $B(u,v)$ on $\mathbf{R}^{n-1}\times\mathbf{R}^n$, defined by

$$B(u,v) = \langle v, \nabla_x\rangle\langle u, \nabla_\xi\rangle\Phi(x,\xi)\big|_{(x^0,\xi^0)},$$

have maximal rank $n-1$.

† Similar arguments (with more detail) may be found in Chapter 8, §1.

As a result, there exists a (unique up to sign) vector $\bar{u} \in \mathbf{R}^n$, $|\bar{u}| = 1$, so that the scalar function

$$x \mapsto \langle \bar{u}, \nabla_\xi \Phi(x, \xi^0) \rangle$$

has a critical point at $x = x^0$. Our further assumption is that this critical point is nondegenerate; i.e., we suppose that the associated $(n-1) \times (n-1)$ quadratic form is nonsingular:

$$\det \left\{ \frac{\partial^2}{\partial x_i \partial x_j} \langle \bar{u}, \nabla_\xi \Phi(x, \xi^0) \rangle \right\} \neq 0 \tag{7}$$

at $x = x^0$.

THEOREM 1. *Under the above assumptions on Φ, the operator* (5) *satisfies the estimate*

$$\|T_\lambda f\|_{L^q(\mathbf{R}^n)} \leq A \lambda^{-n/q} \|f\|_{L^p(\mathbf{R}^{n-1})}, \tag{8}$$

where

$$q = \left(\frac{n+1}{n-1} \right) p' \quad \text{and} \quad 1 \leq p \leq 2;$$

here $p' = (1 - p^{-1})^{-1}$ is the exponent conjugate to p.

Remarks. This theorem is rather intricate and only later will its thrust be clear. For now we limit ourselves to two comments. First, the relation $q = p'(n+1)/(n-1)$ between the exponents and the order of decay $\lambda^{-n/q}$ are best possible, as we shall see in §2.1.1 below. Second, it can be shown that the range $1 \leq p \leq 2$ in Theorem 1 is best possible when the dimension is greater than 2 (see §6.5); however, as is shown in the appendix, when $n = 2$, the result extends to $1 \leq p \leq 4$.

We shall later use an equivalent (dual) version of the result (8):

$$\|T_\lambda^*(f)\|_{L^q(\mathbf{R}^{n-1})} \leq A \lambda^{-n/p'} \|f\|_{L^p(\mathbf{R}^n)}, \qquad q = \left(\frac{n-1}{n+1} \right) p', \tag{8'}$$

for $1 \leq p \leq (2n+2)/(n+3)$; note that $(2n+2)/(n+3)$ is the dual exponent to $2(n+1)/(n-1)$.

1.2.1 The crux of the proof is the result for $p = 2$. By duality, the case $p = 2$ is equivalent to

$$\|T_\lambda^*(f)\|_{L^2(\mathbf{R}^{n-1})} \leq A \lambda^{-n/r'} \|f\|_{L^r(\mathbf{R}^n)}, \quad r' = \frac{2n+2}{n-1}, \ r = \frac{2n+2}{n+3}.$$

Now $\|T_\lambda^*(f)\|^2_{L^2(\mathbf{R}^{n-1})}$ can be written as

$$\int_{\mathbf{R}^n} \int_{\mathbf{R}^n} K_\lambda(\xi, \eta)\, f(\xi)\, \bar{f}(\eta)\, d\xi\, d\eta,$$

with

$$K_\lambda(\xi, \eta) = \int_{\mathbf{R}^{n-1}} e^{i\lambda[\Phi(x,\eta) - \Phi(x,\xi)]} \psi(x, \eta) \bar{\psi}(x, \xi)\, dx. \tag{9}$$

Let U denote the operator with kernel $K_\lambda(\xi, \eta)$; another way of stating the above manipulations is that $U = T_\lambda T_\lambda^*$. Since $\|T_\lambda^* f\|_{L^2} = \langle Uf, f \rangle$, it will suffice to prove, in view of Hölder's inequality, that

$$\|U(f)\|_{L^{r'}(\mathbf{R}^n)} \leq A \lambda^{-2n/r'} \|f\|_{L^r(\mathbf{R}^n)}.$$

The main idea of the proof will be to imbed the operator U in an analytic family of operators $\{U^s\}$ in the strip $(1-n)/2 \leq \mathrm{Re}(s) \leq 1$, so that

$$\begin{cases} \|U^s(f)\|_{L^2(\mathbf{R}^n)} \leq A \lambda^{-n} \|f\|_{L^2(\mathbf{R}^n)}, & \mathrm{Re}(s) = 1 \\ \|U^s(f)\|_{L^\infty(\mathbf{R}^n)} \leq A \|f\|_{L^1(\mathbf{R}^n)}, & \mathrm{Re}(s) = (1-n)/2, \end{cases} \tag{10}$$

while $U^0 = U$.

The construction of the family U^s requires two preliminary steps.

1.2.2 We extend the phase function Φ, initially given on $\mathbf{R}^{n-1} \times \mathbf{R}^n$, to $\mathbf{R}^n \times \mathbf{R}^n$. We shall use the notation $\tilde{x} = (x, y)$, $x \in \mathbf{R}^{n-1}$, $y \in \mathbf{R}^1$, so that $\tilde{x} \in \mathbf{R}^n$. We define $\tilde{\Phi}$ on $\mathbf{R}^n \times \mathbf{R}^n$ by

$$\tilde{\Phi}(\tilde{x}, \xi) = \Phi(x, \xi) + y \Phi_0(\xi) \tag{11}$$

with Φ_0 chosen so that the $n \times n$ Hessian of $\tilde{\Phi}$ is nonvanishing; i.e.,

$$\det \left(\frac{\partial^2}{\partial \tilde{x}_i \partial \xi_j} \tilde{\Phi}(\tilde{x}, \xi) \right) \neq 0. \tag{12}$$

In fact, the matrix $[\partial^2 / \partial x_i \partial \xi_j \, \Phi(x, \xi)]$ already has rank $n-1$, so we only need choose $\Phi_0(\xi)$ so that

$$(\bar{u} \cdot \nabla_\xi) \Phi_0(\xi) \neq 0,$$

where $\bar{u}$ is as in the definition (7).

1.2.3 We fix a $\zeta \in C_0^\infty$ with $\zeta(y) = 1$ for $|y| \leq 1$ and consider the family $\{\alpha_s\}$ of distributions on $\mathbf{R}^1$ that arises by analytic continuation of the family $\{\alpha_s\}$ of functions, initially given when $\mathrm{Re}(s) > 0$ by

$$\alpha_s(y) = \begin{cases} \dfrac{e^{s^2}}{\Gamma(s)} y^{s-1} \zeta(y), & \text{if } y > 0; \\ 0, & \text{if } y \leq 0. \end{cases}$$

For us, the relevance of the factor $e^{s^2}/\Gamma(s)$ is that, first, it vanishes at $s = 0, -1, -2, \ldots$, and second, that it is rapidly decreasing in $s = \sigma + it$ as $|t| \to \infty$, whenever s lies in a fixed strip $a \leq \sigma \leq b$.

We assert, to begin with, that $s \mapsto \alpha_s$ has an analytic continuation (as a distribution-valued function) for all complex s. To see this, let $\phi \in \mathcal{S}$ and let N be a large integer. When $\mathrm{Re}(s) > 0$

$$\begin{aligned}\alpha_s(\phi) &= \int_{-\infty}^{\infty} \alpha_s(y)\,\phi(y)\,dy \\ &= (-1)^N \frac{e^{s^2}}{\Gamma(s)c_N(s)} \int_0^\infty y^{N+s-1}\left(\frac{d}{dy}\right)^N [\zeta(y)\phi(y)]\,dy,\end{aligned} \tag{13}$$

where $c_N(s) = (N+s-1)(N+s-2)\cdots(s)$; the identity (13) is established by an obvious N-fold integration by parts, since $(d/dy)^N[y^{N+s-1}] = c_N(s)y^{s-1}$. From this it is clear that $\alpha_s(\phi)$ continues to $\mathrm{Re}(s) > -N$, and hence to all of $\mathbf{C}$.

Next, observe that

$$\alpha_s(\phi) = \phi(0), \qquad \text{when } s = 0, \tag{14}$$

i.e., that α_0 is the Dirac delta function at the origin. In fact, taking $N = 1$ in (13), letting $s \to 0$, and recalling that $s\Gamma(s) \to 1$ as $s \to 0$, we see that

$$\alpha_0(\phi) = -\int_0^\infty \frac{d}{dy}[\phi(y)\zeta(y)]\,dy = \phi(0)\zeta(0) = \phi(0),$$

establishing (14).

The last fact we shall need about the α_s is the Fourier transform estimate

$$\left|\int \alpha_s(y)e^{iyu}\,dy\right| \le A_\sigma(1+|u|)^{-\sigma}, \quad s = \sigma + it, \quad \sigma \le 1, \tag{15}$$

where the integral on the left is defined by analytic continuation in s. Indeed, let $\tilde{\zeta}$ be a smooth cut-off function with compact support so that $\tilde{\zeta}(y) = 1$ on the support of ζ. Then with $\phi(y) = e^{iyu}\tilde{\zeta}(y)$, we have

$$\int \alpha_s(y)e^{iyu}\,dy = \int \alpha_s(y)\,\phi(y)\,dy.$$

Thus if $-N < \sigma \le -N+1$, the formula (13) easily reduces the proof of (15) to the case where $0 < \sigma \le 1$. Turning to that case, we note that when $|u|$ is not large, the estimate (15) is trivial, since $\alpha_s(y)$ is integrable. For large $|u|$, we write

$$\int \alpha_s(y)e^{iyu}\,dy = \int_0^{1/|u|} + \int_{1/|u|}^\infty .$$

The first integral is obviously $O(\int_0^{|u|^{-1}} y^{\sigma-1}\,dy) = O(|u|^{-\sigma})$. For the second integral, we invoke the corollary in Chapter 8, §1.2 (for $k = 1$). The result is then $\int_{|u|^{-1}}^\infty = O(|u|^{-1}\cdot|u|^{\sigma-1}) = O(|u|^{\sigma})$, so that (15) is proved.

1.2.4 We can now define the analytic family of operators U^s. We write

$$U^s f(\xi) = \int_{\mathbf{R}^n} K^s(\xi,\eta)\,f(\eta)\,d\eta,$$

where $K^s(\xi,\eta) = K^s_\lambda(\xi,\eta)$ is given by the formula

$$K^s(\xi,\eta) = \int_{\mathbf{R}^n} e^{i\lambda[\tilde{\Phi}(\tilde{x},\eta)-\tilde{\Phi}(\tilde{x},\xi)]}\psi(x,\eta)\,\bar{\psi}(x,\xi)\,\alpha_s(y)\,dx\,dy. \tag{16}$$

Recall that $\tilde{\Phi}$ is the extended phase function given by (11); also $\tilde{x} = (x,y) \in \mathbf{R}^n$, with $x \in \mathbf{R}^{n-1}$, $y \in \mathbf{R}^1$. The integral (16) is defined by analytic continuation in s, starting with $\mathrm{Re}(s) > 0$.

Observe that when $\mathrm{Re}(s) = 1$, the definition (16) is given in terms of absolute convergence, and the operator U^s with kernel K^s arises as the kernel of the composition of two operators $S_2 \circ S_1$, where

$$(S_1 f)(\tilde{x}) = \int_{\mathbf{R}^n} e^{i\lambda\tilde{\Phi}(\tilde{x},\eta)}\,\psi(x,\eta)\,\tilde{\zeta}(y)\,f(\eta)\,d\eta$$

and

$$(S_2 g)(\xi) = \int_{\mathbf{R}^n} e^{-i\lambda\tilde{\Phi}(\tilde{x},\xi)}\,\bar{\psi}(x,\xi)\,\zeta(y)\cdot|y|^{s-1}\cdot\frac{e^{s^2}}{\Gamma(s)}\,g(\tilde{x})\,d\tilde{x}.$$

Both fall under the scope of the proposition in §1.1, because

$$\det\left(\frac{\partial^2}{\partial\tilde{x}_i\partial\xi_j}\tilde{\Phi}(\tilde{x},\xi)\right) \ne 0$$

by (12), and the extra factor $|y|^{s-1}e^{s^2}/\Gamma(s)$, which appears in the second operator, is bounded. The result is the L^2 inequality stated in (10).

We next show that

$$|K^s(\xi,\eta)| \le A, \qquad \text{when } \mathrm{Re}(s) = (1-n)/2, \tag{17}$$

from which the (L^1, L^∞) inequality follows easily. Let us write $\hat{\alpha}_s(u)$ for the analytic continuation of

$$\int \alpha_s(y)\,e^{iyu}\,dy.$$

Then because of the formulas (16), (11), and (9), we see immediately that

$$K^s(\xi,\eta) = K_\lambda(\xi,\eta)\cdot\hat{\alpha}_s(\lambda[\Phi_0(\eta) - \Phi_0(\xi)]).$$

However, by (15), $|\hat{\alpha}_s(u)| \leq A(1+|u|)^{(n-1)/2}$ when $\mathrm{Re}(s) = (1-n)/2$, so to prove (17), it suffices to see that

$$|K_\lambda(\xi,\eta)| \leq A(1+\lambda|\xi-\eta|)^{(1-n)/2}. \tag{18}$$

In proving this estimate for K_λ, we may assume that the integrand in (9), and in particular the function ψ appearing there, is supported in a sufficiently small neighborhood of a point (x_0, ξ_0); otherwise the original operator T_λ can be realized as a finite sum of operators of this kind. Next, writing

$$\Psi(x,\xi,\eta) = \Phi(x,\eta) - \Phi(x,\xi),$$

we have that

$$\left(\frac{\partial}{\partial x}\right)^\alpha [\Psi(x,\xi,\eta) - \nabla_\xi \Phi(x,\xi)\cdot(\eta-\xi)] = O(|\eta-\xi|^2), \tag{19}$$

for all multi-indices α.

There are two possibilities that arise:

(i) The direction $(\eta-\xi)$ or its opposite $(\xi-\eta)$ is close to the critical direction $\bar{u}$ appearing in (7), or

(ii) Both of these directions are far from the critical direction $\bar{u}$.

In the first case, the $(n-1)\times(n-1)$ Hessian determinant of

$$x \mapsto \Psi(x,\xi,\eta)$$

exceeds $|\xi-\eta|^{n-1}$ in absolute value, because of (19) and (7). So (18) is a consequence of Proposition 6 in §2.3 of the previous chapter.‡

In the second case, we have that $|\nabla_x \Psi(x,\xi,\eta)| \geq c|\xi-\eta|$, because $\nabla_x\nabla_\xi\Phi(x,\xi)$ has rank $n-1$ and $(\eta-\xi)$ is far from the critical directions $\bar{u}$ and $-\bar{u}$. Then

$$K_\lambda(\xi,\eta) \text{ is rapidly decreasing, as } \lambda|\xi-\eta| \to \infty,$$

as Proposition 4 of Chapter 8, §2.1 shows. Therefore (18), and hence (17), is proved, giving the (L^1, L^∞) estimate in (10).

1.2.5 An interpolation lemma. The proof of the theorem will be concluded by making use of interpolation of operators. We formulate here a version of the theorem needed that is general enough for our applications.

‡ See (21) of Chapter 8 and the note that follows.

We suppose that we are given a family of operators $\{U^s\}$ on the strip $a \leq \mathrm{Re}(s) \leq b$, defined by

$$(U^s f)(x) = \int_{\mathbf{R}^n} k_s(x,y)\, f(y)\, dy,$$

where we assume that the kernels have fixed compact support and are uniformly bounded for $(x,y) \in \mathbf{R}^n \times \mathbf{R}^n$ and $a \leq \mathrm{Re}(s) \leq b$; these are *a priori* assumptions, in the sense that the size of the support and the exact bounds on the k_s do not enter into the conclusions. We also assume that for each (x,y), the function $k_s(x,y)$ is analytic in $a < \mathrm{Re}(s) < b$ and is continuous in the closure $a \leq \mathrm{Re}(s) \leq b$. The significant quantitative assumptions are:

$$\begin{cases} \|U^s(f)\|_{L^{q_0}} \leq M_0 \|f\|_{L^{p_0}}, & \text{when } \mathrm{Re}(s) = a, \\ \|U^s(f)\|_{L^{q_1}} \leq M_1 \|f\|_{L^{p_1}}, & \text{when } \mathrm{Re}(s) = b; \end{cases} \tag{20}$$

here (p_i, q_i) are two pairs of given exponents with $1 \leq p_i, q_i \leq \infty$.

LEMMA.† *Under the assumptions* (20), *we can conclude that*

$$\|U^{a(1-\theta)+b\theta}(f)\|_{L^q} \leq M_0^{1-\theta} M_1^\theta \|f\|_{L^p}, \tag{21}$$

where $0 \leq \theta \leq 1$, $1/q = (1-\theta)/q_0 + \theta/q_1$, *and* $1/p = (1-\theta)/p_0 + \theta/p_1$.

The proof of this lemma is in the same style as that of Theorem 4 in Chapter 4, §5, but simpler, since the function $f^\sharp$ does not intervene. Let q' be the dual to q; to prove (21), it then suffices to show that

$$\left|\int_{\mathbf{R}^n} U^{a(1-\theta)+b\theta}(f)\cdot g\, dx\right| \leq M_0^{1-\theta} M_1^\theta, \tag{22}$$

whenever f and g are simple functions with $\|f\|_{L^p} \leq 1$ and $\|g\|_{L^{q'}} \leq 1$. Now set

$$f_s = |f|^{\alpha(s)}\cdot\frac{f}{|f|}, \qquad \alpha(s) = p\left[\frac{1-s}{p_0} + \frac{s}{p_1}\right],$$

and

$$g_s = |g|^{\beta(s)}\cdot\frac{\bar{g}}{|g|}, \qquad \beta(s) = q'\left[\frac{1-s}{q_0'} + \frac{s}{q_1'}\right],$$

and let

$$I(s) = M_0^{s-1} M_1^{-s} \int_{\mathbf{R}^n} U^{a(1-s)+bs}(f_s)\, g_s\, dx.$$

As a result, $I(s)$ is analytic and bounded in the strip $0 < \mathrm{Re}(s) < 1$, continuous in its closure and satisfies $|I(s)| \leq 1$, when $\mathrm{Re}(s) = 0$ or $\mathrm{Re}(s) = 1$, because of (20). The three-lines lemma then implies that $|I(\theta)| \leq 1$, which is (22), proving (21).

† For a more detailed treatment, see *Fourier Analysis*, Chapter 5, §4.

With this in hand we can finish the proof of the theorem. First, applying it to the operator U^s via (10) gives that $U = U^0$ satisfies

$$\|U^0(f)\|_{L^{r'}(\mathbf{R}^n)} \le A\lambda^{-2n/r'}\|f\|_{L^r(\mathbf{R}^n)}. \tag{23}$$

Here θ is chosen by

$$0 = \left(\frac{1-n}{2}\right)\cdot(1-\theta) + 1\cdot\theta;$$

we take $(p_0, q_0) = (1, \infty)$, $(p_1, q_1) = (2, 2)$, $M_0 = A$, and $M_1 = A\lambda^{-n}$. In other words, we have obtained (23) when

$$\theta = \frac{n-1}{n+1}, \qquad r = \frac{2(n+1)}{n+3}, \qquad \text{and } r' = \frac{2(n+1)}{n-1}.$$

Since $T_\lambda T_\lambda^* = U$, we get (8) with $p = 2$, $q = 2(n+1)/(n-1)$. Finally, another application of the interpolation lemma (this time with $U^s = T_\lambda$ independent of s) and the trivial case $p = 1$, $q = \infty$ of (8) concludes the proof of the theorem.

2. Restriction theorems and Bochner-Riesz summability

2.1 Restriction theorems. We return to the restriction theorems formulated in §4 of the previous chapter and prove a sharp result when S is a hypersurface (i.e., a submanifold of dimension $n-1$) whose Gaussian curvature does not vanish anywhere. We shall show that whenever $1 \le p \le (2n+2)/(n+3)$, the restriction inequality (30) of Chapter 8 holds with $q = 2$. More precisely,

PROPOSITION. *Let $S \subset \mathbf{R}^n$ be a manifold of dimension $n-1$ whose Gaussian curvature is nowhere zero, and let S_0 be a compact subset of S. Then*

$$\left(\int_{S_0} |\widehat{f}(\xi)|^q\, d\sigma(\xi)\right)^{1/q} \le A(S_0)\cdot\|f\|_{L^p(\mathbf{R}^n)}, \qquad f \in \mathcal{S}, \tag{24}$$

whenever

$$1 \le p \le \frac{2n+2}{n+3} \quad \textit{and} \quad q = \left(\frac{n-1}{n+1}\right)p', \quad \textit{where} \quad \frac{1}{p} + \frac{1}{p'} = 1.$$

Note that if $p = (2n+2)/(n+3)$ then $q = 2$, while $p = 1$ gives $q = \infty$.

Proof. After localization to a small neighborhood of a fixed point in S_0 and an appropriate change of variables (moving that point to the origin), we may assume that near that point S is given as the graph

$$x_n = \phi(x_1, \ldots, x_{n-1}) = \phi(x')$$

with $\phi(0) = \nabla\phi(0) = 0$, while

$$\det_{1\le i,j\le n-1}\left[\frac{\partial^2\phi}{\partial x_i \partial x_j}\right](x') \ne 0$$

for $x' = (x_1, \ldots, x_{n-1})$ small (as in Chapter 8, §3.1). Since

$$d\sigma(x) = d\sigma(x', \phi(x')) = (1 + |\nabla\phi|^2)^{1/2}\, dx',$$

proving the inequality is then reduced to showing that

$$\left(\int |\widehat{f}(x', \phi(x'))|^q \tilde{\psi}(x')\, dx'\right)^{1/q} \le A\|f\|_p, \tag{25}$$

where $\tilde{\psi} \in C_0^\infty(\mathbf{R}^{n-1})$ is a suitable nonnegative cut-off function.

For this purpose consider

$$(T_\lambda^* f)(x') = \int_{\mathbf{R}^n} e^{-i\lambda\Phi(x',\xi)}\, \tilde{\psi}(x')\, \psi_0(\xi)\, f(\xi)\, d\xi,$$

where $\Phi(x', \xi) = 2\pi(x'\cdot\xi' + \phi(x')\xi_n)$. Then, as is easily seen, Φ satisfies the conditions formulated in §1.2; in particular, it satisfies (7) with $x_0 = 0$ and $\bar{u} = (0, \ldots, 0, 1)$. We choose $\psi_0 \in C_0^\infty(\mathbf{R}^n)$ with $\psi_0(0) = 1$. According to Theorem 1, we have that

$$\|T_\lambda^*(f)\|_{L^q(\mathbf{R}^{n-1})} \le A\lambda^{-n/p'}\|f\|_{L^p(\mathbf{R}^n)}, \quad \text{with} \quad q = \left(\frac{n-1}{n+1}\right)p';$$

in fact, this is exactly the dual inequality (8′). In this inequality for T_λ^* we re-scale the ξ variable: we replace f by f_λ, $f_\lambda(\xi) = f(\lambda\xi)$, and make the indicated change of variables, noting that $\|f_\lambda\|_{L^p} = \lambda^{-n/p}\|f\|_{L^p}$, and recalling that $\psi_0(\xi/\lambda) \to 1$ as $\lambda \to \infty$. We then get (25), completing the proof of the proposition.

2.1.1 Remarks. 1. The relation $q = [(n-1)/(n+1)]p'$ that appears in the proposition above is best possible. This can be easily seen in the case $n = 2$, taking S to be the unit circle (which is already typical of the general situation treated in the proposition). Note that in this case the inequality is actually equivalent to

$$\int_{1-\delta<|\xi|<1+\delta} |\widehat{f}(\xi)|^q\, d\xi \le c\delta\|f\|^q_{L^p(\mathbf{R}^n)} \tag{26}$$

for small positive δ. To test (26), observe that the shell $1-\delta < |\xi| < 1+\delta$ contains the rectangle

$$R_\delta = \{(\xi_1, \xi_2) : |\xi_1 - 1| < c\delta, |\xi_2| < c\delta^{1/2}\},$$

if c is a sufficiently small constant. Fix a $\widehat{g} \in \mathcal{S}(\mathbf{R}^1)$ with $\widehat{g}(\xi) \geq 1$ when $|\xi| \leq$ c, and define $\widehat{f}$ by

$$\widehat{f}(\xi_1, \xi_2) = \widehat{g}((\xi_1 - 1)/\delta)\, \widehat{g}(\xi_2/\delta^{1/2}).$$

Then of course

$$f(x_1, x_2) = g(\delta x_1)\, e^{-2\pi i x_1}\, g(\delta^{1/2} x_2) \cdot \delta^{3/2},$$

and so (26) becomes

$$\delta^{3/2} \leq c'\delta \cdot \delta^{(3/2)q} \cdot (\delta^{-3/2})^{q/p},$$

because

$$\|f\|_{L^p} = \left(\int |g(\delta x_1)\, g(\delta^{1/2} x_2)|^p\, dx_1\, dx_2\right)^{1/p} = c\delta^{-(3/2)(1/p)}.$$

Unscrambling the exponents leads to the inequality

$$\delta^{1/2(-3/2)q} \leq c\delta^{-(3/2)q/p}.$$

Since this is to hold for $\delta \to 0$, we can see that this requires

$$(3/2)q/p \geq (3/2)q - 1/2,$$

which means $q \leq p'/3$, as asserted. Notice that, as a consequence, the relation among the exponents in (8), $q = p'(n+1)/(n-1)$, is also best possible. Moreover, the scaling argument used in the proposition shows that the order of decay $\lambda^{-n/q}$ appearing in (8) cannot be improved.

2. It is natural to conjecture that the restriction theorem (and in particular the inequality (24)) extends to the range $1 \leq p < 2n/(n+1)$. We shall see below that this indeed holds in the case of two dimensions. Moreover, it is easy to show that no restriction theorem of any kind can hold for $f \in L^p(\mathbf{R}^n)$ when $p \geq 2n/(n+1)$; see §6.19.

2.2 Bochner-Riesz summability. A classical problem in Fourier analysis is to make precise the sense in which the Fourier inversion formula

$$f(x) = \int_{\mathbf{R}^n} \widehat{f}(\xi)\, e^{2\pi i x\cdot\xi}\, d\xi, \qquad \widehat{f}(\xi) = \int_{\mathbf{R}^n} f(x)\, e^{-2\pi i x\cdot\xi}\, dx$$

holds, say for $f \in L^p(\mathbf{R}^n)$. A natural way of formulating this identity is in terms of a "summability method"; for instance, one may assert that

$$f(x) = \lim_{R\to\infty} \int_{|\xi|<R} \widehat{f}(\xi) \left(1 - \frac{|\xi|^2}{R^2}\right)^\delta e^{2\pi i x\cdot\xi}\, d\xi \qquad (27)$$

for suitable δ. The following statements may help put the general question in perspective.

(i) In the case of one dimension, and for the "convergence" problem (i.e., when $\delta = 0$), (27) holds in the L^p norm, $1 < p < \infty$, by M. Riesz's theorem for the Hilbert transform; also (27) is valid almost everywhere for $f \in L^p$, $1 < p < \infty$, and the maximal "partial sum" operator is bounded on these L^p, by Carleson's theorem (in the version of Hunt and Sjölin). Moreover, when $\delta > 0$, the situation is much simpler because then the integral in (27) equals $f * K^\delta_{R^{-1}}(x)$, where $\{K^\delta_{R^{-1}}\}$ is an approximation to the identity satisfying the standard properties needed in Chapter 1, §6.1 and Chapter 2, §2.

(ii) Turning to the situation when $n \geq 2$, we shall here limit ourselves to the question of norm convergence. Matters then reduce to the boundedness in L^p of the multiplier operator

$$f(x) \mapsto (S^\delta f)(x) = \int_{|\xi|\leq 1} \widehat{f}(\xi)\, (1 - |\xi|^2)^\delta\, e^{2\pi i x\cdot\xi}\, d\xi. \qquad (28)$$

This operator can be written as a convolution

$$S^\delta(f) = f * K^\delta$$

with an explicitly given kernel K^δ. When $\delta > (n-1)/2$, we have $K^\delta \in L^1$; hence S^δ is bounded on all L^p, $1 \leq p \leq \infty$. What makes this problem difficult (and relevant to our previous considerations) is that when $\delta \leq (n-1)/2$, the kernel K^δ is no longer integrable and its characteristic properties are reflected in its oscillatory behavior. In the next chapter, we will see that what one can expect in $\mathbf{R}^n$, $n \geq 2$, has to be quite different from what happens in $\mathbf{R}^1$: The operator S^0 (which corresponds to convergence) is bounded on $L^p(\mathbf{R}^n)$ only when $p = 2$.

(iii) Passing to the case $\delta > 0$, it is not difficult to check (see §6.19 below) that the following conditions are necessary for L^p boundedness. First, if we make no further assumption on δ, then we must have

$$\frac{2n}{n+1} \leq p \leq \frac{2n}{n-1};$$

of course $2n/(n-1)$ and $2n/(n+1)$ are conjugate exponents. Second, and more specifically, if $\delta > 0$ and S^δ is bounded on $L^p(\mathbf{R}^n)$, then we must have

$$\frac{2n}{n+1+2\delta} < p < \frac{2n}{n-1-2\delta},$$

assuming also that $\delta < (n-1)/2$.

(iv) We state the positive results known at present. When $n = 2$, matters are completely settled: We shall see below that when $\delta > 0$, S^δ is bounded from $L^p(\mathbf{R}^2)$ to itself for $4/3 \leq p \leq 4$; there is also the companion result that it actually holds in the range $4/(3+2\delta) < p < 4/(1-2\delta)$, whenever $0 < \delta \le 1/2$. When $n \ge 3$, it is therefore reasonable to conjecture that the necessary conditions formulated in (iii) above are also sufficient. However, only the following partial result is known: S^δ is indeed bounded on L^p for the expected range of p's, but with the additional restriction in terms of the exponents r and r' that arose in the proof of Theorem 1 (see §1.2.1).

PROPOSITION. *The operator S^δ, initially defined for $f \in \mathcal{S}$ by* (28), *extends to a bounded operator from $L^p(\mathbf{R}^n)$ to itself whenever*

$$\frac{2n}{n+1+2\delta} < p < \frac{2n}{n-1-2\delta},$$

and

$$1 \le p \le \frac{2(n+1)}{n+3} \quad \textit{or} \quad \frac{2(n+1)}{n-1} \le p \le \infty.$$

Note. The first restriction is equivalent to

$$\delta > \delta(p), \qquad \text{where } \delta(p) = n\left|\frac{1}{p} - \frac{1}{2}\right| - \frac{1}{2};$$

this is the only known necessary condition for the boundedness of S^δ. As mentioned above, when $n = 2$ this condition is in fact sufficient (i.e., we may drop the second assumption on p); a proof may be found in the appendix below. For $n \ge 3$, the second restriction on p has recently been relaxed; see §6.9[a] below. Except for this last extension, our assertions are summarized in Figures 2 and 3; solid lines, and the regions below them, represent known unboundedness (except at $p = 2$, $\delta = 0$), the shaded area above the lines represents known boundedness.[‡]

2.2.1 We use the following lemma in the proof of the proposition.

LEMMA. *With S^δ defined by* (28), *we have that*

$$(S^\delta f)(x) = (f * K^\delta)(x)$$

for $f \in \mathcal{S}$, where

$$K^\delta(x) = \pi^{-\delta}\,\Gamma(1+\delta)\,|x|^{-(n/2)-\delta}\,J_{n/2+\delta}(2\pi|x|),$$

where the Bessel function $J_{n/2+\delta}$ is given by (16) *in Chapter 8.*

[‡] In Figure 3, the boundedness given between $p = 2(n+1)/(n+3)$ and $p = 2(n+1)/(n-1)$ follows from analytic Riesz interpolation between S^δ acting on L^p for those p (with any $\delta > \delta(p)$) and S^0 acting on L^2; see, e.g., the lemma in §1.2.5 or *Fourier Analysis*, Chapter 5, §4.

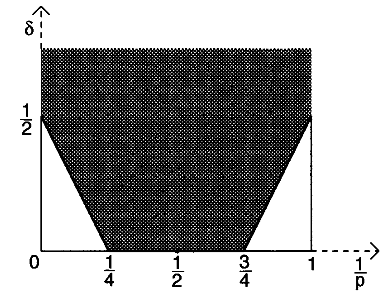

Figure 2. Boundedness of S^δ when $n = 2$.

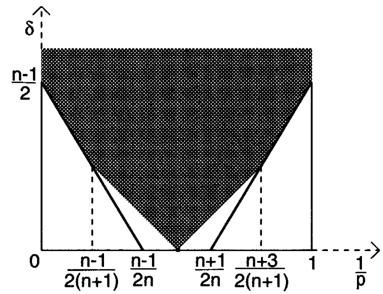

Figure 3. Boundedness of S^δ when $n \ge 3$.

For the identity establishing this lemma, see §6.19.

Note that the lemma, together with the asymptotic formula (15) of Chapter 8, implies that K^δ is bounded, and that

$$K^\delta(x) \sim |x|^{-(n+1)/2-\delta}\left[e^{2\pi i|x|}\sum_{j=0}^{\infty}\alpha_j|x|^{-j} + e^{-2\pi i|x|}\sum_{j=0}^{\infty}\beta_j|x|^{-j}\right] \tag{29}$$

as $|x| \to \infty$, for suitable constants α_j and β_j.

2.2.2 Because of the nature of the kernel K^δ given by the above asymptotic expansion, the study of the operator S^δ will be reduced to the study of an oscillatory operator G_λ, defined as follows. Let ψ be a fixed smooth function of compact support on $\mathbf{R}^n$ that vanishes in a neighborhood of the origin, and set

$$(G_\lambda f)(x) = \int_{\mathbf{R}^n} e^{i\lambda|x-y|}\,\psi(x-y)\,f(y)\,dy. \tag{30}$$

The principal lemma is then

LEMMA. *We have that*

$$\|G_\lambda(f)\|_{L^p} \le A\lambda^{-n/p'}\|f\|_{L^p}, \tag{31}$$

whenever $1 \le p \le (2n+2)/(n+3)$.

To prove (31), we modify G_λ, and consider

$$(\tilde{G}_\lambda f)(x) = \int_{\mathbf{R}^n} e^{i\lambda|x-y|}\,\tilde{\psi}(x,y)\,f(y)\,dy,$$

where now $\tilde{\psi}$ is a smooth cut-off function with compact support for $(x,y) \in \mathbf{R}^n \times \mathbf{R}^n$, whose support again does not intersect the diagonal $\{(x,y) : x = y\}$.

We next split the variable x as $x = (x', x_n)$, with

$$x' = (x_1, \dots, x_{n-1}) \in \mathbf{R}^{n-1}, \qquad x_n \in \mathbf{R}^1,$$

and keep x_n fixed. We write

$$(\tilde{G}_\lambda f)(x', x_n) = (T^*_\lambda f)(x'),$$

where we have set $y = \xi$ in the definition (6).

This leads us to the phase function $\Phi(x', \xi)$ on $\mathbf{R}^{n-1} \times \mathbf{R}^n$ given by

$$\Phi(x',\xi) = -(|x'-\xi'|^2 + |x_n - \xi_n|^2)^{1/2},$$

with x_n fixed, and $\xi = (\xi', \xi_n)$. It is not difficult to verify directly that Φ satisfies the conditions of Theorem 1 set out at the beginning of §1.2. Indeed, the vector $\bar{u}$ arising in (7) may be taken to be

$$\frac{x-y}{|x-y|} = \frac{x-\xi}{|x-\xi|}.$$

We can therefore invoke the theorem and obtain (see (8′))

$$\left(\int_{\mathbf{R}^{n-1}} |(\tilde{G}_\lambda f)(x', x_n)|^q\,dx'\right)^{1/q} \le A\lambda^{-n/p'}\|f\|_{L^p(\mathbf{R}^n)}.$$

Next observe that $q \ge p$, and that the integration in x' above is only over a compact set. Thus

$$\int_{\mathbf{R}^{n-1}} |(\tilde{G}_\lambda f)(x', x_n)|^p\,dx' \le A\lambda^{-np/p'}\|f\|^p_{L^p(\mathbf{R}^n)},$$

and a final integration in x_n (again over a compact set) gives

$$\|\tilde{G}_\lambda(f)\|_{L^p(\mathbf{R}^n)} \le A\lambda^{-n/p'}\|f\|_{L^p(\mathbf{R}^n)}. \tag{32}$$

The passage from the inequality for $\tilde{G}_\lambda$ to that for G_λ is then accomplished by a familiar argument (see, for instance, Chapter 6, §2.3). Indeed, (32) implies that

$$\int_{|x-x^0|\le 1} |(G_\lambda f)(x)|^p\,dx \le A\lambda^{-np/p'}\int_{|x-x^0|\le c} |f|^p\,dx$$

for each x_0, where the constant c is determined by the size of the support of ψ. An integration in x^0 then proves (31), establishing the lemma.

2.2.3 To conclude the proof of the proposition, we note that, because $S^\delta(f) = f * K^\delta$, we can use the asymptotic expansion (29) of K^δ and express the operator S^δ as a finite sum as follows. First there is the principal term, given by a constant multiple of

$$f \mapsto \int_{|y|\ge 1} e^{\pm 2\pi i|y|}\,f(x-y)\,|y|^{-(n+1)/2-\delta}\,dy = (Tf)(x). \tag{33}$$

Next, there are finitely many terms of the same kind, but where the factor $|y|^{-(n+1)/2-\delta}$ is replaced by $|y|^{-(n+1)/2-\delta-j}$ (and hence improved) with $j > 0$. Finally, there is an error term, corresponding to convolution with a kernel belonging to $L^1(\mathbf{R}^n)$. Thus it is clear that we need only consider the main term (33).

Observe that $|y|^{-(n+1)/2-\delta}$ may be written as

$$|y|^{-(n+1)/2-\delta} = \sum_{k=0}^{\infty} 2^{-[(n+1)/2+\delta]k}\cdot\psi(y/2^k), \qquad y \ge 1,$$

where ψ is a smooth function, supported in $1/2 \le |y| \le 2$. Indeed, one need only take

$$\psi(y) = |y|^{-(n+1)/2-\delta}[\eta(y) - \eta(2y)],$$

with $\eta(y) = 1$ for $|y| \le 1$, and $\eta(y) = 0$ for $|y| \ge 2$.† We may therefore write $T = \sum_{k=0}^{\infty} T_k$, with

$$(T_k f)(x) = 2^{-[(n+1)/2+\delta]k}\int e^{2\pi i|y|}\,f(x-y)\,\psi(y/2^k)\,dy.$$

† See also Chapter 6 §4.1.

Next, writing $\|\cdot\|$ for the norm of an operator on $L^p(\mathbf{R}^n)$, an obvious re-scaling shows that

$$\|T_k\| = 2^{-[(n+1)/2+\delta]k}\|G_{2\pi 2^k}\| \cdot 2^{nk}.$$

Therefore by (29), if also $1 \leq p \leq 2(n+1)/(n+3)$,

$$\|T_k\| \leq A2^{-[(n+1)/2+\delta]k} \cdot 2^{-nk/p'} \cdot 2^{nk}. \tag{34}$$

The series on the right side of (34) converges exactly when

$$-\left[\frac{n+1}{2}+\delta\right] - \frac{n}{p'} + n < 0,$$

which after some easy manipulations is seen to be equivalent to $p > \dfrac{2n}{n+1+2\delta}$.

Altogether then, since $T = \sum_{k=0}^{\infty} T_k$, we have the desired conclusion for T, and hence for S^δ, whenever $1 \leq p \leq \dfrac{2(n+1)}{n+3}$ and $\dfrac{2n}{n+1+2\delta} < p$. The result for p in the complementary range $\dfrac{2(n+1)}{n-1} \leq p < \infty$, $p < \dfrac{2n}{n-1-2\delta}$, then follows by duality.

3. *Fourier integral operators: L^2 estimates*

Fourier integral operators have in the last 25 years become an important tool in certain areas of analysis, and in particular in a variety of problems arising in partial differential equations. While several generalizations are possible, in its basic form, a *Fourier integral operator* T is given by

$$(Tf)(x) = \int_{\mathbf{R}^n} e^{2\pi i\Phi(x,\xi)}\, a(x,\xi)\, \hat{f}(\xi)\, d\xi. \tag{35}$$

Here $\hat{f}$ denotes the Fourier transform of f; the function a is a symbol of standard type (i.e., it belongs to one of the classes S^m defined in Chapter 6, §1.3), and we assume that it has compact support in x. The phase Φ is real-valued, homogeneous of degree 1 in ξ, and smooth in (x,ξ), for $\xi \neq 0$, on the support of a. We also assume that Φ satisfies the crucial nondegeneracy condition that, for $\xi \neq 0$,

$$\det\left(\frac{\partial^2\Phi}{\partial x_i \partial \xi_j}\right) \neq 0 \tag{36}$$

on the support of a.

We shall now describe (very briefly) some features of operators of this kind.

(i) The simplest examples arise when $\Phi(x,\xi) = x\cdot\xi$. In that case, and when $a(x,\xi) = 1$ for large ξ, we get essentially the identity operator, which expresses the Fourier inversion formula.[‡] More generally, all pseudo-differential operators are of this form; but these examples are far from typical, as we shall see below.

(ii) More representative examples already arise when we consider (the pair of) phase functions $\Phi_\pm$, given by

$$\Phi_\pm(x,\xi) = x\cdot\xi \pm |\xi|.$$

These occur in two related instances. First, if $d\sigma$ is the induced measure on the unit sphere in $\mathbf{R}^n$, then the Fourier transform formula (25) in Chapter 8, and the asymptotics of Bessel functions (15), also in Chapter 8, show that the convolution operator $Tf = f * d\sigma$ is essentially the sum of two Fourier integral operators of order $-(n-1)/2$; see also Chapter 8, §5.7.

The second instance derives from the wave equation

$$\sum_{j=1}^{m} \frac{\partial^2 u}{\partial x_j^2} = \frac{\partial^2 u}{\partial t^2}, \qquad (x,t) \in \mathbf{R}^n \times \mathbf{R}^1.$$

We consider the solution $u = u(x,t)$ determined by the initial conditions

$$u(x,0) = f_0(x), \qquad \frac{\partial u}{\partial t}(x,0) = f_1(x);$$

Then for each fixed t, $u(x,t) = T_0(f_0) + T_1(f_1)$, where T_0, T_1 are sums of Fourier integral operators (with phases $x\cdot\xi \pm t|\xi|$) of orders 0 and -1, respectively. This follows immediately from the classical representation

$$u(x,t) = \int_{\mathbf{R}^n} \cos(2\pi|\xi|t)\, e^{2\pi i x\cdot\xi}\, \hat{f}_0(\xi)\, d\xi + \frac{1}{2\pi}\int_{\mathbf{R}^n} \frac{\sin(2\pi|\xi|t)}{|\xi|} e^{2\pi i x\cdot\xi}\, \hat{f}_1(\xi)\, d\xi.$$

This second example has a wide generalization: The solution of the initial value problem for a (variable coefficient) hyperbolic equation can be well approximated by Fourier integral operators; see also §6.12. This fact provided an important motivation for the development of the whole subject.

[‡] This is the genesis of the name *Fourier integral operator*.

(iii) A key aspect of Fourier integral operators is connected with the fact that they are not pseudo-local: these operators have kernels whose singular support is not limited to the diagonal. Thus

$$Tf(x) = \int K(x,y)\,f(y)\,dy$$

where (formally)

$$K(x,y) = \int_{\mathbf{R}^n} e^{2\pi i[\Phi(x,\xi)-y\cdot\xi]}\,a(x,\xi)\,d\xi\,;$$

then, as indicated by the localization principle (of Chapter 8, §2.1), $K(x,y)$ can be expected to be singular for those (x,y) where

$$\nabla_\xi[\Phi(x,\xi) - y\cdot\xi] = 0$$

for some ξ.

So if we write $\Phi_\xi = \nabla_\xi(\Phi)$, we are led to consider, for each x, the variety

$$\Sigma_x = \{y : y = \Phi_\xi(x,\xi) \text{ for some } \xi\}. \tag{37}$$

In general, Σ_x is the locus of singularities of the "function" $y \mapsto K(x,y)$. Since Φ_ξ is homogeneous of degree 0, Σ_x is the image via a smooth mapping of the unit sphere, and has dimension at most $n-1$. However, the variety Σ_x can itself be quite singular since we do *not* assume that the rank of the matrix $\Phi_{\xi\xi} = \nabla_\xi\nabla_\xi(\Phi)$ is maximal (or constant).

Indeed, from this point of view the examples in (ii) are still quite special, because there $\text{rank}(\Phi_{\xi\xi}) \equiv n-1$; of course, then Σ_x is a sphere (of radius 1 or t) centered at x. Note that in (i), $\text{rank}(\Phi_{\xi\xi}) = 0$ and $\Sigma_x = \{x\}$.

(iv) Another impetus for the theory is its connection with ideas arising from classical mechanics, in particular with canonical transformations of the (x,ξ) space. It turns out that conjugating with a suitable Fourier integral operator of the form (35) allows one to approximately transform a pseudo-differential operator T_a into another pseudo-differential operator T_b where the symbols $a(x,\xi)$ and $b(x,\xi)$ are related by the canonical transformation whose generating function is Φ, namely

$$b(\Phi_\xi(x,\xi),\xi) = a(x,\Phi_x(x,\xi)).$$

This theorem is part of a broader calculus dealing with the study of the composition of Fourier integral operators. We shall not pursue these matters here, but the reader may consult §6.11 for further references.

3.1 L^2 estimates. The basic result for the L^2 theory is as follows.

PROPOSITION. *Let T be the Fourier integral operator given by* (35), *as described above. We assume also that the symbol a is of order 0, i.e., that $a \in S^0$ (in the terminology of Chapter 6, §1.3), and that a has compact x-support. Then T, initially defined on $\mathcal{S}$, extends to a bounded operator from $L^2(\mathbf{R}^n)$ to itself.*

3.1.1 Proof. We can write $T = S\mathcal{F}$, where $\mathcal{F}(f) = \widehat{f}$ is the Fourier transform, and

$$Sf(x) = \int_{\mathbf{R}^n} e^{2\pi i\Phi(x,\xi)}\,a(x,\xi)\,f(\xi)\,d\xi.$$

By Plancherel's theorem, matters are then reduced to a similar assertion for the operator S. We first take the precaution of modifying S so the the support of its symbol $a(x,\xi)$ is contained in a sufficiently "narrow cone" in the ξ-space; by this we mean that there is a small constant $\underline{c}$, so that whenever ξ and η belong to the cone and $|\eta| \le |\xi|$, then, if we write $\eta = \rho\xi + \eta'$ with η' orthogonal to ξ, we have $|\eta'| \le \underline{c}|\xi|$. Of course, any initially given S can be written as a finite sum of operators corresponding to such narrow cones by using a suitable finite smooth partition of unity in the ξ-space, the elements of which are functions that are homogeneous of degree 0 for large ξ.

We next claim that, whenever ξ and η belong to the same narrow cone,

$$|\nabla_x[\Phi(x,\xi) - \Phi(x,\eta)]| \ge c|\xi - \eta|; \tag{38}$$

here c is a positive constant. This inequality, as well as the others appearing below, is restricted to those x belonging to a compact set in $\mathbf{R}^n$ that contains the x-support of a.

In the sequel, we abbreviate by writing Φ_x for the vector $\nabla_x[\Phi(x,\xi)]$, and $\Phi_{x,\xi}$ for the matrix $\left[\partial^2\Phi/\partial x_i\partial\xi_j\right]$. Then since Φ_x is homogeneous in ξ of degree 1, by Euler's theorem we must have $\Phi_x(x,\xi) = \Phi_{x,\xi}(x,\xi)\cdot\xi$. However, by (36), the matrix $\Phi_{x,\xi}$ is nonsingular; thus

$$|\Phi_{x,\xi}(x,\xi)(u)| \ge c|u| \qquad \text{and} \qquad |\Phi_x(x,\xi)| \ge c|\xi|. \tag{39}$$

To prove (38), we observe that by symmetry in ξ and η, and homogeneity of Φ, it suffices to establish it when $|\xi| = 1$, $|\eta| \le 1$. Assume first that $|\xi - \eta| \le c_1$, where c_1 is another small constant. Then

$$\Phi_x(x,\xi) - \Phi_x(x,\eta) = \Phi_{x,\xi}(x,\xi)\cdot(\xi-\eta) + O(|\xi-\eta|^2),$$

so $|\Phi_x(x,\xi) - \Phi_x(x,\eta)| \ge c|\xi - \eta|$, by the first inequality in (39).

Next, if $|\xi - \eta| \ge c_1$, then write $\eta = \rho\xi + \eta'$ as above, and note that

$$\Phi_x(x,\xi) - \Phi_x(x,\eta) = [\Phi_x(x,\xi) - \Phi_x(x,\rho\xi)] + [\Phi_x(x,\rho\xi) - \Phi_x(x,\rho\xi+\eta')].$$

The first quantity in brackets equals $(1-\rho)\Phi_x(x,\xi)$, whose length is

$$(1-\rho)|\Phi_x(x,\xi)| \geq c(1-\rho) \geq c|\xi-\eta|,$$

by (39); while the second is majorized by

$$O(|\eta'|) \leq O(\underline{c}) = O(\text{small constant} \cdot |\xi-\eta|),$$

by adjusting $\underline{c}$ if necessary. These remarks establish (38).

3.1.2 Having disposed of these preliminaries, we turn to the proof of the boundedness of S, which can be accomplished in a few strokes.

For simplicity, we assume first that the symbol $a(x,\xi)$ has compact support in ξ, but our estimates will be independent of the size of that support. The operator S is then bounded, and we can estimate its norm by estimating the norm of the operator S^*S. We write the latter operator in the form

$$(S^*Sf)(\xi) = \int_{\mathbf{R}^n} K(\xi,\eta)\, f(\eta)\, d\eta \tag{40}$$

where the kernel K is given by

$$K(\xi,\eta) = \int_{\mathbf{R}^n} e^{2\pi i[\Phi(x,\eta)-\Phi(x,\xi)]}\, \bar{a}(x,\xi)\, a(x,\eta)\, dx. \tag{41}$$

Now (38) guarantees that $|\nabla_x[\Phi(x,\eta)-\Phi(x,\xi)]| \geq c|\xi-\eta|$; thus by Proposition 4 in Chapter 8, §2.1, and the fact that $a(x,\xi)$ has compact support in x, we see that for every $N \geq 0$

$$|K(\xi,\eta)| \leq A_N(1+|\xi-\eta|)^{-N}. \tag{42}$$

As a result, if we take $N = n+1$, and apply the lemma in Chapter 7, §2.4.1, we get the boundedness of S^*S, and therefore of S.

To lift the restriction that $a(x,\xi)$ has compact support in ξ, we repeat the familiar device of replacing $a(x,\xi)$ by $a_\varepsilon(x,\xi) = a(x,\xi)\cdot\gamma(\varepsilon\xi)$, where $\gamma \in C_0^\infty$ with $\gamma(0) = 1$; then by what we have just proved, the operators S_ε, with symbols a_ε, have a uniform bound for their norms. Moreover, it is obvious that $S_\varepsilon(f) \to S(f)$ in $\mathcal{S}$, whenever $f \in \mathcal{S}$. This establishes the boundedness of S in general and concludes the proof of the proposition.

3.1.3 Contrary to the special case of pseudo-differential operators, Fourier integral operators whose symbols are of order 0 may not be bounded on any $L^p(\mathbf{R}^n)$, $p \neq 2$. The example of this phenomenon that is easiest to compute directly arises when $n = 3$.

Write $d\sigma$ for the usual measure on the unit sphere $\mathbf{S}^2 \subset \mathbf{R}^3$. According to the remarks in (ii) at the beginning of §3, the operator

$$f \mapsto \frac{\partial}{\partial x_j}(f * d\sigma) \qquad j = 1, 2, \text{ or } 3,$$

is essentially a Fourier integral operator of order 0.

Next let f be a compactly supported function with $f(x) = |x|^{-\alpha}$ for small x, and with f bounded away from the origin. A straightforward calculation gives that

$$|\nabla(f * d\sigma)| \geq A|(1-|x|)|^{1-\alpha},$$

provided $1 < \alpha < 3$. Now $f \in L^p(\mathbf{R}^3)$ when $\alpha p < 3$, while $|(1-|x|)|^{1-\alpha} \notin L^p$ if $\alpha p \geq p$.

Thus, if $1 \leq p < 2$, the choice $\alpha = 1 + 1/p$ shows that

$$f \mapsto \nabla(f * d\sigma)$$

cannot be bounded on L^p; duality extends this observation to $p > 2$. For other examples of this type, see §6.13.

3.1.4 As a consequence of the above proposition (and also by use of similar methods), we can obtain the following variant related to "fractional integration".

PROPOSITION. *Let T be a Fourier integral operator of the type* (35) *whose symbol is of order m, with $-n/2 < m < 0$. Then T, initially defined on $\mathcal{S}$, extends as a bounded operator:*

(a) *from $L^p(\mathbf{R}^n)$ to $L^2(\mathbf{R}^n)$, if $1/p = 1/2 - m/n$.*

(b) *from $L^2(\mathbf{R}^n)$ to $L^q(\mathbf{R}^n)$, if $1/q = 1/2 + m/n$.*

Proof. We first treat case (a).

Note that $T = T_0T_1$, where T_0 is a Fourier integral operator with symbol

$$a_0(x,\xi) = a(x,\xi)\cdot(1+|\xi|^2)^{-m/2}$$

(which is of order 0), and T_1 is the pseudo-differential operator with symbol $(1+|\xi|^2)^{m/2}$ (which of course is of order m). Since $T_0 : L^2 \to L^2$ is bounded (as was just proved), we need only show that $T_1 : L^p \to L^2$ is bounded.

We use the kernel realization of T_1, guaranteed by Chapter 6, §4, Proposition 1, which gives

$$T_1 f(x) = \int k(x-y)\, f(y)\, dy, \tag{43}$$

Monographs in Harmonic Analysis

I. *Introduction to Fourier Analysis on Euclidean Spaces*, by E. M. Stein and G. Weiss

II. *Singular Integrals and Differentiability Properties of Functions*, by E. M. Stein

III. *Harmonic Analysis: Real-Variable Methods, Orthogonality, and Oscillatory Integrals*, by E. M. Stein

Harmonic Analysis: Real-Variable Methods, Orthogonality, and Oscillatory Integrals

ELIAS M. STEIN

with the assistance of Timothy S. Murphy

Princeton University Press
Princeton, New Jersey
1993

where $k(z) = \sum_j k_j(z)$; here the k_j are as in Chapter 6, §4.2 and are independent of x. The sum converges dominatedly to the locally integrable function k, which satisfies the estimate

$$|k(z)| \leq A|z|^{-n-m};$$

as a result, the representation (43) holds as a convergent integral for every x, if $f \in \mathcal{S}$.

We can now apply the Hardy-Littlewood-Sobolev inequality ((31) of Chapter 8) with $\gamma = n + m$, $q = 2$; the resulting relation between exponents is just $1/p = 1/2 - m/n$, and part (a) is proved.

We could prove (b) by duality, if we knew that T^* were of the same form as T, for then $T^* : L^p \to L^2$ would be bounded; p and q are, of course, conjugate exponents. While T^* is essentially of the form (35), establishing this requires a separate argument. It is best, then, to proceed directly with the estimate of T^*. Let us write $T = S\mathcal{F}$ as in §3.1.1, and note that $T^* = \mathcal{F}^* S^*$, so it suffices to see that $S^* : L^p \to L^2$ is bounded. Repeating an orthogonality argument used previously,

$$\|S^*(f)\|_{L^2}^2 = \langle S^* f, S^* f\rangle = \langle SS^* f, f\rangle,$$

so by Hölder's inequality it is enough to prove that

$$SS^* : L^p \to L^q \qquad \text{boundedly.} \tag{44}$$

In showing (44), we assume that the symbol $a(x,\xi)$ has sufficiently small x-support; in any case, our given S can be written as a finite sum of operators whose symbols satisfy this extra restriction. Note next that

$$\nabla_\xi[\Phi(x,\xi) - \Phi(y,\xi)] = \Phi_{x,\xi}(x,\xi)\cdot(x-y) + O(|x-y|^2),$$

so that if we sufficiently restrict the support of a, keeping in mind the nondegeneracy condition (36), we get

$$|\nabla_\xi[\Phi(x,\xi) - \Phi(y,\xi)]| \geq c|x-y|, \tag{45}$$

if x and y are in the support of our symbol a.

We also take the precaution of replacing our symbol $a(x,\xi)$ with

$$a_\varepsilon(x,\xi) = \gamma(\varepsilon\xi)\cdot a(x,\xi),$$

where $\gamma \in C_0^\infty$ and $\gamma(0) = 1$. We note that a_ε satisfies the same differential inequalities that a does, uniformly in ε, $0 < \varepsilon \leq 1$. Dealing with a_ε instead of a allows us to carry out the manipulations required below. Since the bounds obtained will be independent of ε, a passage to the limit as $\varepsilon \to 0$ will give us our desired result. We do not indicate the dependence on ε in the sequel.

Now $(SS^*f)(x) = \int K(x,y)\,f(y)\,dy$, where

$$K(x,y) = \int_{\mathbf{R}^n} e^{2\pi i[\Phi(x,\xi)-\Phi(y,\xi)]}\,a(x,\xi)\,\bar{a}(y,\xi)\,d\xi.$$

We decompose this expression by using the dyadic partition of unity arising in §4.1 of Chapter 6,

$$1 = \sum_{j=-\infty}^{\infty} \widehat{\Psi}_j(\xi).$$

Making the indicated change of variables gives

$$K(x,y) = \sum_{j=-\infty}^{\infty} K_j(x,y),$$

with

$$K_j(x,y) = 2^{nj}\int_{\mathbf{R}^n} e^{2\pi i 2^j[\Phi(x,\xi)-\Phi(y,\xi)]}\,a_j(x,y,\xi)\,d\xi; \tag{46}$$

here $a_j(x,y,\xi) = a(x,2^j\xi)\cdot\bar{a}(y,2^j\xi)\,\widehat{\Psi}(\xi)$, because $\widehat{\Psi}_j(\xi) = \widehat{\Psi}(2^{-j}\xi)$, and $\Phi(x,\xi)$ is homogeneous of degree 1 in ξ.

Note also that $a_j(x,y,\xi)$ is supported in $|\xi| \leq 2$, while the fact that $a \in S^m$ implies that the $2^{2mj}a_j(x,y,\xi)$ are uniformly bounded (as j varies), and the same is true of all the partial derivatives. Then because of (45) we get, by the proposition in Chapter 8, §2.1, that

$$|K_j(x,y)| \leq A_N 2^{nj}\cdot 2^{2mj}\cdot(2^j|x-y|)^{-N}, \qquad N \geq 0. \tag{47}$$

Next, write

$$K = \sum K_j = \sum_{2^j|x-y|\leq 1} + \sum_{2^j|x-y|>1}.$$

For the first sum, use (47) with $N = 0$, and recall that $m > -n/2$. For the second sum, apply (47) with $N = m$, and note also that $m < 0$. Altogether then

$$|K(x,y)| \leq A|x-y|^{-n-2m},$$

and an application of the fractional integration inequality (31) of the previous chapter (now with $\gamma = n + 2m$), shows that (44) is satisfied, completing the proof of the proposition.

Second printing, with corrections and additions, 1995

Published by: Princeton University Press, 41 William Street,
Princeton, New Jersey 08540

In the United Kingdom: Princeton University Press, Chichester, West Sussex

Library of Congress Cataloging-in-Publication Data

Stein, Elias M., 1931–
Harmonic analysis : real-variable methods, orthogonality, and oscillatory integrals / Elias M. Stein, with the assistance of Timothy S. Murphy.

p. cm. — (Princeton mathematical series ; 43)
Includes bibliographical references and index.
ISBN 0-691-03216-5
1. Harmonic analysis. I. Murphy, Timothy S. II. Title.
III. Series.
QA403.3.S74 1993
515′.785–dc20 92–44035

This book has been composed in Computer Modern.

The publisher would like to acknowledge Timothy Murphy for providing the camera-ready copy from which this book was printed. The production of this book made extensive use of free computer software, most notably the TEX typesetting system, X Windows, and GNU Emacs. Northfield Trading L. P. provided computer equipment.

Princeton University Press books are printed on acid-free paper and meet the guidelines for permanence and durability of the Committee on Production Guidelines for Book Longevity of the Council on Library Resources.

Printed in the United States of America

10 9 8 7 6 5 4 3

To my students and collaborators

Contents

Preface

Since the publication of the previous books in this series more than twenty years ago,[†] the subject of harmonic analysis has undergone a vast development. A succession of new departures has in many ways transformed the whole field. It has brought with it additional insights, extending significantly the range of previous ideas. With all of this has come a further clarification of the essential unity linking several of the main areas of analysis.

What has been achieved includes: the broadening of the scope of real-variable methods to encompass, among other matters, the theory of Hardy spaces; the further study of the Fourier transform leading to summability and restriction theorems; the analysis of L^2 methods, exploiting notions of orthogonality and oscillatory integrals; the application of these ideas to situations where geometric properties related to curvature play a role; and other ramifications of some of these concepts in the context of the Heisenberg group. These ideas and their consequences have not only had a profound effect on the domain of Fourier analysis *per se* but have also had a major influence in such areas as partial differential equations, several complex variables, and analysis on symmetric spaces.

It is the objective of this book to try to give an account of the main lines of these developments. Given the sweep of the subject, I think of the task involved as being in some ways akin to the telling of a long and complicated story: an epic tale stretching over several decades, involving various principal characters that appear and reappear (sometimes in disguised form), and following a complex plot with a number of intricate subplots. To pursue this analogy further, I cannot deny that this book is in part autobiographical: as the narrator of the story, I have chosen to recount those matters I know best by virtue of having first-hand knowledge of their unravelling.

A few words about the organization of this book. It is divided into three parts, in accordance with the subject matter. The first five chapters take up real-variable theory; the next six chapters emphasize L^2 methods and oscillatory integrals; the last two chapters introduce analysis on the Heisenberg group and also provide a retrospective view of some of the

[†] See Stein and Weiss [1971] and Stein [1970c], cited in the bibliography.

preceding material. The exposition is guided by the desire that, in each chapter, the material should be developed in the service of the proofs of a few central theorems. In doing this, I have often not formulated the final results in maximal generality, nor have I always chosen the shortest proofs. It is hoped, however, that the reader will ultimately see the advantage of this approach.

It is my pleasure to acknowledge my deep indebtedness to Timothy Murphy for indispensable help in writing this book. He joined me in this effort about three years ago; up to that time only rudimentary progress had been made. From then on, our constant conversations and his many incisive suggestions spurred the work and helped refine the material into its present form. He also took complete charge of the copy editing and typesetting of the manuscript, which he carried out with consummate skill.

I also wish to express my thanks to those others who have aided me: Daryl Geller, Fulvio Ricci, Cora Sadosky, and Christopher Sogge, who made valuable suggestions that have been incorporated in the text; also D. H. Phong, Robert Fefferman, David Jerison, Andrew Bennett, Peter Heller, Andrew Neff, and Der-Chen Chang, who prepared lecture notes of graduate courses I gave at Princeton University during the period 1972 to 1987, on which parts of this book are based.

Lastly, I wish to express my gratitude to all my students and collaborators whose ideas enlighten these pages. This book is in large measure a record of their achievement.

Elias M. Stein
December 1992

Guide to the Reader

The core of the book, which appears in standard type, consists of the thirteen chapters, excluding the appendices and sections titled "Further Results". Written to be as self-contained as possible, its object is to present the main ideas without undue adornment. In addition, the following features should be noted.

Appendices are intended as elaborations of previously treated core subjects and are given with substantial sketches of proofs.

Further Results are meant to survey the vast number of additional extensions and cognate topics; these are often presented as mere reformulations of theorems in the cited literature, sometimes also with some indication of proof.

Previous monographs in this series† are cited from time to time for helpful background material. A number of interesting related topics, not pursued in this book, can also be found there, as well as earlier, more elemental (and possibly more transparent) approaches to some of the matters treated here.

† Stein and Weiss [1971] and Stein [1970c]; in the sequel, these books will be referred to simply as *Fourier Analysis* and *Singular Integrals*.

Harmonic Analysis

Real-Variable Methods, Orthogonality, and Oscillatory Integrals

4. Fourier integral operators: L^p estimates

We have seen that a Fourier integral operator of order 0 is, in general, not bounded from L^p to itself, if $p \neq 2$. To have this boundedness requires some smoothing, the degree of which depends on the exponent p. The result expressing this is as follows:

THEOREM 2. *Let T be a Fourier integral operator, as in* (35), *whose symbol a is of order m (i.e., $a \in S^m$), with $(1-n)/2 < m \leq 0$. Then T, initially defined on $\mathcal{S}$, extends to a bounded operator from $L^p(\mathbf{R}^n)$ to itself, whenever*

$$\left|\frac{1}{2} - \frac{1}{p}\right| \leq \frac{-m}{n-1}. \tag{48}$$

Remarks. This is the optimal conclusion, as a modification of the example in §3.1.3 shows; further details are in §6.13. Note that when $n = 1$, the appropriate requirement is $m \leq 0$, and in this case the result is much simpler; see also §6.20. Finally, observe that under the restriction (48), the condition $-(n-1)/2 < m$ implies that $1 < p < \infty$.

4.1 Scheme of the proof. The main point of the proof is to show that when the symbol a has order $-(n-1)/2$, then the operator is a bounded mapping from the Hardy space $H^1(\mathbf{R}^n)$ to $L^1(\mathbf{R}^n)$. In order to explain the principal ideas of the argument we shall first review the idea of the proof of the corresponding result for singular integrals and pseudo-differential operators.[†]

Suppose a is an atom associated to a ball B; it suffices to show that

$$\int_{\mathbf{R}^n} |Ta(x)|\, dx \leq A. \tag{49}$$

The region of integration in (49) is divided into two parts: the "region of influence" of B, namely the ball B^* concentric with B having (say) twice the radius, and its complement ${}^cB^*$. Note that B^* carries the preponderance of the effect of the atom (supported in B), because of the pseudo-local properties of T.

For the estimate on B^*, one uses the L^2 theory of T, together with the fact that

$$|B^*| \leq c|B|. \tag{50}$$

On the complement of B^*, the inequality

$$\int_{{}^cB^*} |K(x,y) - K(x,y')|\, dx \leq A, \qquad \text{if } y, y' \in B, \tag{51}$$

† As in Chapter 3, §3.1 and Chapter 6, §5.1.

is crucial; here K is the kernel of the operator T. This inequality is closely connected with the fact that

$$\int_{\mathbf{R}^n} |K_j(x,y)| \leq A, \tag{52}$$

where K_j are the kernels arising from K via the Littlewood-Paley decomposition of the frequency domain.[‡]

Turning to the case of general Fourier integral operators, the description of the region of influence of B must now take into account the nature of the more general singular sets Σ_x given by (37). It is therefore reasonable to expect that B^* should here be something like

$$B^* = \{x : \operatorname{dist}(\Sigma_x, B) \leq c\delta\},$$

where δ is the radius of B; the actual B^* is more complicated, due to the fact that the sets Σ_x are not regular. Since in general Σ_x may be $n-1$ dimensional, the estimate (50) (i.e., $|B^*| \leq c\delta^n$) has to be replaced by

$$|B^*| \leq c\delta, \qquad \text{for small } \delta, \tag{53}$$

and a corresponding change in the L^2 estimate for T is needed. This is given by the (L^p, L^2) inequality in §3.1.4.

The analogue of (52) requires a *second* dyadic decomposition, superimposed on the first: each dyadic shell $2^{j-1} \leq |\xi| \leq 2^{j+1}$ is partitioned into "thin" truncated cones of thickness roughly $2^{j/2}$; each such truncated cone is essentially an elongated rectangle whose major axis has length $\sim 2^j$, while all the other sides have length $\sim 2^{j/2}$. Roughly $2^{(n-1)j/2}$ such elements are needed to cover the shell $2^{j-1} \leq |\xi| \leq 2^{j+1}$ (see Figure 1 in the introduction to this chapter).

The role of the truncated cones can be understood as follows. When ξ ranges over one of these cones, the phase $\Phi(x,\xi) - y \cdot \xi$ is essentially *linear* in ξ: it is well approximated by $[\Phi_\xi(x,\xi_j^\nu) - y] \cdot \xi$, where ξ_j^ν is a fixed unit direction. Thus the integral (67) below (with $\bar{\xi} = \xi_j^\nu$), which gives the kernel, is basically a Fourier transform and can then be easily estimated.

4.2 We now give the exact description of the influence set B^* associated to a ball B, whose radius is assumed to be at most 1. First, for each positive integer j, we consider a (roughly) equally spaced set of points with grid length $2^{-j/2}$ on the unit sphere $\mathbf{S}^{n-1}$; that is, we fix a collection $\{\xi_j^\nu\}_\nu$ of unit vectors, $|\xi_j^\nu| = 1$, that satisfy:

(i) $|\xi_j^\nu - \xi_j^{\nu'}| \geq 2^{-j/2}$, if $\nu \neq \nu'$.

(ii) If $\xi \in \mathbf{S}^{n-1}$, then there exists a ξ_j^ν so that $|\xi - \xi_j^\nu| < 2^{-j/2}$.

‡ As in Chapter 6, §4.2 and §5.4.

To do this, simply take a maximal collection $\{\xi_j^\nu\}$ for which (i) holds. Notice that there are at most $c2^{j(n-1)/2}$ elements in the collection $\{\xi_j^\nu\}_\nu$.

Now let $B = B(\bar{y}, \delta)$ be the ball about $\bar{y}$ with radius $\delta \le 1$. We define the "rectangles" $R_j^\nu = R_j^\nu(B)$ in the x-space as follows. First, we let the rectangles $\tilde{R}_j^\nu$ in the y-space be given by

$$\tilde{R}_j^\nu = \{y : |y - \bar{y}| \le \bar{c}2^{-j/2}, |\pi_j^\nu(y - \bar{y})| \le \bar{c}2^{-j}\},$$

where π_j^ν is the orthogonal projection in the direction ξ_j^ν and $\bar{c}$ is a large constant (independent of j), to be fixed later. Next, the mapping

$$x \mapsto y = \Phi_\xi(x, \xi)$$

has, for each ξ, a nonvanishing Jacobian, according to (36). So if $2^{-j} \le 1$, we take R_j^ν to be the inverse under Φ_ξ, with $\xi = \xi_j^\nu$, of the rectangle $\tilde{R}_j^\nu$:

$$R_j^\nu = \{x : |\bar{y} - \Phi_\xi(x, \xi_j^\nu)| \le \bar{c}2^{-j/2}, |\pi_j^\nu(\bar{y} - \Phi_\xi(x, \xi_j^\nu))| \le \bar{c}2^{-j}\}. \tag{54}$$

Now let $B^* = \bigcup_{2^{-j}\le\delta} \bigcup_\nu R_j^\nu$; then

$$|B^*| \le \sum_{2^{-j}\le\delta} \sum_\nu |R_j^\nu| \le c \sum_{2^{-j}\le\delta} \sum_\nu |\tilde{R}_j^\nu|$$
$$\le c \sum_{2^{-j}\le\delta} 2^{-j(n+1)/2} \cdot 2^{j(n-1)/2} = c \sum_{2^{-j}\le\delta} 2^{-j} \le c\delta,$$

and the property (53) holds for B^*.

4.3 We turn to the estimate for $T(a)$ where a is an atom supported in the ball B. To begin with, if the radius of B exceeds 1, the estimate is easy, because of our assumption that the symbol of T has compact x-support. Indeed,

$$\int |T(a)|\,dx \le c\|T(a)\|_{L^2} \le c'\|a\|_{L^2}.$$

The first inequality holds because $T(a)$ has fixed compact support; the second follows from the L^2 boundedness property already proved (under the weaker assumption that the symbol has order 0). Now since a is an atom, $|a(x)| \le |B|^{-1}$, we have $\|a\|_{L^2} \le |B|^{-1/2} \le c$ and (49) holds in this case.

Next, assume that radius $B = \delta \le 1$. Then

$$\int_{B^*} |Ta|\,dx \le \|Ta\|_{L^2} \cdot |B^*|^{1/2} \le c\delta^{1/2}\|Ta\|_{L^2}.$$

However, by part (a) of the proposition in §3.1.4, $\|Ta\|_{L^2} \le A\|a\|_{L^p}$, if

$$\frac{1}{p} = \frac{1}{2} + \frac{n-1}{2n}, \tag{55}$$

since we are dealing with an operator whose symbol has order $m = -(n-1)/2$. Now $\|a\|_{L^p} \le |B|^{-1+1/p}$, since $|a(x)| \le |B|^{-1}$ and a is supported in B. Thus we get

$$\int_{B^*} |Ta|\,dx \le A\delta^{1/2} \cdot \delta^{n(-1+1/p)} = A,$$

since (55) is equivalent to $1/2 + n(-1+1/p) = 0$. We therefore have that

$$\int_{B^*} |Ta|\,dx \le A. \tag{56}$$

4.4 Further dyadic decomposition. We now detail the second "dyadic" decomposition of the ξ-space that is needed.† We recall the unit vectors $\{\xi_j^\nu\}$ used above; they give an essentially uniform grid on the unit sphere, with separation $2^{-j/2}$. Let Γ_j^ν denote the corresponding cone in the ξ-space whose central direction is ξ_j^ν, i.e.,

$$\Gamma_j^\nu = \{\xi : |\xi/|\xi| - \xi_j^\nu| \le 2 \cdot 2^{-j/2}\}.$$

We can construct an associated partition of unity: it is given by functions χ_j^ν, each homogeneous of degree 0 in ξ and supported in Γ_j^ν, with

$$\sum_\nu \chi_j^\nu(\xi) = 1 \qquad \text{for all } \xi \ne 0 \text{ and all } j, \tag{57}$$

and

$$|\partial_\xi^\alpha \chi_j^\nu(\xi)| \le A_\alpha 2^{|\alpha|j/2}|\xi|^{-|\alpha|}. \tag{58}$$

In fact, fix a smooth, nonnegative function ϕ with $\phi(u) = 1$ for $|u| \le 1$, and $\phi(u) = 0$ for $|u| \ge 2$. Let

$$\eta_j^\nu(\xi) = \phi(2^{j/2}[\xi/|\xi| - \xi_j^\nu]),$$

and define $\chi_j^\nu = \eta_j^\nu \cdot (\sum_j \eta_j^\nu)^{-1}$. Then, because of properties (i) and (ii) enjoyed by the unit vectors $\{\xi_j^\nu\}$, it is an easy matter to verify (57) and (58).

† This decomposition is schematized in the figure that appears near the beginning of the present chapter.

From (57) follows the refined Littlewood-Paley decomposition:

$$1 = \widehat{\Psi}_0(\xi) + \sum_{j=1}^{\infty} \sum_{\nu} \chi_j^\nu(\xi) \cdot \widehat{\Psi}_j(\xi). \tag{59}$$

With this decomposition, we define the operators T_j^ν by

$$T_j^\nu f(x) = \int_{\mathbf{R}^n} e^{2\pi i \Phi(x,\xi)}\, a_j^\nu(x,\xi)\, \widehat{f}(\xi)\, d\xi,$$

where $a_j^\nu(x,\xi) = \chi_j^\nu(\xi) \cdot \widehat{\Psi}_j(\xi) \cdot a(x,\xi)$; we also define corresponding operators T_j, having symbols $a_j(x,\xi) = \widehat{\Psi}_j(\xi) \cdot a(x,\xi)$. Clearly

$$T_j = \sum_\nu T_j^\nu. \tag{60}$$

We let K_j denote the kernel of the operator T_j. The key estimates we shall make are:

$$\int_{\mathbf{R}^n} |K_j(x,y)|\, dx \le A, \qquad \text{all } y \in \mathbf{R}^n, \tag{61}$$

$$\int_{\mathbf{R}^n} |K_j(x,y) - K_j(x,y')|\, dx \le A|y-y'| \cdot 2^j, \quad \text{all } y, y' \in \mathbf{R}^n, \tag{62}$$

$$\int_{^cB^*} |K_j(x,y)|\, dx \le \frac{A \cdot 2^{-j}}{\delta}, \text{ if } y \in B \text{ and } 2^{-j} \le \delta, \tag{63}$$

where the bound A is independent of j, y, y', and δ; recall that the radius of B is δ.

Because of (60), it suffices to make similar estimates for the kernels of the T_j^ν, which we denote by K_j^ν; however, for these K_j^ν the corresponding right-hand sides must incorporate a further factor of $2^{-j(n-1)/2}$, since there are essentially $2^{j(n-1)/2}$ terms in the decomposition (60). We now turn to these estimates.

4.5 Majorization of the kernels. By the definition of T_j^ν, we see that its kernel is given by

$$K_j^\nu(x,y) = \int e^{2\pi i[\Phi(x,\xi) - y\cdot\xi]}\, a_j^\nu(x,\xi)\, d\xi.$$

To simplify the writing of the estimates for $K_j^\nu(x,y)$, we set $\bar{\xi} = \xi_j^\nu$, and choose axes in the ξ-space so that ξ_1 is in the direction of $\bar{\xi}$ and $\xi' = (\xi_2, \dots, \xi_n)$ is perpendicular to $\bar{\xi}$.

Now $\Phi(x,\xi) - y\cdot\xi = [\Phi_\xi(x,\bar{\xi}) - y]\cdot\xi + [\Phi(x,\xi) - \Phi_\xi(x,\bar{\xi})\cdot\xi]$. We set

$$h(\xi) = \Phi(x,\xi) - \Phi_\xi(x,\bar{\xi}) \cdot \xi$$

and claim the following two estimates for $h(\xi)$:

$$\left| \left(\frac{\partial}{\partial \xi_1}\right)^N h(\xi) \right| \le A_N \cdot 2^{-Nj}, \tag{64}$$

$$|(\nabla_{\xi'})^N h(\xi)| \le A_N \cdot 2^{-Nj/2}. \tag{65}$$

These are to hold if $N \ge 1$ and ξ is restricted to the support of $a_j^\nu(x,\xi)$; the latter requires that, in our coordinates,

$$2^j \le |\xi| \le 2^{j+1}, \qquad \text{and} \qquad |\xi'| \le c2^{j/2}.$$

Since $\Phi(x,\xi)$ is homogeneous of degree 1 in ξ,

$$\Phi_\xi(x,\xi) \cdot \xi = \Phi(x,\xi);$$

hence $h(\xi_1, 0, \dots, 0) = 0$. Moreover, it is clear that

$$(\nabla_\xi h)(\xi_1, 0 \dots, 0) = 0.$$

As a result

$$\left(\frac{\partial}{\partial \xi_1}\right)^N h(\xi_1, 0, \dots, 0) = \nabla_{\xi'} \left(\frac{\partial}{\partial \xi_1}\right)^N h(\xi_1, 0, \dots, 0) = 0.$$

Thus

$$\left(\frac{\partial}{\partial \xi_1}\right)^N h(\xi_1, \xi') = O(|\xi'|^2 \cdot |\xi|^{-N-1}),$$

since h is homogeneous of degree 1. However, when $|\xi'| \le c2^{j/2}$ and $|\xi| \sim 2^j$, we have that

$$|\xi'|^2 \cdot |\xi|^{-N-1} = O(2^j \cdot 2^{-(N+1)j}) = O(2^{-Nj}),$$

and (64) is proved.

Again, $(\nabla_{\xi'} h)(\xi_1, 0, \cdot, 0) = 0$, so

$$\nabla_{\xi'} h(\xi_1, \xi') = O(|\xi'| \cdot |\xi|^{-1}),$$

which gives (65) when $N = 1$. For $N \ge 2$, we have that

$$|(\nabla_\xi)^N h(\xi)| \le A|\xi|^{1-N},$$

by homogeneity, and $2^{(1-N)j} \le 2^{-Nj/2}$, so (65) is proved for all $N \ge 1$.

At this stage we should point out that (58) has an improvement when we differentiate in the ξ_1-direction. In fact, we have

$$\left| \left(\frac{\partial}{\partial \xi_1}\right)^N \chi_j^\nu(\xi) \right| \le A_N |\xi|^{-N}, \qquad N \ge 1. \tag{66}$$

This is because in the cone of support of χ_j^ν we can write

$$\frac{\partial}{\partial \xi_1} = \frac{\partial}{\partial r} + O(2^{-j/2}) \cdot \nabla_\xi$$

where $\partial/\partial r$ is the radial derivative; since χ is homogeneous of degree 0, $(\partial/\partial r)^N \chi = 0$.

4.6 We now rewrite K_j^ν as

$$K_j^\nu(x,y) = \int_{\mathbf{R}^n} e^{2\pi i[\Phi_\xi(x,\bar{\xi})-y]\cdot\xi}\, b_j^\nu(x,\xi)\,d\xi \tag{67}$$

where

$$b_j^\nu(x,\xi) = e^{2\pi i h(\xi)}\, a_j^\nu(x,\xi) = e^{2\pi i h(\xi)}\, \chi_j^\nu(\xi)\cdot \widehat{\Psi}_j(\xi)\cdot a(x,\xi)$$

We next introduce the operator L defined by

$$L = I - 2^{2j}\frac{\partial^2}{\partial \xi_1^2} - 2^j \nabla_{\xi'}.$$

Because of (58), (66), (64), (65), and the fact that a is a symbol in $S^{-(n-1)/2}$, we get that

$$|L^N(b_j^\nu(x,\xi))| \le A_N\cdot 2^{-j(n-1)/2}.$$

However, $L^N(e^{2\pi i[\Phi_\xi(x,\bar{\xi})-y]\cdot\xi})$ equals

$$\{1+4\pi^2 2^{2j}|(\Phi_\xi(x,\bar{\xi})-y)_1|^2+4\pi^2 2^j|(\Phi_\xi(x,\bar{\xi})-y)'|^2\}^N\cdot e^{2\pi i[\Phi_\xi(x,\bar{\xi})-y]\cdot\xi}.$$

We introduce this in (67), pass the differentiations onto $b_j^\nu(x,\xi)$, and note that the support of $b_j^\nu(x,\xi)$ has volume at most $c2^j\cdot 2^{j(n-1)/2}$. Thus we obtain that

$$|K_j^\nu(x,y)| \le c2^j\{1+2^j|(\Phi_\xi(x,\xi_j^\nu)-y)_1|+2^{j/2}|(\Phi_\xi(x,\xi_j^\nu)-y)'|\}^{-2N}. \tag{68}$$

Here we have replaced $\bar{\xi}$ by its original name ξ_j^ν; also $(\cdot)_1$ indicates the component in the direction ξ_j^ν, and $(\cdot)'$ denotes the orthogonal component.

In carrying out the estimate for $\int |K_j^\nu(x,y)|\,dx$, we use the majorization (68), as well as the change of variables

$$x \mapsto \Phi_\xi(x,\xi_j^\nu),$$

whose Jacobian is bounded from below, as was indicated above. The result is

$$\int |K_j^\nu(x,y)|\,dx \le A2^j\int (1+|2^j(x-y)_1|+|2^{j/2}(x-y)'|)^{-2N}\,dx$$
$$\le A2^{-j(n-1)/2},$$

if we choose N so that $2N>n$. Therefore

$$\int |K_j^\nu(x,y)|\,dx \le A2^{-j(n-1)/2}. \tag{69}$$

A similar estimate holds for $\nabla_y K_j^\nu(x,y)$, once we observe that the differentiation in y introduces factors bounded by 2^j. As a result,

$$\int |\nabla_y K_j^\nu(x,y)|\,dx \le A2^j\cdot 2^{-j(n-1)/2},$$

and so

$$\int |K_j^\nu(x,y)-K_j^\nu(x,y')|\,dx \le A|y-y'|\cdot 2^j\cdot 2^{-j(n-1)/2}. \tag{70}$$

4.7 Let us now estimate

$$\int_{^cB^*} |K_j^\nu(x,y)|\,dx$$

when B is a ball of radius δ, centered at $\bar{y}$, and $2^{-j}\le\delta$. Suppose k is the integer so that $2^{-k}<\delta\le 2^{-k+1}$. Then, as set out in §4.2, there is a unit vector ξ_k^μ, so that $|\xi_j^\nu-\xi_k^\mu|\le 2^{-k/2}$. Since $B^* = \bigcup_{2^{-\ell}\le\delta}\bigcup_\nu R_\ell^\nu$, we have the inclusion ${}^cB^*\subset {}^cR_k^\mu$. However, by (54) for $x\in {}^cR_k^\mu$,

$$2^k|\pi_k^\mu(\Phi_\xi(c,\xi_k^\mu)-\bar{y})|+2^{k/2}|\Phi_\xi(x,\xi_k^\mu)-\bar{y}|\ge\bar{c}.$$

If $y\in B$, then $|y-\bar{y}|\le 2^{-k+1}$, and since $\bar{c}$ is assumed sufficiently large, we get as a consequence

$$2^j|(\Phi(x,\xi_j^\nu)-y)_1|+2^{j/2}|(\Phi_\xi(x,\xi_j^\nu)-y)'|\ge c2^{(j-k)}$$

when $j\ge k$. Inserting this in the bound (68) and arguing as before, we obtain

$$\int_{^cB^*}|K_j^\nu(x,y)|\,dx \le A\,2^j\,2^{(j-k)}\int (1+|2^j(x-y)_1|+|2^{j/2}(x-y)'|)^{1-2N}\,dx.$$

If we choose N so that $2N-1>n$, the result is

$$\int_{^cB^*}|K_j^\nu(x,y)|\,dx \le A\cdot\frac{2^j}{\delta}\cdot 2^{-j(n-1)/2}, \qquad y\in B. \tag{71}$$

Summing the inequalities (69) to (71) in ν, and taking into account that there are essentially $2^{j(n-1)/2}$ terms involved, gives us (61)–(63), as desired.

4.8 With these majorizations for the kernels K_j, it is an easy matter to conclude our estimate of $T(a)$, for an atom a. We have

$$T(a) = \sum_{j=0}^{\infty} T_j(a) = \sum\nolimits_1 + \sum\nolimits_2,$$

where $\sum_1$ is taken over those j with $2^j \leq \delta^{-1}$, and $\sum_2$ is taken over those j with $2^j > \delta^{-1}$; here δ is the radius of the ball B associated to the atom a.

For the second sum, we use (63), which yields

$$\sum\nolimits_2 \int_{^cB^*} |T_j(a)|\,dx \leq A\Big(\int |a(y)|\,dy\Big)\cdot\Big(\sum_{2^j>\delta^{-1}} 2^{-j}\Big)\cdot\delta^{-1} \leq A.$$

For the first sum, we use (62), and write

$$T_j(a) = \int K_j(x,y)\,a(y)\,dy = \int_B [K_j(x,y) - K_j(x,\bar{y})]\,a(y)\,dy,$$

using the fact that $\int a(y)\,dy = 0$. Thus

$$\int |(T_j a)(x)|\,dx \leq A\cdot 2^j\cdot\|a\|_1\cdot\delta,$$

and

$$\sum\nolimits_1 \int_{\mathbf{R}^n} |T_j(a)|\,dx \leq A\sum_{2^j\leq\delta^{-1}} 2^j\cdot\delta \leq A.$$

Altogether then

$$\int_{^cB^*} |T(a)|\,dx \leq A,$$

which, combined with (56), gives us (49), and establishes that T maps H^1 to L^1 boundedly.

We next remark that a similar result holds for the operator T^* (the adjoint of T): it is also a bounded mapping $H^1 \to L^1$. The only changes in the argument that are needed are as follows. First, to ensure the boundedness $T^* : L^p \to L^2$ (as in §4.3), we need part (b) of the proposition in §3.4, instead of part (a). Second, taking adjoints reverses the roles of x and y in the formulae for the kernels. Thus, the definition of the rectangles R_j^ν (given by (54)) must be modified to read

$$R_j^\nu = \{x : |x - \Phi_\xi(\bar{y},\xi_j^\nu)| \leq \bar{c}2^{-j/2}, |\pi_j^\nu[x - \Phi_\xi(\bar{y},\xi_j^\nu)]| \leq \bar{c}2^{-j}\}.$$

Once these changes are made, the proof proceeds as before.

We can also invoke the duality between H^1 and BMO (see Chapter 4, §1) to assert that when $f \in L^2 \cap L^\infty$, $Tf \in$ BMO, and

$$\|Tf\|_{\mathrm{BMO}} \leq A\|f\|_{L^\infty}. \tag{72}$$

Indeed, $\langle Tf, g\rangle = \langle f, T^*g\rangle$ holds for $f \in L^2 \cap L^\infty$ and g in the dense subspace H_a^1 of H^1 (defined in Chapter 4, §1.2). From the above, it is clear that $g \mapsto \langle Tf, g\rangle$ extends to a bounded linear functional on H^1; therefore $Tf \in$ BMO, and (72) is established. A similar argument also shows that (72) is valid, if T is replaced by T^*.

4.9 Conclusion of the proof. All our statements so far have been made under the assumption that the symbol of our Fourier integral operator has order $m = -(n-1)/2$. We now pass to the case when $-(n-1)/2 < m < 0$ by complex interpolation. We will use the theorem set out in §5.2 of Chapter 4. To do this, we define the analytic family of operators T_s in the strip $0 < \mathrm{Re}(s) \leq 1$ by

$$T_s = e^{(s-\theta)^2}\int_{\mathbf{R}^n} e^{2\pi i\Phi(x,\xi)}\,a(x,\xi)\,(1+|\xi|^2)^{\gamma(s)/2}\,\hat{f}(\xi)\,d\xi,$$

where $\gamma(s) = -m - \dfrac{s(n-1)}{2}$ and $\theta = \dfrac{-2m}{n-1}$; note that $\gamma(\theta) = 0$.‡

When $\mathrm{Re}(s) = 0$, the symbol $a(x,\xi)\cdot(1+|\xi|^2)^{\gamma(s)/2}$ has order 0, and thus each T_s is bounded from L^2 to itself. Moreover, the argument used to prove this (see §3) shows that bounds for only finitely many derivatives of the symbol are involved. The bounds of the corresponding derivatives of $a(x,\xi)\cdot(1+|\xi|^2)^{\gamma(s)/2}$, with $s = it$, have at most polynomial growth in t. Given the rapid decrease, as $|t| \to \infty$, of the factor $e^{(s-\theta)^2}$, we obtain therefore

$$\|T_{it}(f)\|_{L^2} \leq A\|f\|_{L^2}, \qquad -\infty < t < \infty. \tag{73}$$

Next, when $\mathrm{Re}(s) = 1$, the symbol $a(x,\xi)\cdot(1+|\xi|^2)^{\gamma(s)/2}$ has order $-(n-1)/2$. The results proved above for H^1 and BMO then apply for these s, and again the bounds involve only finitely many derivatives of the symbol, and hence are of polynomial growth in t. As a result we have

$$\|T_{1+it}(f)\|_{\mathrm{BMO}} \leq A\|f\|_{L^\infty}, \quad f \in L^2\cap L^\infty, \quad t \in \mathbf{R}. \tag{74}$$

The application of the interpolation theorem then gives

$$\|T_\theta(f)\|_{L^p} \leq A\|f\|_{L^p}, \tag{75}$$

‡ When $n = 1$, the argument has to be modified, but is simpler: analytic interpolation is not needed. See also §6.20.

with $\theta = 1 - 2/p$. However $T_{1-2/p} = T$, and the relation between θ and p gives $1/2 - 1/p = -m/(n-1)$. Using the dual operator T^* instead of T gives the similar result when $1/p - 1/2 = -m/(n-1)$, and so our theorem is proved when $|1/2 - 1/p| = -m/(n-1)$. Note that the case $|1/2 - 1/p| < -m/(n-1)$ follows trivially: if a symbol is of order m_1, it is also of order m, whenever $m > m_1$.

The proof of the theorem is therefore concluded.

5. *Appendix: Restriction theorems in two dimensions*

5.1 For the case $n = 2$, we extend here the theorem in §1.2 to the full range of exponents.

THEOREM. *Under the assumptions of Theorem 1, when $n = 2$, we have that*

$$\|T_\lambda(f)\|_{L^q(\mathbf{R}^2)} \le A\lambda^{-2/q}\|f\|_{L^p(\mathbf{R}^1)} \tag{76}$$

where $q = 3p'$, and $1 \le p < 4$.

Notice that the limiting case $p = 4$ corresponds to $q = 4$. The argument here is restricted to two dimensions by virtue of the special role played by the exponent 4 ($= 2 \times 2$), as will be seen below. Indeed, the idea of the proof is to write

$$[T_\lambda f(\xi)]^2 = \int\int_{\mathbf{R}^2} e^{i\lambda[\Phi(x_1,\xi)+\Phi(x_2,\xi)]}\,\psi(x_1,\xi)\,\psi(x_2,\xi)\,f(x_1)\,f(x_2)\,dx_1\,dx_2 \tag{77}$$

and to consider (77) as an oscillatory integral with phase function

$$\Phi(x_1,\xi) + \Phi(x_2,\xi),$$

defined for $(x,\xi) \in \mathbf{R}^2 \times \mathbf{R}^2$.

5.2 The difficulty with this approach is that the Hessian

$$\det\left(\frac{\partial}{\partial x_i}\frac{\partial}{\partial \xi_j}[\Phi(x_1,\xi)+\Phi(x_2,\xi)]\right)$$

vanishes when $x_1 = x_2$. In fact, it equals

$$\det\begin{pmatrix}\Phi''_{\xi_1}(x_2,\xi) & \Phi'_{\xi_1}(x_2,\xi)\\ \Phi''_{\xi_2}(x_2,\xi) & \Phi'_{\xi_2}(x_2,\xi)\end{pmatrix}\cdot(x_1-x_2) + O(|x_1-x_2|^2),$$

where the primes denote differentiation with respect to x and the subscripts denote differentiation with respect to ξ. However, our condition (7) amounts to the nonvanishing of the determinant multiplying $(x_1 - x_2)$ above. Thus we are led to make a change of variables

$$(x_1, x_2) \mapsto (s_1, s_2)$$

so that

$$\det\left(\frac{\partial(s_1,s_2)}{\partial(x_1,x_2)}\right) = x_1 - x_2.$$

For this purpose, let $s_1 = x_1 + x_2$, $s_2 = x_1 \cdot x_2$; then indeed

$$\det\left(\frac{\partial(s_1,s_2)}{\partial(x_1,x_2)}\right) = x_1 - x_2.$$

Moreover, set $\tilde\Phi(s,\xi) = \Phi(x_1,\xi) + \Phi(x_2,\xi)$; then $\tilde\Phi$ is smooth in (s,ξ) (the smoothness in s results from the fact that $\Phi(x_1,\xi) + \Phi(x_2,\xi)$ is symmetric in x_1 and x_2), and

$$\det\left(\frac{\partial^2\tilde\Phi(s,\xi)}{\partial s_i \partial \xi_j}\right) \ne 0,$$

when x_1 is near x_2. Note that $ds_1\,ds_2 = |x_1 - x_2|\,dx_1\,dx_2$; thus

$$[T_\lambda f(\xi)]^2 = 2\int_{\mathbf{R}^2} e^{i\lambda\tilde\Phi(s,\xi)}\,\tilde\psi(s,\xi)\,F(s)\,ds \tag{78}$$

where $\tilde\psi(s,\xi) = \phi(x_1,\xi)\cdot\phi(x_2,\xi)$, and

$$F(s) = \frac{f(x_1)\cdot f(x_2)}{|x_1 - x_2|};$$

the multiplication by 2 occurs because the mapping from the x-space to the s-space is two-to-one.

5.3 To (78) we apply the following variant of the proposition in §1.1. If $G(\xi)$ is defined by

$$G(\xi) = \int_{\mathbf{R}^2} e^{i\lambda\tilde\Phi(s,\xi)}\,\tilde\psi(s,\xi)\,F(s)\,ds,$$

then

$$\|G\|_{L^{r'}(\mathbf{R}^2)} \le A\lambda^{-2/r'}\|F\|_{L^r(\mathbf{R}^2)} \tag{79}$$

whenever $1 \le r \le 2$, with $1/r' + 1/r = 1$. In fact, when $r = 2$, (79) is just (3) when $n = 2$. Also, when $r = 1$, the inequality (79) is trivial. The result for $1 \le r \le 2$ then follows by interpolation.

5.4 To conclude the proof of the theorem, set $2r' = q$, then $\|G\|_{L^{r'}(\mathbf{R}^2)} = \|T_\lambda f\|_q^2$. Moreover,

$$\begin{aligned}\|F\|^r_{L^r(\mathbf{R}^2)} &= \int |F(s)|^r\,ds_1\,ds_2\\ &= \int |f(x_1)\cdot f(x_2)|^r|x_1-x_2|^{-r}\,ds_1\,ds_2\\ &= \frac{1}{2}\int |f(x_1)|^r|f(x_2)|^r|x_1-x_2|^{1-r}\,dx_1\,dx_2.\end{aligned}$$

However, by the theorem of fractional integration (see (31) in Chapter 8)

$$\left|\int g(x_1)\, g(x_2) |x_1 - x_2|^{-\gamma}\, dx_1\, dx_2\right| \leq \|g\|_s^2,$$

if $1 - 1/s = 1/s - 1 + \gamma$. Setting $g = |f|^r$, $\gamma = r - 1$, we have $2 - 2/s = r - 1$. Also $\|g\|_s = \|f\|_{L^p}^r$, where $p = sr$. With $\gamma = r - 1$, the limitation $\gamma < 1$ is equivalent to $r < 2$. Note that

$$\frac{3}{q} = \frac{3}{2r'} = \frac{3r-3}{2r} = 1 - \frac{1}{p},$$

since $(3-r)/2 = 1/s$, $p = 2r/(3-r)$. Therefore $3/p = 1/p'$, and the restriction $1 \leq r < 2$ is equivalent to $1 \leq p < 4$. The theorem is therefore proved.

5.5 The corollaries dealing with restrictions of Fourier transforms and Bochner-Riesz summability can now be established in two dimensions for the full range of exponents.

COROLLARY 1. *Suppose $S \subset \mathbf{R}^2$ is a curve whose curvature is nowhere zero and S_0 is a compact subset of S. Then*

$$\left(\int_{S_0} |\hat{f}(\xi)|^q\, d\sigma(\xi)\right)^{1/q} \leq A(S_0) \cdot \|f\|_{L^p(\mathbf{R}^2)}, \qquad f \in \mathcal{S},$$

whenever $3q = p'$, $1 \leq p < 4/3$.

COROLLARY 2. *The operator S^δ, defined by (28), extends to a bounded operator from $L^p(\mathbf{R}^2)$ to itself, for $4/3 \leq p \leq 4$, whenever $\delta > 0$; and more generally to the range*

$$\frac{4}{3+2\delta} < p < \frac{4}{1-2\delta},$$

whenever $0 < \delta \leq 1/2$.

These results are proved in the same way as the corresponding propositions in §2.1 and §2.2 are deduced from Theorem 1.

6. Further results

A. Oscillatory integrals

6.1 We remark that estimates for oscillatory integrals of the second kind (such as given §1.1) for

$$(T_\lambda f)(\xi) = \int_{\mathbf{R}^n} e^{i\lambda\Phi(x,\xi)}\, \psi(x,\xi)\, f(x)\, dx$$

imply corresponding decay estimates for

$$I(\lambda) = \int_{\mathbf{R}^n} e^{i\lambda\phi(x)}\, \psi_0(x)\, dx.$$

The simplest example of this is as follows. Assume that $\left|\det\left(\frac{\partial^2\phi}{\partial x_i \partial x_j}\right)\right| \geq 1$ on the support of ψ_0. Then one can conclude that $I(\lambda) = O(\lambda^{-n/2})$ as $\lambda \to \infty$.[†]

To prove the assertion, let $\Phi(x,\xi) = \phi(x-\xi)$ and choose $\psi(x,\xi) = \psi_0(x-\xi)\,\psi_1(x)$, for an appropriate $\psi_1 \in C_0^\infty(\mathbf{R}^n)$. One then verifies that

$$\frac{\partial}{\partial \xi_j} T_\lambda = T_\lambda \frac{\partial}{\partial \xi_j} + T'_\lambda,$$

where T'_λ has the same phase function as T_λ. Repeated applications of this, and the fact that

$$|F(0)| \leq c \sum_{|\alpha| \leq k} \|\partial_\xi^\alpha F(\xi)\|_{L^2},$$

with $k > n/2$, allow us to prove our conclusion as a consequence of the proposition in §1.1.

6.2 Pursuing the analogy between oscillatory integrals of the first and second kinds, we also have the following.

(a) In analogy with §5.11(a) of the previous chapter, we can obtain the decay $O(\lambda^{-n/2})$, even when the Hessian of Φ vanishes, by inserting an appropriate mitigating factor in the definition of T_λ. Thus define

$$(T'_\lambda f)(\xi) = \int_{\mathbf{R}^n} e^{i\lambda\Phi(x,\xi)}\, |h(x,\xi)|^\mu\, \psi(x,\xi)\, f(x)\, dx,$$

where $h(x,\xi) = \det\left(\frac{\partial^2\Phi}{\partial x_i \partial \xi_j}\right)$. Then

$$\|T'_\lambda(f)\|_{L^2(\mathbf{R}^n)} \leq A\lambda^{-n/2}\, \|f\|_{L^2(\mathbf{R}^n)}, \qquad \text{as } \lambda \to \infty,$$

as long as $\mu \geq 5n/2$.

(b) A variant of (a) arises when we replace the determinant of the matrix $\left(\frac{\partial^2\Phi}{\partial x_i \partial \xi_j}\right)$ by its Hilbert-Schmidt norm $\left\|\frac{\partial^2\Phi}{\partial x \partial \xi}\right\| = \left(\sum_{i,j=1}^n \left|\frac{\partial^2\Phi}{\partial x_i \partial \xi_j}\right|^2\right)^{1/2}$. Then if

$$(T''_\lambda f)(\xi) = \int_{\mathbf{R}^n} e^{i\lambda\Phi(x,\xi)} \left\|\frac{\partial^2\Phi}{\partial x \partial \xi}\right\|^\mu \psi(x,\xi)\, f(x)\, dx,$$

one has

$$\|T''_\lambda(f)\|_{L^2(\mathbf{R}^n)} \leq A\lambda^{-1/2}\, \|f\|_{L^2(\mathbf{R}^n)}, \qquad \text{as } \lambda \to \infty,$$

whenever $\mu \geq 2n + 1/2$.

[†] Compare with §1.2 and §2.3 of Chapter 8 for the cases where $n = 1$, and where $n > 1$ with ψ supported in a small neighborhood of a critical point of ϕ.

(c) As a consequence of (b), one can prove an analogue of the proposition in §2.2 of the previous chapter. Suppose some derivative $\dfrac{\partial^2\Phi}{\partial x_i \partial\xi_j}$ does not vanish to infinite order at any point in the support of ψ. Then if

$$(T_\lambda f)(\xi) = \int_{\mathbf{R}^n} e^{i\lambda\Phi(x,\xi)}\, \psi(x,\xi)\, f(x)\, dx,$$

the norm of T_λ is $O(\lambda^{-\varepsilon})$ as $\lambda \to \infty$, for some $\varepsilon > 0$.

For (a), see Sogge and Stein [1986]. Part (b) can be proved by similar techniques. To prove (c) note that, under the hypotheses, we have

$$\int_{\mathbf{R}^n}\int_{\mathbf{R}^n} \left\| \frac{\partial^2\Phi}{\partial x \partial \xi} \right\|^{-2\delta} |\psi(x,\xi)|^2\, dx\, d\xi < \infty$$

for some $\delta > 0$; as a result

$$f \mapsto \int_{\mathbf{R}^n} e^{i\lambda\Phi(x,\xi)} \left\| \frac{\partial^2\Phi}{\partial x\partial\xi} \right\|^{-\delta} \psi(x,\xi)\, f(x)\, dx$$

is bounded uniformly in λ. An interpolation with (a) then gives the desired result.

6.3 In several applications, the oscillatory integrals above arise in combination with kernels of singular integral operators. To be more specific, let T be a bounded operator from $L^2(\mathbf{R}^n)$ to itself that is representable by a distribution kernel K: $(Tf)(x) = \int K(x,y)\, f(y)\, dy$ for $f \in \mathcal{S}$, where K satisfies the estimates $|\partial_y^\beta \partial_x^\alpha K(x,y)| \le A|x-y|^{-n-|\alpha|-|\beta|}$ (see, e.g., the discussion in Chapter 7 of such operators). Let $\Phi(x,y)$ be a real smooth phase function, let $\psi \in C_0^\infty(\mathbf{R}^n \times \mathbf{R}^n)$, and assume that $\det\left(\dfrac{\partial^2\Phi}{\partial x_i \partial y_j}\right)$ has no zeros on the support of ψ. Consider the operator

$$(T_\lambda f)(x) = \int_{\mathbf{R}^n} e^{i\lambda\Phi(x,y)}\, K(x,y)\, \psi(x,y)\, f(y)\, dy,$$

(defined by $\langle T_\lambda f, g\rangle = \int\int e^{i\lambda\Phi(x,y)}\, K(x,y)\, \psi(x,y)\, f(y)\bar{g}(x)\, dy\, dx$). Then the norm of T_λ (as an operator on L^2) remains bounded as $\lambda \to \infty$.

Phong and Stein [1986a]. The result in Chapter 12, §5.1, where $\Phi(x,y)$ is a real bilinear form on $\mathbf{R}^n \times \mathbf{R}^n$, can be viewed as a model case of this theorem.

6.4 There are a number of variants of the conclusions in §6.3.

(a) Let K be as in §6.3, and let $P(x,y)$ be any real polynomial on $\mathbf{R}^n \times \mathbf{R}^n$ of degree d. Then

$$(Tf)(x) = \int_{\mathbf{R}^n} e^{iP(x,y)}\, K(x,y)\, f(y)\, dy$$

is bounded from L^2 to itself; the bound for T may be chosen to depend only on K and d and not on the coefficients of P.

(b) The conclusion stated in §6.3 holds when the hypothesis

$$\det\left(\frac{\partial^2\Phi}{\partial x_i \partial y_j}\right) \neq 0$$

is replaced by the weaker assumption that, at each point, there is an i and a j so that $\dfrac{\partial^2\Phi}{\partial x_i\partial y_j}$ does not vanish to infinite order there.

(c) The operators in (a) and (b) are also bounded (with the same uniformities) from $L^p(\mathbf{R}^n)$ to itself, for $1 < p < \infty$.

For (a), see Ricci and Stein [1987]. Note that, in the translation-invariant case, this result is essentially contained in §5.23(a) of the previous chapter. The theorem in (b) is due to Pan [1991a]; there is a close relation between this result and the statement §6.2(c) above. In connection with the L^p boundedness of these operators, it should be mentioned that there are also limiting results for $p = 1$; these can be formulated in terms of appropriate variants of the Hardy spaces H^1 or weak-type $(1,1)$ results. For the statements and proofs see Phong and Stein [1986a], Stein [1986], Ricci and Stein [1987], Chanillo and Christ [1987], Pan [1991b]; a particular example is discussed in §6.22 below.

6.5 The estimate (8) in Theorem 1 cannot be improved beyond the range $1 \le p \le 2$, when $n \ge 3$.

To see this, let $n = 3$. Then there is an appropriate Φ and a bounded f having compact support so that

$$\|T_\lambda f\|_q \ge c\lambda^{-1/2-1/q}, \qquad \text{as } \lambda \to \infty.$$

This is consistent with (8) only if $q \ge 4$ (i.e., $p \le 2$).

Indeed, for $x \in \mathbf{R}^2$ and $\xi = (\xi', \xi_3) \in \mathbf{R}^3$ (here $\xi' \in \mathbf{R}^2$), take

$$\Phi(x,\xi) = x\cdot\xi' + \frac{1}{2}\langle A(\xi_3)x, x\rangle$$

where, for each ξ_3, $A(\xi_3)$ is a real, symmetric 2×2 matrix, depending smoothly on ξ_3. We now impose two conditions on $\xi_3 \mapsto A(\xi_3)$.

The first is that $\dfrac{dA(\xi_3)}{d\xi_3}$ is invertible for each ξ_3. This condition guarantees the basic hypothesis (7) of Theorem 1. The second requirement is that $\operatorname{rank}(A(\xi_3)) \equiv 1$ for all ξ_3. It is easy to check that these two conditions are compatible, and that indeed there are smooth functions $\xi_3 \mapsto A(\xi_3)$ that satisfy both simultaneously. Now let $f(x) = 1$ on the support of ψ. Then

$$(T_\lambda f)(\xi) = \int_{\mathbf{R}^2} e^{i\lambda\Phi(x,\xi)}\, \psi(x)\, dx.$$

Let $S = \{\xi = (\xi', \xi_3) : \xi' \in \operatorname{range}(A(\xi_3))\}$. In view of our assumptions on $\operatorname{rank}(A(\xi_3))$, we see that S is a smooth hypersurface. Note that if $\xi \in S$, the quadratic function $x \mapsto \Phi(x,\xi)$ has a critical point, and moreover the rank of $\dfrac{\partial^2\Phi(x,\xi)}{\partial x_i \partial x_j}$ is exactly 1. Thus if $\xi \in S$, we can show by the method of stationary phase that

$$|(T_\lambda f)(\xi)| \approx \lambda^{-1/2} \qquad \text{as } \lambda \to \infty.$$

The estimate also holds in a tubular neighborhood of S whose radius is a small multiple of λ^{-1}. The result is that

$$\|T_\lambda f\|_{L^q} \geq c\lambda^{-1/2}\lambda^{-1/q},$$

and the result is proved.

See Bourgain [1991b], where it is also shown that for a certain class of Φ one does have

$$\|T_\lambda f\|_{L^q} \leq c\lambda^{-n/q}\|f\|_{L^\infty},$$

for some q, with $q < 2(n+1)/(n-1)$.

B. Restriction theorems and Bochner-Riesz summability

6.6 When $n = 2$, the sharp (L^p, L^q) restriction theorem in §5.5 (for curves with nowhere vanishing curvature) can be extended to curves in the plane that have finite type (in the sense of Chapter 8, §3.2). In fact, if $S \subset \mathbf{R}^2$ is such a curve, having type k, and S_0 is a fixed compact subset of S, then

$$\|\widehat{f}\|_{L^{q,p}(S_0,d\sigma)} \leq A\|f\|_{L^p(\mathbf{R}^2)},$$

where $(k+1)q = p'$, $1 \leq p < 4/3$, and $q > 1$.[‡] Note that $L^{q,p} \subset L^q$ only when $p \leq q$. Thus, strictly speaking, we get an (L^p, L^q) result only when, in addition, we have $p \leq (k+2)/(k+1)$.

See Sjölin [1974], Sogge [1987a].

6.7 There is another approach to the proof of Bochner-Riesz summability in $\mathbf{R}^2$ for $4/3 \leq p \leq 4$, as stated in §5.5, Corollary 2. It depends on different ideas that are interesting in their own right, the more so since, as it has turned out, their thrust can be adapted to a variety of other situations. We now outline the argument involved, allowing ourselves some slight oversimplifications.

Fix a smooth bump function ϕ on $\mathbf{R}^+$, supported in $[1/2, 1]$, and let S_k denote the multiplier operator

$$\widehat{S_k f}(\xi) = \phi(2^k(1-|\xi|)) \cdot \widehat{f}(\xi).$$

Note that the multiplier is supported in the shell $2^{-k-1} \leq 1-|\xi| \leq 2^{-k}$. Since $S^\delta(f)$ is essentially $\sum_k 2^{-k\delta} S_k(f)$, it suffices to prove

$$\|S_k(f)\|_{L^4} \leq A_\varepsilon 2^{\varepsilon k}\|f\|_{L^4}, \qquad \text{for every } \varepsilon > 0. \tag{$*$}$$

To do this, one utilizes the following three ideas.

‡ Here $L^{q,p}$ is a Lorentz space; a definition can be found in Chapter 6, §7.10.

(i) Let $\mathcal{R}^k$ denote the collection of all rectangles, centered at the origin, having side lengths 2^k and $2^{k/2}$, whose orientation is arbitrary. For any $R \in \mathcal{R}^k$, let $(A_R f)(x) = |R|^{-1}\int_R f(x-y)\,dy$. Then

$$\|\sup_{R\in\mathcal{R}^k} |(A_R f)(x)|\,\|_{L^2(\mathbf{R}^2)} \leq A_\varepsilon 2^{\varepsilon k}\|f\|_{L^2(\mathbf{R}^2)}.$$

(ii) A consequence of (i) is the following vector-valued inequality for L^4:

$$\Big\|\Big(\sum_j |A_{R_j}(f_j)|^2\Big)^{1/2}\Big\|_{L^4(\mathbf{R}^2)} \leq A_\varepsilon 2^{\varepsilon k}\Big\|\Big(\sum_j |f_j|^2\Big)^{1/2}\Big\|_{L^4(\mathbf{R}^2)},$$

where $\{R_j\}_{j=1}^\infty$ is any collection of rectangles chosen from $\mathcal{R}^k$.

(iii) An additional (easy) square function inequality is needed. Let P_j denote the "projection" given as a multiplier operator by

$$\widehat{P_j f}(\xi) = \chi(\xi_2 - j)\cdot\widehat{f}(\xi),$$

where χ is a smooth bump function that is scaled to the length $2^{-k/2}$, i.e., $\chi(u) = \chi_0(2^{k/2}u)$. Then

$$\Big\|\Big(\sum_j |P_j f|^2\Big)^{1/2}\Big\|_{L^p(\mathbf{R}^2)} \leq A\|f\|_{L^p(\mathbf{R}^2)}, \qquad 2 \leq p \leq \infty,$$

with A independent of k.

We postpone momentarily a discussion of the proofs of assertions (i)-(iii), and show first how they imply $(*)$. We may assume that the spectrum of f is contained in the sector $|\arg(\xi)| \leq \pi/4$, and we divide the shell $2^{-k-1} \leq 1-|\xi| \leq 2^{-k}$ (which has thickness 2^{-k-1}) into disjoint sectors of angle width $2^{-k/2}$; this is akin to the "second" dyadic decomposition in §4.4. Indeed, if we consider $S_k P_j$, we see that the support of its multiplier is basically a rectangle r_j containing one of these sectors. Each rectangle r_j has side lengths essentially 2^{-k} and $2^{-k/2}$ in the ξ_1 and ξ_2 directions, respectively. Now if we choose χ appropriately, we have

$$\sum_j P_j = I \qquad \text{and} \qquad S_k = \sum_j S_k P_j = \sum_j P_j S_k.$$

The key observation is that

$$\Big\|\sum_j P_j S_k f\Big\|_{L^4}^4 \leq \int_{\mathbf{R}^2}\Big(\sum_j |P_j S_k f(x)|^2\Big)^2 dx,$$

which is based on the geometrical fact that, for each $\xi \in \mathbf{R}^2$, there are at most a fixed number of pairs (j, j') with $\xi \in r_j + r_{j'}$. Next, let R_j be the dual rectangle to r_j (so R_j has dimensions $\approx 2^k \times 2^{k/2}$). Since the multiplier of $P_j S_k$ is supported in r_j, one can prove that (essentially)

$$|(P_j S_k f)(x)| \leq c[A_{R_j}(|P_j f|)](x).$$

Now invoke (ii), from which it follows that

$$\|S_k f\|_{L^4} \leq c2^{\varepsilon k}\Big\|\Big(\sum_j |P_j f(x)|^2\Big)^{1/2}\Big\|_{L^4}.$$

Then applying (iii) yields $(*)$.

For the proof of (i), see §3.11 in the next chapter. The assertion (ii) is a simple consequence. Indeed,

$$\|(\sum_j |(A_{R_j} f_j)(x)|^2)^{1/2}\|_{L^4}^2 = \sup_{\|\omega\|_{L^2}=1} \sum_j \int_{\mathbf{R}^2} |(A_{R_j} f_j)(x)|^2\, \omega(x)\, dx.$$

However, $|A_{R_j} f_j|^2 \le A_{R_j}(|f_j|^2)$ and

$$\int [A_{R_j}(|f_j|^2)](x)\, \omega(x)\, dx = \int |f_j(x)|^2\, (A_{R_j}\omega)(x)\, dx.$$

Summing in j and using (i) then gives (ii) via Schwarz's inequality.

To prove (iii), note that $\sum_{j\in\mathbf{Z}} e^{ij\theta}\chi(\xi_2 - j)$ is the Fourier transform of a measure $d\mu_\theta$ on $\mathbf{R}^2$ whose total variation is bounded independently of θ and the scaling factor $2^{k/2}$. Write $T_\theta f = f * d\mu_\theta$. Then

$$\frac{1}{2\pi}\int_0^{2\pi} |(T_\theta f)(x)|^2\, d\theta = \sum_j |(P_j f)(x)|^2.$$

The result follows because, when $p \ge 2$, the inequality $\|T_\theta f\|_{L^p} \le A\|f\|_{L^p}$ implies

$$\left\|\left(\frac{1}{2\pi}\int_0^{2\pi} |(T_\theta f)(x)|^2\, d\theta\right)^{1/2}\right\|_{L^p(dx)} \le A\|f(x)\|_{L^p(dx)}.$$

The initial ideas leading to the above proof are in C. Fefferman [1973b], and were developed by Córdoba [1977], [1979], [1982]. Several variants and applications are in §6.8 below and in Chapter 11, §4.11–§4.13.

6.8 We consider the almost everywhere and dominated convergence of the Bochner-Riesz means when $n > 1$; the classical case of one dimension is different and is discussed in §6.10 below. As in §2.2, we set $S^\delta(f) = f * K^\delta$, and

$$S_R^\delta f = f * K_{R^{-1}}^\delta = \int_{|\xi|<R} \hat{f}(\xi)\,(1 - |\xi|^2/R^2)^\delta\, e^{2\pi i x\cdot\xi}\, d\xi,$$

for suitable f.

By well-known arguments, the convergence of $S_R^\delta(f)$ to f (as $R \to \infty$) in L^p norm is equivalent to the L^p boundedness of the operator S^δ, and thus holds under the conditions given by the proposition in §2.2 (when $n \ge 3$) and by Corollary 2 in §5.4 (when $n = 2$). For the almost everywhere convergence, one considers the maximal operator S_*^δ, defined by

$$(S_*^\delta f)(x) = \sup_{R>0} |(S_R^\delta f)(x)|.$$

As opposed to norm convergence, where duality between L^p and $L^{p'}$ is operative, the problem of dominated convergence splits into two seemingly different questions: the case $p \ge 2$, and the case $p < 2$.

In the first case, the results and methods used are close to those used for norm convergence. In particular, for $n = 2$, we can assert that

$$\|S_*^\delta(f)\|_{L^p} \le A\|f\|_{L^p}, \qquad \text{for } 2 \le p \le 4, \qquad \text{if } \delta > 0,$$

and, more generally, that this inequality holds in the range $2 \le p < \infty$, whenever $p < 4(1-2\delta)^{-1}$ and $0 < \delta \le 1/2$.

To prove this, for which the key case is $p = 4$, one utilizes square functions.† Consider the operator

$$(G^\alpha f)(x) = \left(\int_0^\infty |(S_R^{\alpha+1} f)(x) - (S_R^\alpha f)(x)|^2 \frac{dR}{R}\right)^{1/2}.$$

Then the proof of the result can be reduced to the inequality

$$\|G^\alpha f\|_{L^4(\mathbf{R}^2)} \le A_\alpha \|f\|_{L^4(\mathbf{R}^2)}, \qquad \text{if } \alpha > 1/2,$$

and this in turn is equivalent to the estimate

$$\|s_k f\|_{L^4(\mathbf{R}^2)} \le A_\varepsilon 2^{-k/2} 2^{\varepsilon k}\|f\|_{L^4(\mathbf{R}^2)}, \qquad \text{for each } \varepsilon > 0 \text{ and } k \ge 1, \qquad (**)$$

where $(s_k f)(x) = (\int_0^\infty |f * \Phi_t^{(k)}|^2 \frac{dt}{t})^{1/2}$ and

$$\widehat{\Phi^{(k)}}(\xi) = \phi(2^k(1 - |\xi|)),$$

with ϕ as in §6.7. Note that $\|s_k f\|_{L^2} \approx 2^{-k/2}\|f\|_{L^2}$ because $\int_0^\infty |\widehat{\Phi^{(k)}}(t\xi)|^2 \frac{dt}{t} \approx 2^{-k}$. The proof of $(**)$ has many points in common with the proof of $(*)$ in §6.7. Further details are in Carbery [1983]; see also the refinement in Carbery [1985].

Several additional remarks are in order. First, a partial analogue of this result (when $p \ge 2$) for $n \ge 3$ is in §6.9 below. Second, in the case $p \le 2$, the best maximal result that is known at present is the assertion

$$\|S_*^\delta f\|_{L^p} \le A\|f\|_{L^p}, \qquad \text{for } \delta > (n-1)[1/p - 1/2];$$

see *Fourier Analysis*, Chapter 7. It would be interesting to determine whether, in fact, the inequality is valid for the larger ranges

$$\frac{4}{3+2\delta} < p < 2, \qquad \text{when } 0 < \delta < \frac{1}{2} \text{ and } n = 2,$$

and

$$\frac{2n}{n+1+2\delta} < p < 2, \qquad \text{for } 0 < \delta < \frac{n-1}{2} \text{ and } n \ge 3.$$

A result in this direction, for the "lacunary" version of S_*^δ, is in Córdoba and Lopez-Melero [1981], and Igari [1981].

---

† For background on the use of square functions in the context of summability, see *Fourier Analysis*, Chapter 7, §5.

6.9 (i) The following relation between the (L^p, L^2) restriction property and the boundedness of S^δ and S^δ_* holds in $\mathbf{R}^n$, $n \geq 2$. If p_0 is such that the (L^{p_0}, L^2) restriction (for the sphere $\mathbf{S}^{n-1} \subset \mathbf{R}^n$) is valid, then $S^\delta : L^p \to L^p$ is bounded in the corresponding range, i.e., when $1 \leq p \leq p_0$ and for optimal δ, $\dfrac{2n}{n+1+2\delta} < p < \dfrac{2n}{n-1-2\delta}$. The fact that we can take $p_0 = \dfrac{2n+2}{n+3}$ (see §2.1) gives the range $\delta > \dfrac{n-1}{2(n+1)}$.

(ii) For $p \geq 2$, the result (i) can be strengthened to state that

$$\int_{\mathbf{R}^n} |(S^\delta f)(x)|^2 \psi(x)\, dx \leq A \int_{\mathbf{R}^n} |f|^2 \{[M(\psi^r)](x)\}^{1/r}\, dx,$$

where $\psi \geq 0$, $r = \dfrac{n+1}{2}$, $\delta > \dfrac{n-1}{2(n+1)}$, and M is the usual Hardy-Littlewood maximal operator.

(iii) A similar inequality holds with S^δ replaced by the maximal operator S^δ_* (or the square function $G^{\delta-1/2}$) discussed in §6.8. As a result,

$$\|S^\delta f\|_{L^p(\mathbf{R}^n)} \leq A\|f\|_{L^p(\mathbf{R}^n)}, \qquad \text{if } p_0' \leq p < \frac{2n}{n-1-2\delta} \leq \infty.$$

To prove (i), we take our cue from §6.7 and the decomposition $S^\delta = \sum 2^{-k\delta} S_k$ used there. This suggests that we decompose the kernel K^δ of S^δ as $K^\delta = \sum_{k=0}^{\infty} K_k$, with $K_k(x) = K^\delta(x) \cdot \eta(2^{-k}x)$, $k \geq 1$, where η is an appropriate *radial* C^∞ function, supported in the annulus $1/4 \leq |x| \leq 1$.

It will suffice to see that

$$\|f * K_k\|_{L^{p_0}(\mathbf{R}^n)} \leq A 2^{-\varepsilon k} \|f\|_{L^{p_0}(\mathbf{R}^n)}, \tag{$***$}$$

for some $\varepsilon > 0$, as long as $\dfrac{2n}{n+1+2\delta} < p_0$.

Now the multiplier m_k corresponding to K_k is given by

$$m_k(\xi) = 2^{nk}[(1-|\xi|^2)^\delta_+ * \widehat{\eta}](2^k \xi).$$

From this, it is easy to see that m_k is radial and, in view of the smoothness of $(1-|\xi|^2)^\delta_+$ away from $|\xi| = 1$ (using also that it has compact support), we have the estimate

$$|m_k(\xi)| \leq A_N 2^{-Nk}(1+|\xi|)^{-N}, \qquad \text{all } N \geq 0,$$

whenever $|1-|\xi|| \geq 1/2$.

Next, since the kernel K_k is supported in the ball of radius 2^k, it suffices to prove $(***)$ for functions supported in this ball.[‡] Supposing f to be so supported, we have that

$$\|f * K_k\|_{p_0} \leq C 2^{nk[1/p_0 - 1/2]} \|f * K_k\|_{L^2}.$$

[‡] For this point see, e.g., the argument invoked immediately following (32), or §3.2 in Chapter 6.

Now

$$\|f * K_k\|^2_{L^2} = \int_{|1-|\xi||\leq 1/2} |\widehat{f}(\xi)|^2 |m_k(\xi)|^2\, d\xi + \int_{|1-|\xi||>1/2} .$$

The second integral is easily treated by the decay properties of $m_k(\xi)$ for $|1-|\xi|| > 1/2$ described above. For the first integral, we use the restriction theorem and the radiality of m_k. Thus

$$\begin{aligned} &\int_{|1-|\xi||\leq 1/2} |\widehat{f}(\xi)|^2 |m_k(\xi)|^2\, d\xi \\ &\leq \sup_{1/2<r<3/2} \int_{|\xi|=1} |\widehat{f}(r\xi)|^2\, d\sigma(\xi) \times \int_{1/2\leq r\leq 3/2} |m_k(r)|^2 r^{n-1}\, dr \\ &\leq c\|f\|^2_{L^{p_0}(\mathbf{R}^n)} \|K_k\|^2_{L^2} \leq c\|f\|^2_{L^{p_0}(\mathbf{R}^n)} \cdot 2^{-k(1+2\delta)}, \end{aligned}$$

since $|K^\delta(x)| \leq A|x|^{-\frac{n+1}{2}-\delta}$, by (29). This establishes $(***)$ with $\varepsilon = \delta + 1/2 - n[1/p_0 - 1/2]$.

The original proof that an (L^p, L^2) restriction theorem gives the corresponding boundedness of S^δ is in C. Fefferman [1970]. The argument above is due to the author and may be found in C. Fefferman [1973b]. The proofs of (ii) and (iii) are based on some similar ideas; see Christ [1985b].

6.9[a] When $n \geq 3$, the restriction theorem for spheres and the boundedness of S^δ have recently been extended beyond the ranges given in §2.1 and §2.2. In particular, when $n = 3$, the result given there for $p \geq 4$ has been shown to hold for $p \geq 4 - \varepsilon$, with $\varepsilon = 8/75$.

Bourgain [1991a]; related results are in §6.5 and in §3.1[a] of the next chapter.

6.10 For the case $n = 1$, the results of almost everywhere and dominated summability (more precisely, convergence) are contained in a theorem that originated in the work of Carleson [1966] and was modified and extended by Hunt [1968] and Sjölin [1971]. The theorem in question can be formulated for any number of dimensions, but it applies to the problem of summability only when $n = 1$.

Suppose K_0 is a Calderón-Zygmund kernel (e.g., of the type described in Chapter 1, §8.19); assume that K_0 is homogeneous of degree $-n$, is smooth away from the origin, and has integral zero over any sphere centered at the origin. Let

$$(T_\lambda f)(x) = \text{p.v.} \int_{\mathbf{R}^n} e^{2\pi i \lambda y \cdot x} K_0(y) f(x-y)\, dy,$$

and set $(T_* f)(x) = \sup_{\lambda \in \mathbf{R}^n} |(T_\lambda f)(x)|$. Then we have

$$\|T_*(f)\|_{L^p} \leq A_p \|f\|_{L^p}, \qquad 1 < p < \infty.$$

This result applies to S_R^δ, $\delta = 0$, in one dimension, since in this case

$$(S_R^0 f)(x) = \int_{-R}^{R} e^{2\pi i x\cdot\xi}\,\widehat{f}(\xi)\,d\xi.$$

If we take $K_0(x) = 1/\pi x$ to be the kernel of the Hilbert transform, we see that $S_R^0 = (T_R - T_{-R})/2i$.

C. Fourier integral operators

6.11 We indicate briefly some further aspects of Fourier integral operators.

The basic notion of Fourier integral operators we treated, i.e., those of the form

$$(Tf)(x) = \int_{\mathbf{R}^n} e^{2\pi i\Phi(x,\xi)}\,a(x,\xi)\,\widehat{f}(\xi)\,d\xi,$$

can be extended as follows. Consider first a "symbol", that is, a function $b(x,y,\tau)$ defined for $(x,y,\tau)\in\mathbf{R}^n\times\mathbf{R}^n\times\mathbf{R^N}$ (here N need not equal n); we assume that it has order μ, in the sense that

$$|\partial_{x,y}^\beta\partial_\tau^\alpha b(x,y,\tau)| \le A_{\alpha,\beta}(1+|\tau|)^{\mu-|\alpha|},$$

and we suppose that b has compact support in x and y. Next, consider a "phase" function $\phi(x,y,\tau)$ that is real-valued, homogeneous of degree 1 in τ, and smooth when $\tau\ne 0$. We suppose that it satisfies the basic nondegeneracy condition that the $(N+n)\times(N+n)$ matrix of second derivatives

$$\begin{pmatrix}\phi''_{\tau\tau} & \phi''_{x\tau}\\ \phi''_{\tau y} & \phi''_{xy}\end{pmatrix}$$

is invertible on $\mathrm{supp}(b(x,y,\tau))$. The generalization of our T is then an operator $\tilde{T}$ of the form

$$(\tilde{T}f)(x) = \int_{\mathbf{R}^n}\int_{\mathbf{R^N}} e^{2\pi i\phi(x,y,\tau)}\,b(x,y,\tau)\,f(y)\,dy\,d\tau,$$

when the integral is suitably defined. Observe that T is essentially a special case of $\tilde{T}$ that arises when $N=n$ and $\phi(x,y,\tau)=\Phi(x,\tau)-y\cdot\tau$; the nondegeneracy condition required of ϕ is then equivalent to (36).

What is crucial for these operators are the (local) canonical transformations from $T^*(\mathbf{R}^n)\backslash\{0\}$ to itself[†] determined by Φ and ϕ. For the first, it is the mapping $(\nabla_\xi\Phi,\xi)\mapsto(x,\nabla_x\Phi)$, while for the second, it is the mapping $(y,-\nabla_y\phi)\mapsto(x,\nabla_x\phi)$ defined on the set where $\nabla_\tau\phi=0$. Indeed, it is a basic fact that, essentially, operators of the type $\tilde{T}$ can be written as finite sums of operators of the kind T, associated to the same canonical transformations. In this reduction, the symbols a have order $m=\mu-\dfrac{n-N}{2}$, where μ is the order of the symbol b.

† Here $T^*(\mathbf{R}^n)$ is the cotangent space of $\mathbf{R}^n$.

The above, and the requirements that one should be able to compose such operators, take adjoints, etc., leads one to re-cast their definition with homogeneous canonical relations as the starting point. Important stages in the development of the theory are the works of Lax [1957], Maslov [1965], Hörmander [1968], Egorov [1969], and Eskin [1970]; see also the systematic treatment in Hörmander [1971], as well as the expositions in Taylor [1981], Treves [1982], M. Beals, C. Fefferman, R. Grossman [1983], and Sogge [1993].

Two other aspects should be mentioned. First, the use of Fourier integral operators to conjugate pseudo-differential operators into simpler forms; examples where this idea is applied may be found in Egorov [1969], Nirenberg and Treves [1971], in addition to the just-cited references. A second aspect, which represents the original motivation for the subject, is the possibility of expressing the solution of the initial-value problem for a strictly hyperbolic equation using these operators. This is treated in §6.12.

6.12 Let $P(x,t;\partial_x,\partial_t)$ be a partial differential operator of order m in $\mathbf{R}^{n+1}$, with $x\in\mathbf{R}^n$, $t\in\mathbf{R}$. We write $p(x,t;\xi,\tau)$ for the top-order part of the symbol of P; it is a homogeneous polynomial (in ξ and τ) of degree m. We assume that P is *strictly hyperbolic* in t, in the sense that, for each (x,t,ξ) with $\xi\ne0$, the polynomial $p(x,t;\xi,\tau)$ has m real distinct roots $\tau_1(x,\xi,t),\dots,\tau_m(x,\xi,t)$.

For small t, one solves the "eikonal" equation

$$\frac{\partial}{\partial t}\Phi_j(x,\xi,t)=\tau_j(x,\nabla_x\Phi_j,t),\qquad\text{with }\Phi_j(x,\xi,t)|_{t=0}=x\cdot\xi,$$

for $j=1,\dots,m$.

Then if the symbols $a_j^k(x,\xi,t)$, which depend smoothly on t, are suitably chosen, with $a_j^k\in S^{-k}$, one can assert the following: The solution of the equation $P(x,t;\partial_x,\partial_t)u(x,t)=0$, with initial condition $\dfrac{\partial^k}{\partial t^k}u(x,t)|_{t=0}=f_k(x)$, is given for small t by

$$u(x,t)=\sum_{\substack{1\le j\le m\\0\le k\le m-1}} T_{j,k}(f_k),$$

modulo an infinitely smoothing operator. Here $T_{j,k}$ is a Fourier integral operator, depending smoothly on the parameter t, defined by

$$(T_{j,k}f)(x)=\int_{\mathbf{R}^n} e^{2\pi i\Phi_j(x,\xi,t)}\,a_j^k(x,\xi,t)\,\widehat{f}(\xi)\,d\xi.$$

It should be remarked that when $m=2$ (that is, when $P(u)=0$ gives a variable-coefficient wave equation), then one also has the important property that

$$\mathrm{rank}(\nabla_\xi\nabla_\xi\Phi_j(x,\xi,t))=n-1$$

for t small, $t\ne0$.

For further details, see the references in §6.11 and Duistermaat [1973].

6.13 The fact that various L^p estimates for are sharp can be seen as a consequence of the following assertions.

(a) Suppose $\gamma(\xi)$ is a smooth function that equals $|\xi|^{-\alpha}$ for large ξ. Then the distribution

$$f_\alpha = \int_{\mathbf{R}^n} e^{2\pi i|\xi|} e^{2\pi i x\cdot\xi}\,\gamma(\xi)\,d\xi$$

equals a function $f_\alpha(x)$ when $|x| \neq 1$; this function is smooth there and, moreover,

$$|f_\alpha(x)| \approx |1-|x||^{\alpha-n/2-1/2} \qquad \text{as } |x| \to 1,$$

as long as $\alpha < (n+1)/2$.

(b) With $\gamma(\xi)$ as above, but now $\alpha < n$, the distribution

$$g_\alpha = \int_{\mathbf{R}^n} e^{2\pi i x\cdot\xi}\,\gamma(\xi)\,d\xi$$

equals a function $g_\alpha(x)$ for $x \neq 0$. Moreover, for appropriate c_α, $g_\alpha(x) - c_\alpha|x|^{\alpha-n}$ is everywhere smooth, and $g_\alpha(x)$ is rapidly decreasing at infinity.

(c) As a result, suppose T is a Fourier integral operator (of the type (35)) of order 0, with phase $\Phi(x,\xi) = |\xi| + x\cdot\xi$, and whose symbol $a(x,\xi)$ equals $a_1(x)$, where $a_1 \in C_0^\infty(\mathbf{R}^n)$ and $a_1(x) \neq 0$ for $|x| = 1$. Then T is not bounded from $L^p(\mathbf{R}^n)$ to itself, if $p \neq 2$. More generally, suppose T is the Fourier integral operator of order m, $(1-n)/2 \leq m \leq 0$, having the same phase but with symbol $a(x,\xi) = a_1(x)\cdot\gamma(\xi)$, where $\alpha = -m$. Then T is not bounded from $L^p(\mathbf{R}^n)$ to itself for $\left|\frac{1}{2} - \frac{1}{p}\right| > \frac{-m}{n-1}$.

To show that $g_\alpha(x) - c_\alpha|x|^{\alpha-n}$ is smooth, note that it is the Fourier transform of a compactly supported distribution in view of the formula for the Fourier transform of $|x|^{\alpha-n}$.‡ The fact that $g_\alpha(x)$ is rapidly decreasing at infinity follows by repeated integration by parts.

To prove (a), we observe that, because of (25) in Chapter 8 (or §6.19 below), f_α is essentially

$$2\pi|x|^{(2-n)/2}\int_1^\infty \gamma(u)\,e^{2\pi i u}\,J_{(n-2)/2}(2\pi|x|u)\,u^{n/2}\,du.$$

Thus, by the asymptotic formula (15) in Chapter 8, the main contribution to the above integral is

$$c|x|^{(1-n)/2}\int_1^\infty e^{2\pi i u(1-|x|)}\,u^{n/2-1/2-\alpha}\,du,$$

from which the assertion follows.

To prove (c), let us first consider the case $p < 2$. Let $f = Tg$, where T has symbol $a_1(x)\cdot\gamma(\xi)$, with $\gamma(\xi) = |\xi|^m$ for large ξ. If $g = g_\alpha$, then $f = a_1(x) f_{\alpha-m}$. For a given m, $(1-n)/2 < m \leq 0$, if $1/2 - 1/p > -m/(n-1)$, we see that if α is chosen so that $(\alpha - m - n/2 - 1)p = -1$, then $(\alpha - n)p < n$. Thus $f \notin L^p$, while $g \in L^p$. The proof for $p < 2$ follows from this, once we observe that the dual of T is essentially its complex conjugate.

‡ See, e.g., *Fourier Analysis*, Chapter 4, §4.

6.14 Let $\Phi(x,\xi)$ be a phase function that satisfies the usual restriction $\det\{\Phi_{x,\xi}\} \neq 0$, together with the special assumption that $\operatorname{rank}\{\Phi_{\xi\xi}\} = n-1$ when $\xi \neq 0$. Let $K(x,y)$ denote the kernel of the corresponding Fourier integral operator

$$K(x,y) = \int_{\mathbf{R}^n} e^{2\pi i(\Phi(x,\xi) - y\cdot\xi)}\,a(x,\xi)\,d\xi,$$

where we assume that the symbol a belongs to $S^{-(n+1)/2}$. Then for each y, the function $K(\cdot,y)$ (when properly defined) belongs to $\mathrm{BMO}(\mathbf{R}^n)$, uniformly in y. The same is true when the roles of x and y are reversed.

To prove this assertion, suppose that $y = 0$, and write

$$K(x) = \int_{\mathbf{R}^n} e^{2\pi i\Phi(x,\xi)}\,a(x,\xi)\,d\xi.$$

Let S denote the singular support of K, which is

$$S = \{x : \Phi_\xi(x,\xi) = 0 \text{ for some } \xi \neq 0\}.$$

Then by our assumption on $\operatorname{rank}\{\Phi_{\xi\xi}\}$, S is a smooth $(n-1)$ dimensional manifold.

We claim next that $|\nabla K(x)| \leq c[d(x)]^{-1}$ for $x \notin S$; here d denotes the distance from S. Indeed, passing to polar coordinates $\xi = r\theta$, $\theta \in \mathbf{S}^{n-1}$, we can write

$$\nabla K(x) = \int_0^\infty \int_{\mathbf{S}^{n-1}} e^{2\pi i r\Phi(x,\theta)} a_1(x,\theta,r)\,r^{n-1}\,dr\,d\sigma(\theta)$$

with $a_1 \in S^{(1-n)/2}$, modulo a smooth error term. Now the integral

$$\int_{\mathbf{S}^{n-1}} e^{2\pi i r\Phi(x,\theta)} a_1(x,\theta,r)\,d\sigma(\theta)$$

can be treated by the method of stationary phase, since $\theta \mapsto \Phi(x,\theta)$, as a function on $\mathbf{S}^{n-1}$, has a nonvanishing Hessian at every point. This gives an asymptotic expression for the integral over $\mathbf{S}^{n-1}$ whose main term is

$$\frac{e^{2\pi i r\Phi(x,\theta')}\,a_1(x,\theta',r)}{c(x,\theta')}\cdot r^{(1-n)/2},$$

where $\theta' = \theta'(x)$ is the critical point of $\Phi(x,\theta)$ when x is near S. Note that since $\det\{\Phi_{x,\xi}\} \neq 0$, we get that $|\Phi(x,\theta')| \approx d(x)$, so the estimate $|\nabla K(x)| \leq c/d(x)$ then follows from the fact that

$$\left|\int_1^\infty e^{iru}\,b(r)\,dr\right| \leq \frac{A}{|u|},$$

if $\left|\dfrac{\partial^k}{\partial r^k} b(r)\right| \leq c_k r^{-k}$, when $k = 0,1,2$.

One proves in the same way that $|K(x) - K(x')| \leq c$, if x and x' are any pair of points with $|x - x'| \approx d(x) \approx d(x')$. Finally one observes that since S is smooth, these inequalities imply that $K \in \mathrm{BMO}$.

6.15 The fractional integration theorem of Hardy-Littlewood and Sobolev can be restated to assert that a pseudo-differential operator of order $-\alpha$, $0 < \alpha < n$, maps $L^p(\mathbf{R}^n)$ to $L^q(\mathbf{R}^n)$, if $1/p - 1/q = \alpha/n$ and $1 < p, q < \infty$ (see Chapter 8, §4.2). We consider now the generalization of this result for a Fourier integral operator T (as given by (35)), whose symbol a has order $m = -\alpha$.

(a) When $p \leq 2 \leq q$, $T : L^p(\mathbf{R}^n) \to L^q(\mathbf{R}^n)$ is bounded for p and q satisfying the same relation as in the theorem of fractional integration, that is, $1/p - 1/q = \alpha/n$ and $1 < p, q < \infty$.

(b) When $1 < p \leq q \leq 2$, $T : L^p \to L^q$ when $\frac{n}{p} - \frac{1}{q} = \alpha + \frac{n-1}{2}$. A similar (dual) result holds in the range $2 \leq p \leq q < \infty$.

(c) In both cases (a) and (b) the conclusions extend to $p = 1$, if the space $L^1(\mathbf{R}^n)$ is replaced by the Hardy space $H^1(\mathbf{R}^n)$.

The assertion (a) is already essentially contained in §3.1.4. One uses the decomposition $T = T_0 T_1$ arising in the proposition there, but where T_0 has order $-\alpha_0$ and T_1 has order $-\alpha_1$, with $\frac{\alpha_0}{n} = \frac{1}{2} - \frac{1}{q}$ and $\frac{\alpha_1}{n} = \frac{1}{p} - \frac{1}{2}$. To prove (b) one interpolates, using the technique in §4.9, between the (L^p, L^p) conclusion given by Theorem 2 and the fact that if T has order $-n/2$ then $T : H^1 \to L^2$ (for this last point, see also Chapter 3, §5.21).

6.16 We continue our discussion of the analogue of the fractional integration theorem for Fourier integral operators, but here we make the special assumption that $\operatorname{rank}(\Phi_{\xi\xi}) = n - 1$; here Φ is the phase of T. This hypothesis is particularly relevant when considering estimates for second-order hyperbolic equations (see §6.12). In this case, if

$$(Tf)(x) = \int_{\mathbf{R}^n} e^{2\pi i \Phi(x,\xi)}\, a(x,\xi)\, \widehat{f}(\xi)\, d\xi$$

and $a \in S^{-\alpha}$, we can obtain a stronger conclusion than §6.15 for the range $p \leq 2 \leq q$, but no improvement when $1 < p < q \leq 2$ or $2 \leq p < q < \infty$.

The unified way of stating this is as follows. The operator T, whose symbol has order $-\alpha$, is bounded from $L^p(\mathbf{R}^n)$ to $L^q(\mathbf{R}^n)$ if $1 < p, q < \infty$ and

$$\alpha + \frac{n-1}{2} = \frac{n}{p} - \frac{1}{q}, \qquad \text{when } q \leq p',$$
$$\alpha - \frac{n-1}{2} = \frac{1}{p} - \frac{n}{q}, \qquad \text{when } q \geq p';$$

here p' is the exponent conjugate to p.

Again, the conclusion extends to $p = 1$ if we replace $L^1(\mathbf{R}^n)$ by the Hardy space $H^1(\mathbf{R}^n)$.

The above assertions are proved by interpolating (as in §4.9) between the (L^p, L^p) estimates given by Theorem 2 and the fact that, for our particular phase, the operator maps H^1 to L^∞ when $\alpha = (n+1)/2$. This last fact is a consequence of §6.14 and the duality of H^1 and BMO.

It should also be pointed out that the results stated here and in §6.15 are sharp in the sense that, for a given α and p, the exponent q cannot be increased. This observation is a consequence of §6.13.

There are many papers applying L^p estimates for Fourier integral operators to hyperbolic equations that are connected with the results in §6.15, §6.16, and §6.17; see, e.g., Brenner [1977], M. Beals [1982], Kapitanskii [1990], Seeger, Sogge, and Stein [1991].

6.17 Let T be a Fourier integral operator of order m, as in (35).

(a) $T : L^p_k(\mathbf{R}^n) \to L^p_{k'}(\mathbf{R}^n)$ if $1 < p < \infty$ and $k' = k - m - (n-1)\cdot|1/2 - 1/p|$.

(b) $T : \Lambda_\gamma(\mathbf{R}^n) \to \Lambda_{\gamma'}(\mathbf{R}^n)$ where $\gamma' = \gamma - m - (n-1)/2$, if both γ and γ' are positive.

For (a), the essential conclusion is the special case $k = k' = 0$ given by Theorem 2. The proof of (b) is implicit in the proof of that theorem. Further details are in Seeger, Sogge, and Stein [1991].

D. Miscellaneous topics

6.18 The common distributions arising in the theory of Fourier integral operators are the *Lagrangian distributions* (Hörmander [1971]). It turns out that each such distribution belongs locally to some Hardy space H^p. More precisely, suppose $\phi(x,\tau)$ is a real phase function defined for $(x,\tau) \in \mathbf{R}^n \times \mathbf{R}^N$ that is homogeneous of degree 1 in τ, and smooth when $\tau \neq 0$. Assume ϕ is non-degenerate, in that $\nabla\phi = \nabla_{x,\tau}\phi \neq 0$, and that the gradients $\nabla\frac{\partial\phi}{\partial\tau_1}, \dots, \nabla\frac{\partial\phi}{\partial\tau_N}$ are linearly independent on the set where $\nabla_\tau\phi(x,\tau) = 0$, $\tau \neq 0$. Suppose $a(x,\tau)$ is a symbol, in the sense that

$$|\partial_x^\beta \partial_\tau^\alpha a(x,\tau)| \leq A_{\alpha,\beta}(1 + |\tau|)^{m-|\alpha|}$$

for some m, and suppose that $a(x,\tau)$ has compact support in x. Then the distribution u, given by

$$u = \int_{\mathbf{R}^N} e^{i\phi(x,\tau)}\, a(x,\tau)\, d\tau,$$

belongs to $H^p_{\text{loc}}(\mathbf{R}^n)$ for some $p > 0$; the distribution u is defined via the linear functional

$$\psi \mapsto \int_{\mathbf{R}^n \times \mathbf{R}^N} e^{i\phi(x,\tau)}\, a(x,\tau)\, \psi(x)\, dx\, d\tau, \qquad \psi \in \mathcal{S};$$

the integral can be made to converge, using integration by parts and the assumption $\nabla\phi \neq 0$.

To prove our assertion, let S be the "singular set" of u, defined by

$$S = \{x \in \mathbf{R}^n : \nabla_\tau\phi(x,\tau) = 0 \text{ for some } \tau \neq 0\}.$$

Let $\tilde{S} = \{(x,\tau) \in \mathbf{R}^n \times \mathbf{R}^N : \nabla_\tau\phi(x,\tau) = 0, |\tau| = 1\}$. By our assumptions on ϕ, it follows that $\tilde{S}$ is a smooth submanifold of $\mathbf{R}^n \times \mathbf{R}^N$ of dimension $n-1$. Since S is the projection of $\tilde{S}$ on $\mathbf{R}^n$, it can be seen by a simple covering argument that $|S_\delta| \leq c\delta$ as $\delta \to 0$; here S_δ denotes the set of points whose distance from S is at most δ. Moreover, if $x \notin S$, then again by our assumptions on ϕ, we have that $|\nabla_\tau\phi(x,\tau)| \geq c d(x)$, where $d(x)$ is the distance of x from S. Thus an integration by parts shows that $|u(x)| \leq A d(x)^{-M}$ with $M > \max(0, N+m)$. If we now invoke §5.18 of Chapter 3, we get the desired conclusion.

6.19 The Fourier transform of a radial function is radial. The following identities are related to this fact.

(a) If $f(x) = f_0(|x|)$ is a radial function on $\mathbf{R}^n$, its Fourier transform is given by

$$\widehat{f}(\xi) = 2\pi|\xi|^{(2-n)/2} \int_0^\infty J_{(n-2)/2}(2\pi|\xi|r)\, f_0(r)\, r^{n/2}\, dr;$$

here $J_{n-2/2}$ is a Bessel function (as defined in Chapter 8, §1.4.1).

(b) If K^δ is the kernel given by

$$K^\delta(x) = \int_{|\xi|\le 1} (1-|\xi|^2)^\delta\, e^{2\pi i x\cdot\xi}\, d\xi,$$

then $K^\delta(x) = \pi^{-\delta}\Gamma(1+\delta)|x|^{-\delta-n/2} J_{\delta+n/2}(2\pi|x|)$.

(c) As a consequence, if $n \ge 2$ and $p \ge 2n/(n+1)$ (or if $n=1$ and $p>1$), there can be no (L^p, L^q) restriction theorem for the unit sphere.

(d) It follows from (b) that S^δ is not bounded from $L^p(\mathbf{R}^n)$ to itself, if

$$p \le \frac{2n}{n+1+2\delta}, \qquad \text{or} \quad p \ge \frac{2n}{n-1-2\delta}.$$

The statement (a) is merely a reformulation of the identity (25) of the previous chapter. The assertion (b) then follows from this and the identity

$$J_{\mu+\nu+1}(t) = \frac{t^{\nu+1}}{2^\nu\Gamma(\nu+1)} \int_0^1 J_\mu(ts)\,(1-s^2)^\nu\, ds,$$

which in turn is a consequence of formula (16) in Chapter 8 and the Beta integral identity

$$u^{\delta+\beta} = \frac{\Gamma(\delta+\beta+1)}{\Gamma(\delta+1)\Gamma(\beta)} \int_0^u (u-s)^{\beta-1}\, s^\delta\, ds.$$

For (a) and (b), see also Chapter 4 of *Fourier Analysis.*

Assertion (c) follows because $J_{(n-2)/2}(2\pi r)\cdot r^{n/2} \notin L^{p'}(r^{n-1}\,dr)$ if $p \ge 2n/(n+1)$; here $1/p'+1/p=1$. To prove (d), note that if f is the characteristic function of a small ball centered at the origin, then because of (b) and (29), $|(S^\delta f)(x)| \ge c|x|^{-\delta-(n+1)/2}$ (on the average) for large x, and thus $|S^\delta(f)| \notin L^p$, if $p \le 2n/(n+1+2\delta)$. The result for $p \ge 2n/(n-1-2\delta)$ follows by duality. Statement (d) goes back to Herz [1954].

6.20 In one dimension, Fourier integral operators are essentially trivial modifications of pseudo-differential operators. More precisely, let $\Phi_\pm(x) = \Phi(x,\pm 1)$. Then the condition (36) is the statement that $d\Phi_\pm/dx \ne 0$, so the mappings $\Phi_\pm$ are local diffeomorphisms of $\mathbf{R}^1$. Note that $\Phi(x,\xi) = |\xi|\Phi_\pm(x)$ with the sign determined by sign(ξ). Therefore we see that the operator T defined by (35) can be written as

$$(Tf)(x) = (S_+f)[\Phi_+(x)] + (S_-f)[\Phi_-(x)] + (Ef)(x).$$

Here $S_\pm$ are pseudo-differential operators with symbols $a_\pm$, where $a_+(x,\xi) = a(x,\xi)$ if $1 \le \xi$ and $a_+(x,\xi) = 0$ if $\xi \le -1$; the symbol a_- has the symmetric property. The error operator E has a kernel in $C_0^\infty(\mathbf{R}^1\times\mathbf{R}^1)$.

6.21 The restriction and summability theorems proved in §2.1 and §2.2 can be extended to the setting where $\mathbf{R}^n$ is replaced by a compact n-dimensional Riemannian manifold M. Let Δ denote the Laplace-Beltrami operator on M, and let P_{λ_j} be the orthogonal projection on the eigenspace of Δ corresponding to the eigenvalue $-\lambda_j$, $\lambda_j \ge 0$.

(a) The analogue of the (L^p, L^2) restriction theorem is the norm estimate

$$\Big\| \sum_{\lambda<\lambda_j^{1/2}<\lambda+1} P_{\lambda_j} f \Big\|_{L^2(M)} \le A(1+\lambda)^{\delta(p)} \|f\|_{L^p(M)}$$

for $1 \le p \le \dfrac{2(n+1)}{n+3}$ and $\delta(p) = n\left|\dfrac{1}{p}-\dfrac{1}{2}\right| - \dfrac{1}{2}$.

(b) The analogue of the proposition in §2.2 is the uniform estimate

$$\|S_R^\delta f\|_{L^p(M)} \le A\|f\|_{L^p(M)}$$

with A independent of R, as long as $\delta > \delta(p)$ and $1 \le p \le \dfrac{2(n+1)}{n+3}$ or $\dfrac{2(n+1)}{n-1} \le p \le \infty$. Here

$$S_R^\delta f = \sum_{0\le\lambda_j<R} \left(1-\frac{\lambda_j}{R}\right)^\delta P_{\lambda_j} f.$$

Sogge [1987b], [1988]. Further related results may be found in Christ and Sogge [1988], Seeger and Sogge [1989].

Note that if we take M to be the n-torus $\mathbf{T^n}$, then, re-scaling the assertions in (a) and (b) and passing to the limit, we are led back to the results in §2.1 and §2.2.

6.22 As we have seen, the method of oscillatory integrals and the related use of T^*T are quite useful in proving L^2 and L^p estimates. It turns out, moreover, that these ideas can also be utilized in the limiting case of proving weak-type $(1,1)$ estimates. We illustrate this in a simple example.

Consider in $\mathbf{R}^1$ the operator $Tf = f * K$, where $K = \text{p.v.}\, e^{ix^2}/x$. This operator is bounded on $L^2(\mathbf{R}^n)$; the proof is contained essentially in Chapter 8, §5.23. One can prove that T is of weak-type $(1,1)$ by the following scheme.

For $\alpha > 0$, let $f = g + \sum b_j$ be a Calderón-Zygmund decomposition of f at height α (such as in Chapter 1, §4), where b_j is supported in the interval I_j. Let I_j^* denote an expansion of I_j by a large fixed factor. Disregarding the b_j for which the corresponding I_j has length at most 1 (for which the argument is a little simpler), it suffices to see that for the remaining I_j we have

$$\int_{^c\cup I_j^*} |\Sigma(Tb_j)(x)|^2\, dx \le A\alpha\|f\|_{L^1}.$$

Now let $K_\lambda(x) = \chi(x/\lambda)\cdot e^{ix^2}/x$, where $\chi(u) = 1$ if $|u| \geq 2$, and $\chi(u) = 0$ if $|u| \leq 1$. If λ_j is the length of I_j, then $T(b_j) = T_{\lambda_j}(b_j)$ outside I_j^*; thus it suffices to prove that

$$\|\sum T_{\lambda_j}(b_j)\|_{L^2}^2 \leq A\alpha \|f\|_{L^1}.$$

However, the left side is dominated by $\sum_{j,k} |\langle T_{\lambda_j}^* T_{\lambda_k} b_k, b_j\rangle|$, and for the kernel of $T_\lambda^* T_\mu$ one can obtain the bound

$$\min(\lambda^{-1}, \mu^{-1}, |x-y|^{-2}), \qquad \lambda, \mu \geq 1.$$

From this it is not difficult to see that $\sum_{j,k} |\langle T_{\lambda_j}^* T_{\lambda_k} b_k, b_j\rangle| \leq A\alpha\|f\|_{L^1}$, establishing the assertion.

Arguments of this kind originated in C. Fefferman [1970]. The particular result whose proof is sketched here is due to Chanillo, Kurtz, and Sampson [1986]; see also Chanillo and Christ [1987].

Notes

§1, §2, and §5. The sharp restriction theorem in two dimensions (Corollary 1 in the appendix) is in C. Fefferman [1970] and grew out of a discussion with the author; see also Zygmund [1974]. The n-dimensional (L^p, L^2) restriction theorem for the sphere was proved by the author in 1967 (unpublished) for $1 \leq p < 4n/(3n+1)$; for $1 \leq p < (2n+2)/(n+3)$ by Tomas [1975]; and then in the same year by the author (unpublished) for $1 \leq p \leq (2n+2)/(n+3)$. The last argument has been used in several related contexts by Strichartz [1977] and Greenleaf [1981]. It is elaborated in Stein [1986] to prove Theorem 1.

Preceding some of these developments about restriction theorems was the important conclusion obtained in C. Fefferman [1970]: Whenever an (L^p, L^2) restriction theorem holds for $\mathbf{S}^{n-1}$, the "best" Bochner-Riesz assertion also holds for this p. This was followed by the result of Carleson and Sjölin [1972], which gave the sharp summability result in two dimensions for $\delta > 0$, $4/3 \leq p \leq 4$ (Corollary 2 in the appendix). Shortly thereafter, C. Fefferman [1971b] obtained the counterexample for $\delta = 0, p \neq 2$; see §2.5 in the next chapter. The common formulation of the restriction theorem and the summability theorem (as in the theorem in the appendix) is due to Hörmander [1973]; the proposition in §1.1 also originates there. Another exposition of the theory of Bochner-Riesz means (and several related topics) may be found in K. M. Davis and Y.-C. Chang [1987].

§3 and §4. Theorem 2 is in Seeger, Sogge, and Stein [1991]; for earlier versions, corresponding essentially to the case when rank $\Phi_{\xi\xi} \equiv n-1$, see Stein [1971], Peral [1980], Miyachi [1980b], M. Beals [1982]. The general form of the theorem requires the second dyadic decomposition, which is implicit in C. Fefferman [1973b]. For the background on Fourier integral operators, see the references cited in §6.11.

CHAPTER X

Maximal Operators: Some Examples

The general question of the differentiability of integrals in $\mathbf{R}^n$ represents one of the main issues in real-variable theory. One broad formulation of the problem is as follows. For what collections $\mathcal{C}$ of sets $\{C\}$, is it true that for "all" f

$$\lim_{\operatorname{diam}(C)\to 0} \frac{1}{|C|}\int_C f(x-y)\,dy = f(x) \quad \text{a.e. ?}$$

Closely related to this is the problem of the boundedness on L^p of the corresponding maximal operator $M_{\mathcal{C}}$, defined by

$$(M_{\mathcal{C}} f)(x) = \sup_{C\in\mathcal{C}} \frac{1}{|C|}\int_C |f(x-y)|\,dy.$$

Let us first observe that if we make the special assumption that the sets in $\mathcal{C}$ have *bounded eccentricity*, in the sense that the ratio (in volume) between the smallest ball containing C to the largest ball contained in C is uniformly bounded over $C \in \mathcal{C}$, then this question has (essentially) already been resolved in §3.1 of Chapter 2.[†] Thus the main problem that needs to be considered is when the sets are allowed to be "thin", that is, when their eccentricities may be unbounded. It is fair to state that, although many collections $\mathcal{C}$ of this kind present themselves naturally, at present one can decide the question only in very specific circumstances.

An example illustrates the nature and difficulty of the general problem. In $\mathbf{R}^2$, we take $\mathcal{C}$ to be the three-parameter collection of all rectangles centered at the origin. Then $M_{\mathcal{C}}$ is not bounded on any L^p, $p < \infty$, and the almost-everywhere limit conclusion may fail, even for characteristic functions of sets of finite measure. On the other hand, if the collection $\mathcal{C}$ is restricted to centered rectangles whose major axes point in a fixed direction, or in an infinite number of suitable "lacunary" directions, then the conclusion holds for all L^p, $p > 1$, but fails completely when $p = 1$.

What lies behind these negative results can best be understood in terms of a fundamental construction of Besicovitch which yields a set

[†] Of course, if $\mathcal{C}$ is the collection of all balls centered at the origin, then $M_{\mathcal{C}}$ is the usual maximal operator M, treated in Chapter 1.

of measure zero that contains line segments in all possible directions. As such, it begins to reveal the possible complexity of sets in two or more dimensions. Inherent in this situation (and in distinction to one dimension) is the geometric fact that a continuum of "singular" directions come into play.

We begin our exposition by giving a detailed description of the Besicovitch set. Next, we present a fundamental maximal principle that is relevant here and is also useful in other contexts. It states (in a precise sense) that almost-everywhere results imply corresponding (weak-type) maximal inequalities. An application of this principle to the Besicovitch set then gives the negative solution to the problem of differentiation with respect to rectangles of arbitrary orientation. Incidentally, the Besicovitch set can also be used in other circumstances; among these are: proving the failure of "unrestricted" convergence (taken in its widest sense) for Poisson integrals on symmetric spaces (in particular, for the tube domain corresponding to a circular cone in $\mathbf{R}^3$) and a construction which shows that the characteristic function of the ball in $\mathbf{R}^n$, $n \geq 2$, gives a Fourier multiplier that is unbounded on L^p, for all $p \neq 2$.

Having visited the counterexamples and difficulties that exist, it is then our task to begin the road back. We shall see that, indeed, there are a number of interesting positive results that can be obtained. This can be achieved if we supplement the technique of covering lemmas with the use of an implicit orthogonality. Here square functions turn out to play a key role. In §3 below, we sketch the proof, using these methods, of the L^p boundedness of the maximal operator associated to a family of rectangles having prescribed lacunary directions. The important situation when the sets in $\mathcal{C}$ have some curvature, for which oscillatory integrals are needed, will be the subject of the next chapter.

1. The Besicovitch set

Shortly after its initial construction, it was noted that the Besicovitch set could be used to give a solution to the Kakeya "needle problem". The question was to find a plane set of least area in which a segment of unit length could be moved so that it pointed in all possible directions. While this historical aspect has remained something of a curiosity, the Besicovitch set has come to play an increasingly significant role in real-variable theory and Fourier analysis. Indeed, our accumulated experience allows us to regard the structure of this set as, in many ways, representative of the complexities of two-dimensional sets, in the same sense that Cantor-like sets already display some of the typical features that arise in the one-dimensional case.

Our presentation of the Besicovitch set will be in terms of a union of a large number of congruent thin rectangles in the plane, which have a

high degree of overlap. Each rectangle will have side lengths 1 and 2^{-N}, where N is a large fixed integer. Given such a rectangle R, we define its *reach* $\tilde{R}$, to be the rectangle obtained by translating R two units, in the positive direction, along the longer side of R. The relation between R and $\tilde{R}$ is displayed in Figure 1.

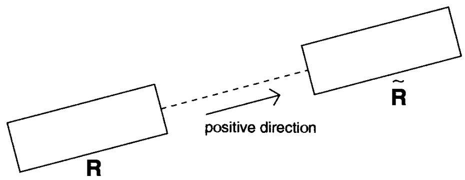

Figure 1. The reach $\tilde{R}$ of the rectangle R.

THEOREM 1. *Given any $\varepsilon > 0$, there exists an integer $N = N_\varepsilon$, and 2^N rectangles $R_1, \ldots, R_{2^N}$, each having side lengths 1 and 2^{-N}, so that*

(i) $\left|\bigcup_{j=1}^{2^N} R_j\right| < \varepsilon$, *and*

(ii) *The $\tilde{R}_j$ are mutually disjoint, $j = 1, \ldots, 2^N$, and so*

$$\left|\bigcup_{j=1}^{2^N} \tilde{R}_j\right| = 1.$$

Here $\tilde{R}_j$ denotes the "reach" of R_j, as defined above.

1.1 While the properties of the Besicovitch set used below are most easily stated in terms of rectangles, for the actual construction it seems necessary to work with triangles. More precisely, one proceeds as follows. An initial triangle is partitioned into a large number of smaller triangles, obtained by equally dividing the base of the original triangle. The main point then is that these subtriangles can be translated so that their union has small measure.

The translation procedure used arises as a series of iterations of a basic procedure, which we will now describe.

1.1.2 We start with a triangle T and construct from it a figure $\Phi(T)$, which is the union of translates of two subtriangles of T. To specify the construction, we fix a constant of proportionality α with $1/2 < \alpha < 1$. Suppose that T is the triangle ABC, whose base AB lies along the x-axis. We bisect the base AB at M, obtaining thereby two subtriangles: the "left" triangle AMC and the "right" triangle MBC (Figure 2).

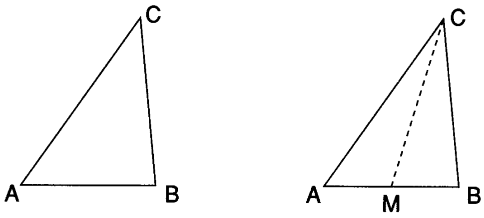

Figure 2. Bisecting the triangle T.

Now we translate the "right" triangle leftwards to obtain the overlapping figure $\Phi(T)$ (Figure 3).

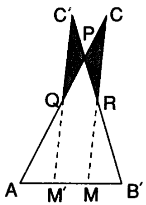

Figure 3. The overlapping figure $\Phi(T)$.

Here $M'B'C'$ is the translate of MBC; the left triangle AMC has remained fixed. The figure $\Phi(T)$ is the union of two parts: the smaller triangle $\Phi_h(T) = AB'P$ (which is similar to the original triangle ABC), and the union $\Phi_a(T)$ of the two small shaded triangles.

We shall call $\Phi_h(T)$ the "heart" of the figure $\Phi(T)$ and $\Phi_a(T)$ the "arms" of $\Phi(T)$. The constant α is the side-length ratio between the (similar) heart triangle $\Phi_h(T)$ and original triangle T. It determines the figure $\Phi(T)$, once the triangle T is given. We observe first that

$$|\Phi_h(T)| = \alpha^2 \cdot |T|. \tag{1}$$

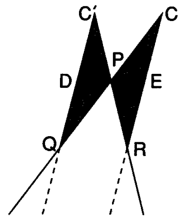

Figure 4. The "arms" $\Phi_a(T)$ of $\Phi(T)$.

Next, we fix our attention on $\Phi_a(T)$ (Figure 4). We draw the line DE parallel to the base AB' that passes through the intersection point P. It divides $\Phi_a(T)$ into four triangles. A moment's reflection shows that the triangle $C'DP$ is similar to the "right" triangle $C'M'B'$, with ratio $1-\alpha$. Also, PER is congruent to the triangle $C'DP$. Moreover, the triangle PEC is similar to the "left" triangle AMC, with ratio $1-\alpha$; also the triangle PEC is congruent to PDQ. Altogether then:

$$|\Phi_a(T)| = 2(1-\alpha)^2 \cdot |T|. \tag{2}$$

Therefore

$$|\Phi(T)| = [\alpha^2 + 2(1-\alpha)^2] \cdot |T|. \tag{3}$$

We will use this process to generate our monster, which will have a tiny heart and many arms.

1.1.2 We now describe the result of an "n-fold iteration" of this construction. Starting from the fixed triangle ABC, we subdivide the base AB into 2^n equal subintervals, with division points

$$A = A_0, A_1, \ldots, A_{2^n} = B.$$

In this way, our original triangle is divided into 2^{n-1} disjoint smaller triangles, namely

$$A_{2j}A_{2j+2}C, \qquad 0 \le j < 2^{n-1};$$

the base of such a triangle has midpoint A_{2j+1}.

Now with α fixed throughout, we construct the figure $\Phi(A_{2j}A_{2j+2}C)$ as above, for each of the 2^{n-1} triangles. In so doing, we obtain 2^{n-1} "hearts" and also 2^{n-1} pairs of "arms", in the above terminology. By our construction, the right side of $\Phi_h(A_{2j}A_{2j+2}C)$ is parallel to the line CA_{2j+2}, as is the left side of $\Phi_h(A_{2j+2}A_{2j+4}C)$ (here $0 \le j < 2^{n-1}-1$). Thus the triangle $A_{2j+2}A_{2j+4}C$ can be moved leftwards so that the left side of $\Phi_h(A_{2j+2}A_{2j+4}C)$ coincides with the right side of $\Phi_h(A_{2j}A_{2j+2}C)$. Carrying out such a translation for all the triangles $A_{2j}A_{2j+2}C$, $0 < j < 2^{n-1}$, we can incorporate each of these 2^{n-1} hearts into one composite heart, which is similar to the original triangle ABC.

To summarize, we have translated the 2^n subtriangles of ABC, forming a figure that we call $\Psi_1(ABC)$. As stated above, it contains a "heart", namely the disjoint union of the translates of the

$$\Phi_h(A_{2j}A_{2j+2}C).$$

The rest of $\Psi_1(ABC)$ consists of the union of the translated

$$\Phi_a(A_{2j}A_{2j+2}C),$$

which we refer to as the "arms" of $\Psi_1(ABC)$. Observe that there can be considerable overlap among these arms, although we will not take advantage of this.

Since $|A_{2j}A_{2j+2}C| = 2^{-n+1}|ABC|$, (1) gives us that

$$|\text{heart of } \Psi_1(ABC)| = \sum_{j=0}^{2^{n-1}-1} |\Phi_h(A_{2j}A_{2j+2}C)| = \alpha^2|ABC|.$$

Also

$$|\text{arms of } \Psi_1(ABC)| \le \sum |\Phi_a(A_{2j}A_{2j+2}C)| = 2(1-\alpha)^2 \cdot |ABC|,$$

by (2). Thus

$$|\Psi_1(ABC)| \le [\alpha^2 + 2(1-\alpha)^2] \cdot |ABC|. \tag{4}$$

We now iterate the construction leading to Ψ_1 as follows. The heart of $\Psi_1(ABC)$ is given to us as the union of 2^{n-1} triangles, and so we carry out the above process on the heart of $\Psi_1(ABC)$ with n replaced by $n-1$; then we re-translate all 2^n of the original triangles $A_jA_{j+1}C$, $0 \le j < 2^n$, to obtain the figure $\Psi_2(ABC)$. The area of its heart is then $\alpha^2 \cdot \alpha^2 \cdot |ABC|$; also, the area of the additional arms generated at this stage will not exceed $2(1-\alpha^2)\alpha^2 \cdot |ABC|$. We continue in this way, finally obtaining $\Psi_n(ABC)$.

It follows from (4) that

$$|\Psi_n(ABC)| \le [\alpha^{2n} + 2(1-\alpha^2) + 2(1-\alpha^2)\alpha^2 + \cdots + 2(1-\alpha^2)\alpha^{2n-2}] \cdot |ABC|.$$

However

$$\begin{aligned} 2(1-\alpha^2) + \cdots + 2(1-\alpha^2)\alpha^{2n-2} &\le 2(1-\alpha^2)\sum_{j=0}^{\infty}\alpha^{2j} \\ &= \frac{2(1-\alpha^2)}{1-\alpha^2} \le 2(1-\alpha). \end{aligned}$$

Therefore

$$|\Psi_n(ABC)| \le [\alpha^{2n} + 2(1-\alpha)] \cdot |ABC|. \tag{5}$$

The set $\Psi_n(ABC)$ is essentially the Besicovitch set we are after. Note that if we take α close to 1 and then n large, the factor $\alpha^{2n}+2(1-\alpha)$ can be made as small as we wish.

1.1.3 The following observation gives the thrust of the preceding construction.

Let T_j, $0 \le j < 2^n$, denote the triangles $A_jA_{j+1}C$ that make up ABC, and let T_j' denote the corresponding translated triangles comprising $\Psi_n(ABC)$. Of course, the T_j have a common vertex C; let C_j be the corresponding vertices of the T_j'. Denote by T_j^* the triangles obtained by reflecting the T_j' through C_j. While we have seen that the triangles T_j' overlap to a high degree, the reflected triangles T_j^* are mutually disjoint.

In fact, if T_{j_2} was originally to the right of T_{j_1}, then by construction T_{j_2} was moved leftwards (relative to T_{j_1}), so C_{j_2} is to the left of C_{j_1}. The relative positions of T_{j_1}' and T_{j_2}' are then as in Figure 5, from which the disjointness of $T_{j_1}^*$ and $T_{j_2}^*$ is clear.

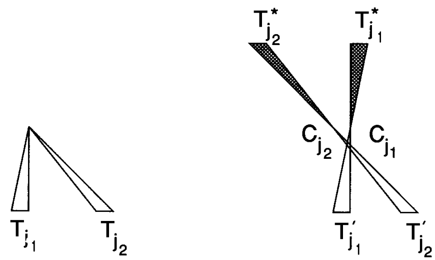

Figure 5. Reflected subtriangles are disjoint.

1.1.4 To complete our construction of the Besicovitch set, we pass from the triangles above to rectangles. We begin by fixing the original triangle ABC to be the equilateral triangle whose altitude has length 2. Next, suppose T_j' is one of the triangles making up $\Psi_n(ABC)$. We draw a line from its vertex C_j to the midpoint of its base, marking off the points P_1 and P_2 on it at distances 1/2 and 3/2 from the vertex. We let R_j denote the rectangle whose major axis is P_1P_2, whose side lengths are 1 and 2^{-N}; here $N = n + c_1$, where c_1 is a fixed large integer (see Figure 6). Since the angle at the vertex C_j' is larger than $c_2 \cdot 2^{-n}$, for some small positive constant c_2, we can always choose c_1 large enough so that $R_j \subset T_j'$. Now let $\tilde{R}_j \subset T_j^*$ be the reflection of R_j through C_j.

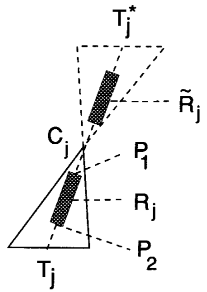

Figure 6. End of the proof of Theorem 1.

We have therefore 2^n rectangles R_j of dimension (1×2^{-N}), so that their "reaches" $\tilde{R}_j$ are disjoint. To pass to the corresponding $2^N = 2^{c_1} \cdot 2^n$ such rectangles, notice that the figure $\Psi_n(ABC)$, together with all its reflected triangles, belongs to a fixed compact set. Taking 2^{c_1} disjoint translates of such sets finally gives us 2^N rectangles, with side lengths 1 and 2^{-N}, which are contained in a set of measure at most

$$2^{c_1} \cdot [\alpha^{2n} + 2(1-\alpha)] \cdot |T|,$$

while the corresponding "reach" rectangles are all disjoint. If we take α sufficiently close to 1, and $N = n + c_1$ sufficiently large, we can make this measure arbitrarily small, completing the proof of the theorem.

2. Maximal functions and counterexamples

We turn to the main topic of this chapter and the next, namely the

study of averages of functions. In its translation-invariant version, the principal problem can be loosely stated as follows.

Say we are given a collection $\mathcal{C} = \{C\}$ of sets. Then for what classes of functions f do we have that, as C ranges over $\mathcal{C}$,

$$\lim_{\operatorname{diam}(C)\to 0} \frac{1}{|C|} \int_C f(x-y)\,dy = f(x) \tag{6}$$

for almost every x?

We have already learned that results of this kind are the consequence of suitable estimates for related maximal functions. In the setting at hand, we are led to consider the corresponding maximal function defined by

$$(M_{\mathcal{C}} f)(x) = \sup_{C\in\mathcal{C}} \frac{1}{|C|} \int_C |f(x-y)|\,dy. \tag{7}$$

Let us recall the instances of this encountered earlier. These were first, in Chapter 1, when $\mathcal{C}$ was essentially the collection of all balls centered at the origin and second, in §3 of Chapter 2, when $\mathcal{C}$ was an arbitrary collection of balls (whose centers might have been far from the origin). In both of these cases one can use a weak-type $(1,1)$ inequality for the maximal function to derive the almost-everywhere results of the type (6).

Our first main result will be that, conversely, in the general setting, qualitative results of the kind (6) imply quantitative conclusions about the maximal operator (7). This can be formulated as follows.

Suppose we are given a countable collection $\{d\mu_j\}_j$ of finite nonnegative measures on $\mathbf{R}^n$, supported in a fixed compact set. We form the maximal operator $\mathcal{M}$, defined by

$$(\mathcal{M}f)(x) = \sup_j |(f * d\mu_j)(x)|. \tag{8}$$

PROPOSITION 1. *Suppose p is given, with $1 \le p < \infty$. Assume that for each $f \in L^p(\mathbf{R}^n)$, we have*

$$(\mathcal{M}f)(x) < \infty \quad \textit{for some set of } x \textit{ having positive measure.} \tag{9}$$

Then it follows that the mapping $f \mapsto \mathcal{M}(f)$ is of weak-type (p,p); that is, there is a constant A so that

$$|\{x : (\mathcal{M}f)(x) > \alpha\}| \le \frac{A}{\alpha^p} \|f\|_{L^p}^p, \tag{10}$$

for all $f \in L^p(\mathbf{R}^n)$ and for every $\alpha > 0$.‡

‡ Generalizations may be found in §3.4–§3.6 below.

2.1 The proof of the proposition uses a Borel-Cantelli type lemma which says (roughly) that, given a suitable collection $\{E_j\}$ of sets in $\mathbf{R}^n$, with

$$\sum_j |E_j| = \infty,$$

we can "randomize" their positions by translations so that they are almost "independent", and as a consequence cover every point of $\mathbf{R}^n$ infinitely often. More precisely, the following holds.

LEMMA. *Let $\{E_j\}$ be a collection of subsets of a fixed compact set, with $\sum_j |E_j| = \infty$. Then there exists a sequence of translates $F_j = E_j + x_j$ so that*

$$\limsup\{F_j\} = \bigcap_{k=1}^{\infty}\Big(\bigcup_{j=k}^{\infty} F_j\Big) = \mathbf{R}^n,$$

except for a set of measure zero.

2.1.1 The proof of the lemma is based on the following observation: Suppose A_1 and A_2 are two subsets of the unit cube $Q \subset \mathbf{R}^n$. Then there is an $h \in \mathbf{R}^n$ so that

$$|A_1 \cap (A_2 - h)| \geq 2^{-n}|A_1| \cdot |A_2|. \tag{11}$$

Indeed, let χ_{A_1} and χ_{A_2} denote the characteristic functions of these sets, and let

$$\eta(x) = \int \chi_{A_1}(y)\,\chi_{A_2}(y+x)\,dy.$$

Clearly $\int \eta(x)\,dx = |A_1| \cdot |A_2|$. Moreover, η is supported in the cube of side length 2 (centered at the origin), so there is an h in this cube with

$$\eta(h) = \int \chi_{A_1}(y)\,\chi_{A_2}(y+h)\,dy \geq 2^{-n} \cdot |A_1| \cdot |A_2|,$$

and (11) holds for this h.

2.1.2 We return to the proof of the lemma and assume, as we may, that the E_j are, after a suitable translation, supported in the unit cube Q. We will construct, first, translates $F_j = E_j + x_j$, so that, except for a set of measure zero, we can cover the unit cube at least once, i.e.,

$$\bigcup_j F_j \supset Q. \tag{12}$$

Take $F_1 = E_1$. Now suppose inductively that $F_1, \ldots, F_{j-1}$ have been determined. Let $A_1 = Q \cap {}^c(F_1 \cup \cdots \cup F_{j-1})$, $A_2 = E_j$. Then construct

$A_2 - h = E_j - h = F_j$, according to the requirement (11) above. Let $p_j = |Q \cap (F_1 \cup \cdots \cup F_j)|$ be the portion of the unit cube Q covered by $F_1 \cup \cdots \cup F_j$. Then

$$\begin{aligned} p_j &= p_{j-1} + |Q \cap {}^c(F_1 \cup \cdots \cup F_{j-1}) \cap F_j| \\ &= p_{j-1} + |A_1 \cap (A_2 - h)| \geq p_{j-1} + 2^{-n} \cdot |A_1| \cdot |E_j| \\ &= p_{j-1} + 2^{-n}(1 - p_{j-1}) \cdot |E_j|. \end{aligned}$$

As a result,

$$p_j - p_{j-1} \geq 2^{-n}(1 - p_{j-1}) \cdot |E_j|. \tag{13}$$

Now the p_j are monotonically increasing and are bounded above by 1. If $\lim_j p_j < 1$ then, since $\sum |E_j| = \infty$, the series

$$\sum (p_j - p_{j-1})$$

would have to diverge, giving a contradiction. Therefore $\lim p_j = 1$ and (12) is established.

We next decompose our collection $\{E_j\}$ into a countably infinite number of subcollections so that, over each subcollection, the sum of the measures diverges. Using the analogue of (12) for appropriate translates of the unit cube, it is then easily seen that almost every point of $\mathbf{R}^n$ can be covered infinitely often by translates of our sets E_j.

2.1.3 With the lemma now established, we return to the proof of the proposition. Let B be a ball that contains the convolution of the unit cube Q and a compact set that contains the support of the measures $d\mu_j$. Thus, if f is supported in Q, then $f * d\mu_j$ is supported in B, and hence $\mathcal{M}f$ is also supported in B.

We first show that our desired conclusion (10) holds whenever f is supported in Q; i.e., that there exists a constant A so that

$$|\{x \in B : \mathcal{M}f(x) > \alpha\}| \leq \frac{A}{\alpha^p}\|f\|_{L^p}^p, \tag{14}$$

for every $f \in L^p$ supported in Q and every $\alpha > 0$.

If this were not the case, we could find, for each k, a positive number α_k, and a $g_k \in L^p$, with g_k supported in the unit cube, so that

$$|\{x \in B : \mathcal{M}g_k(x) > \alpha_k\}| \geq \frac{2^k}{\alpha_k^p}\|g_k\|_{L^p}^p.$$

Now replace g_k by $g_k' = kg_k/\alpha_k$. Then we have a sequence g_k' in L^p with

$$\frac{|\{x \in B : \mathcal{M}g_k'(x) > k\}|}{\|g_k'\|_{L^p}^p} \to \infty, \quad \text{as } k \to \infty.$$

Taking subsequences of the g_k' (with possible repetitions) allows us to find a sequence f_k in L^p, and a sequence of constants $R_k \to \infty$, so that, taking

$$E_k = \{x \in B : \mathcal{M}f_k(x) > R_k\},$$

we have

$$\sum_k |E_k| = \infty, \qquad \text{while} \quad \sum_k \|f_k\|_{L^p}^p < \infty. \tag{15}$$

Notice that since we assumed that the measures $d\mu_j$ are nonnegative, we can suppose that the functions f_k, appearing in (15), are also nonnegative.

Next, according to the lemma, we can find x_k so that the sets $F_k = E_k + x_k$ have the property that

$$\limsup\{F_k\} = \mathbf{R}^n, \text{ a.e. .}$$

Let $\tilde{f}_k(x) = f_k(x + x_k)$, and set $F(x) = \sup \tilde{f}_k(x)$. Then first

$$\mathcal{M}(F) \geq \sup_k \mathcal{M}(\tilde{f}_k),$$

while $\mathcal{M}(\tilde{f}_k) > R_k$ on F_k. Thus $\mathcal{M}(F) = \infty$ almost everywhere.

Now $|F|^p \leq \sum_k |\tilde{f}_k|^p$, so

$$\|F\|_{L^p}^p \leq \sum_k \|\tilde{f}_k\|_{L^p}^p = \sum_k \|f_k\|_{L^p}^p < \infty,$$

and therefore $F \in L^p(\mathbf{R}^n)$. This violates our assumption (9), and hence (14) is proved by contradiction.

From this it is now an easy matter to obtain the full conclusion (10). For each $j \in \mathbf{Z}^n$, let Q_j denote the cube $Q + j$, and let B_j denote the ball $B + j$. Then Q_j is essentially a disjoint covering of $\mathbf{R}^n$, while $\bigcup B_j$ covers each point of $\mathbf{R}^n$ at most N times, for some fixed N. Write

$$f = \sum_j f_j = \sum_j f \cdot \chi_{Q_j}.$$

Then $\mathcal{M}(f) \leq \sum_j \mathcal{M}(f_j)$, while $\mathcal{M}(f_j)$ is supported in B_j. Using (14) above gives

$$|\{x \in B_j : \mathcal{M}f_j(x) > \alpha\}| \leq \frac{A}{\alpha^p}\|f_j\|_{L^p}^p.$$

Thus

$$\begin{aligned}|\{x \in \mathbf{R}^n : \mathcal{M}f(x) > \alpha\}| &\leq \frac{AN}{(\alpha/N)^p}\sum_j \|f_j\|_{L^p}^p \\ &= \frac{AN^{p+1}}{\alpha^p}\sum_j \|f_j\|_{L^p}^p = \frac{AN^{p+1}}{\alpha^p}\|f\|_{L^p}^p,\end{aligned}$$

and the proposition is proved.

2.2 Counterexample for rectangles with arbitrary orientations. We now apply the construction of the Besicovitch set, together with the maximal principle given in the proposition just proved, to show that differentiation of the integral (in the sense of (6) above) does not hold for rectangles having arbitrary orientations.

We place ourselves in $\mathbf{R}^2$ and take for our collection $\mathcal{C}$ the set $\mathcal{R}$ of *all* rectangles centered at the origin. For any rectangle $R \in \mathcal{R}$, we write $\operatorname{diam}(R)$ for its diameter; this quantity is of course comparable with the length of the longest side of R. By the *orientation* of R, we mean the angle made by its longer side with the x-axis.

COROLLARY 1. *For each p with $1 \leq p < \infty$, there exists an $f \in L^p(\mathbf{R}^2)$ so that*

$$\limsup_{\substack{\operatorname{diam}(R)\to 0\\ R\in\mathcal{R}}} \frac{1}{|R|}\int_R f(x-y)\,dy = \infty, \qquad a.e.\ x. \tag{16}$$

Here $\limsup\limits_{\substack{\operatorname{diam}(R)\to 0\\ R\in\mathcal{R}}}$ *is defined to be* $\inf\limits_{\delta>0} \sup\limits_{\substack{\operatorname{diam}(R)<\delta\\ R\in\mathcal{R}}}$.

Let us consider first the $\mathcal{M}$ defined by

$$\sup_{\operatorname{diam}(R)<8} \frac{1}{|R|}\left|\int_R f(x-y)\,dy\right|.$$

Then clearly, by restricting ourselves to rectangles with rational side lengths and rational orientations, we see that

$$\mathcal{M}f(x) = \sup_j |(f * d\mu_j)(x)|,$$

where $\{d\mu_j\}$ is a countable collection of positive measures supported in a fixed compact set. Our proposition therefore applies and shows that there is an $f \in L^p$ with $\mathcal{M}(f) = \infty$ a.e., unless the mapping $\mathcal{M}$ is of weak-type (p,p).

We now use the Besicovitch set E given by Theorem 1 to show that no such weak inequality can hold. Indeed, take $f = \chi_E$, where $E = \bigcup_{j=1}^{2^N} R_j$; here the R_j are as in §1.1.4. Then $\|\chi_E\|_{L^p}^p = |E| < \varepsilon$. Moreover, it is clear that if x belongs to any of the "reach" rectangles $\tilde{R}_j$, then there is a rectangle R, centered at x, with $\operatorname{diam}(R) \leq 8$, so that

$$|R \cap R_j| \geq \frac{1}{12}|R|.$$

Hence $\mathcal{M}\chi_E \geq 1/12$ in the set $\bigcup_j \tilde{R}_j$, which has measure 1. Thus, for any $A > 0$, we can find a set E so that

$$|\{x : \mathcal{M}\chi_E > \alpha\}| \leq \frac{A}{\alpha^p}\|\chi_E\|_{L^p}^p$$

is violated, so $\mathcal{M}$ is not of weak-type (p,p). As remarked above, this shows that there is an $f \in L^p$ with $\mathcal{M}f(x) = \infty$ almost everywhere.

Next, we define

$$\mathcal{M}_\delta f(x) = \sup_{\mathrm{diam}(R)<\delta} \frac{1}{|R|}\left|\int_R f(x-y)\,dy\right|,$$

so that $\mathcal{M} = \mathcal{M}_8$; a simple dilation of $\mathbf{R}^2$ shows that for each $j \in \mathbf{N}$, there is an $f_j \in L^p(\mathbf{R}^2)$ with $\mathcal{M}_{1/j}(f_j) = \infty$ almost everywhere. Clearly, we can take $f_j \geq 0$ and $\|f_j\|_{L^p} \leq 2^{-j}$. Writing $f = \sum f_j$, we have $f \in L^p$ and

$$\limsup_{\mathrm{diam}(R)\to 0} \frac{1}{|R|}\int_R f(x-y)\,dy = \infty \qquad \text{for a.e. } x,$$

completing the proof of the corollary.

Remark. By a similar construction, we can show that in the present context (6) also fails, even for characteristic functions of sets of finite measure. See §3.3.

2.3 Rectangles with sides parallel to the axes. Having seen the negative results for the family of rectangles with all possible orientations, it may be worthwhile to compare this to what happens when this family is greatly restricted, by limiting ourselves to rectangles with a fixed orientation.

For this purpose, let $\mathcal{R}_0$ denote the set of rectangles in $\mathbf{R}^2$ that are centered at the origin and whose sides are parallel to the coordinate axes. We take $\mathcal{C} = \mathcal{R}_0$, and consider the maximal operator $\mathcal{M}_{\mathcal{R}_0}$ given by (7). Then $\mathcal{M}_{\mathcal{R}_0}$ is bounded on $L^p(\mathbf{R}^2)$, for $1 < p \leq \infty$; moreover, the result (6) also holds for this family, whenever f is locally in L^p, $p > 1$. All of this is a consequence of iterating the usual one-dimensional L^p theory.†

On the other hand, such results fail at $p = 1$.

COROLLARY 2. *There exists an $f \in L^1(\mathbf{R}^2)$ so that*

$$\limsup_{\mathrm{diam}(R)\to 0} \frac{1}{|R|}\int_R f(x-y)\,dy = \infty \qquad \textit{for a.e. } x, \tag{17}$$

where R ranges over $\mathcal{R}_0$.

We wish to use the proposition stated at the beginning of §2. To do so, we must work with a countable collection of measures that are supported in a single compact set. Thus, to prove the corollary, it suffices to show that the more restricted maximal function

$$f(x) \mapsto \sup_R \frac{1}{|R|}\left|\int_R f(x-y)\,dy\right|$$

† See Chapter 2, §5.20; there we have written $M_S = M_{\mathcal{R}_0}$.

is not of weak-type $(1,1)$, where R ranges over those rectangles in $\mathcal{R}_0$ whose side lengths are rational and whose diameter is at most four. This is equivalent to showing that

$$\mathcal{M}f(x) = \sup_{\substack{\mathrm{diam}(R)\leq 4\\ R\in\mathcal{R}_0}} \frac{1}{|R|}\left|\int_R f(x-y)\,dy\right|$$

is not of weak-type $(1,1)$.

The easiest way to see that this $\mathcal{M}$ cannot be of weak-type $(1,1)$ is to consider the Dirac delta function; we ask whether

$$|\{x : \mathcal{M}f > \alpha\}| \leq \frac{A}{\alpha}\int |f|\,dx \qquad \text{for all } \alpha > 0 \tag{18}$$

can hold, replacing f by the unit mass at the origin. To this end, let $\{f_j\}$ be a sequence of nonnegative functions so that f_j is supported on $|x| < 1/j$ and $\int f_j\,dx = 1$, for all j. The weak limit of the f_j is then the Dirac measure $d\mu$.

We now use the simple observation that, for each $x = (x_1, x_2)$, if R is a rectangle centered at the origin with side lengths ℓ_1, ℓ_2 and $\ell_1 > 2|x_1|$, $\ell_2 > 2|x_2|$, then R contains x in its interior. Therefore, if R is any rectangle in $\mathcal{R}_0$,

$$\lim_{j\to\infty} \frac{1}{|R|}\int_R f_j(x-y)\,dy = \frac{1}{|R|}\int_R d\mu(x-y) = \frac{1}{|R|}\chi_R.$$

Restricting ourselves to R with $\mathrm{diam}(R) \leq 4$, we have

$$\mathcal{M}(d\mu)(x) \geq \frac{1}{4|x_1|\cdot|x_2|},$$

whenever $|x| \leq 1$.

As a result, if (18) were true, we would have

$$|\{x : |x| \leq 1, (4|x_1 x_2|)^{-1} > \alpha\}| \leq A/\alpha. \tag{19}$$

The set appearing on the left side of (19) can be expressed more simply as

$$\{x : |x| \leq 1, |x_1 x_2| \leq (4\alpha)^{-1}\},$$

which is the subset of the unit disk "inside" of the two hyperbolas $\pm x_1 x_2 = (4\alpha)^{-1}$. The area of this set is $\sim (1/\alpha)\log\alpha$, as $\alpha \to \infty$, which violates (19). This completes the proof of Corollary 2.

2.4 Some remarks about Poisson integrals. Poisson integrals are themselves prime examples of averages of functions, and their study has played a vital role in the development of our subject. We wish here to illustrate (albeit in a sketchy way) the above results in terms of these concepts.

2.4.1 Let Γ be a regular cone[‡] in $\mathbf{R}^n$, and consider the tube domain T_Γ in $\mathbf{C}^n$ given by

$$T_\Gamma = \{x + iy \in \mathbf{C}^n : y \in \Gamma\}.$$

Each such domain has a naturally associated Poisson kernel $P_y(x)$; with it we can form the "Poisson integral"

$$u(x,y) = (f * P_y)(x),$$

for appropriate functions f defined on $\mathbf{R}^n$. We give three examples.

(i) $\Gamma_1 \subset \mathbf{R}^1$, where Γ_1 is the positive half-line. Then T_{Γ_1} is the usual upper half-plane,

$$P_y(x) = \frac{y}{\pi(x^2+y^2)},$$

and $u(x,y)$ is the classical Poisson integral of f.

(ii) $\Gamma_2 \subset \mathbf{R}^2$, where Γ_2 is the first quadrant; Γ_2 is the product of two half-lines, $\Gamma_2 = \Gamma_1^{(1)} \times \Gamma_1^{(2)}$. The Poisson kernel P_y is also a product:

$$P_y(x) = \frac{1}{\pi^2} \cdot \frac{y_1}{(x_1^2+y_1^2)} \cdot \frac{y_2}{(x_2^2+y_2^2)}$$

(iii) $\Gamma_3 \subset \mathbf{R}^3$, where Γ_3 is the "forward light-cone":

$$\Gamma_3 = \{(y_1,y_2,y_3) : y_3 > (y_1^2+y_2^2)^{1/2}\}.$$

Here

$$P_y(x) = \frac{c \cdot (y,y)^{3/2}}{[(\langle x,x\rangle - \langle y,y\rangle)^2 + 4\langle x,y\rangle]^{3/2}},$$

where $\langle \cdot,\cdot \rangle$ is the indefinite form

$$\langle u,v\rangle = u_3v_3 - u_1v_1 - u_2v_2$$

and c is a constant.[†]

There is a close connection between the behavior of the Poisson integrals $f * P_y$ as $y \to 0$, and the averages over rectangles discussed above. The analogue of question (6) is the issue of whether

$$\lim_{\substack{y\to 0\\ y\in\Gamma}} (f * P_y)(x) = f(x) \quad \text{for a.e. } x. \tag{20}$$

The analogue of the maximal operator (7) is then

$$P_*^\Gamma f(x) = \sup_{y\in\Gamma} |f * P_y(x)|. \tag{21}$$

The relations one can prove among these are as follows. We limit ourselves to f that are nonnegative.

‡ A regular cone Γ is an open convex cone whose dual cone (the set of ξ so that $\xi \cdot x > 0$ for all $x \in \Gamma$) has a nonempty interior.

† For these formulas, and for further background about Poisson integrals, see *Fourier Analysis*, chapters 2 and 3.

In case (i)

$$(P_*^{\Gamma_1} f)(x) \approx (Mf)(x). \tag{22}$$

Here M is the standard one-dimensional maximal operator (we may write $M = M_{\mathcal{C}}$, where $\mathcal{C}$ is the set of all intervals centered at the origin). The notation $\approx$ means that the quantities on either side are comparable, with bounds that are independent of the function f and the point x.

As a result, there are the usual L^p and L^1 inequalities for $P_*^{\Gamma_1}$, and also convergence a.e. in (20) whenever f is locally integrable.[‡]

In case (ii)

$$(P_*^{\Gamma_2} f)(x) \approx (M_{\mathcal{R}_0} f)(x), \tag{23}$$

where, as in §2.3, $\mathcal{R}_0$ is the set of rectangles in $\mathbf{R}^2$, centered at the origin, with sides parallel to the coordinate axes. Thus we have L^p inequalities for $P_*^{\Gamma_2}$ (and the convergence (20)) when $f \in L^p(\mathbf{R}^2)$, $p > 1$, but no positive results at $p = 1$.

In case (iii)

$$(P_*^{\Gamma_3} f)(x) \geq c M_{\mathcal{R}}(f')(x_1,x_2)\chi(x_3) \tag{24}$$

if $f(x_1,x_2,x_3) = f'(x_1,x_2)\chi(x_3)$, where χ is the characteristic function of the unit interval, and $c > 0$. Here $\mathcal{R}$ is the collection of rectangles with all orientations (as in §2.2), and so we have no positive results in this case for any L^p. Further details may be found in §3.7 below.

2.4.2 One may ask what happens when the approach region used above (as $y \to 0$, we allow all $y \in \Gamma$), called the "unrestricted approach", is replaced by the restricted approach, where y is required to lie in a proper subcone Γ_0 of Γ.[†] In this new situation the usual L^p, $p \geq 1$ results hold for both examples (ii) and (iii). Indeed, the maximal operator

$$f \mapsto \sup_{y\in\Gamma_0} |(f * P_y)(x)|$$

is of weak-type $(1,1)$ because it falls under the scope of the singular approximations to the identity treated by Proposition 2 in §4 of Chapter 2. See also Chapter 2, §5.18 and Chapter 13, §7.11.

2.4.3 A crucial element responsible for the negative results for the Poisson integrals in the Γ_3 case above, as well as for the maximal operator $M_{\mathcal{R}}$, is that we are considering the supremum over a continuum of singular directions. In these examples, the singular directions are parametrized by a circle and are inherent in the rotational invariance of the operators. The situation is similar for the counterexample for the "ball multiplier" in $\mathbf{R}^2$, which is our next concern; there, every direction normal to the boundary will again play the role of a singular direction.

‡ We remark that if f is not required to be positive then $P_*^{\Gamma_1} f \in L^p$ exactly when $f \in H^p$ (as in Chapter 3) but Mf is *never* in L^p when $p \leq 1$.

† That is, the closure of Γ_0 is contained in $\Gamma \cup \{0\}$.

2.5 Counterexample for the ball multiplier. We shall now consider the Fourier multiplier given by the characteristic function of a ball, i.e., the "partial sum" operator S^δ for $\delta = 0$, given by (28) of the previous chapter. Thus we are concerned with the operator S, defined for $f \in L^2(\mathbf{R}^n)$ by

$$Sf(x) = \int_{|\xi|<1} \widehat{f}(\xi) e^{2\pi i x\cdot\xi}\, d\xi. \tag{25}$$

PROPOSITION. *Suppose $n \geq 2$ and $p \neq 2$. Then the operator S, initially defined on $L^2 \cap L^p$, is not extendable to a bounded operator from $L^p(\mathbf{R}^n)$ to itself.*

2.5.1 For any ball $B \subset \mathbf{R}^n$, denote by S_B the multiplier operator associated to the ball B:

$$S_B f(x) = \int_B \widehat{f}(\xi) e^{2\pi i x\cdot\xi}\, d\xi.$$

Similarly, if u is a unit vector in $\mathbf{R}^n$, let S^u denote the multiplier operator corresponding to the half-space in $\mathbf{R}^n$ whose normal direction is u, i.e.,

$$S^u f(x) = \int_{\xi\cdot u>0} \widehat{f}(\xi) e^{2\pi i x\cdot\xi}\, d\xi.$$

We shall see that if an inequality of the form

$$\|Sf\|_{L^p} \leq A_p \|f\|_{L^p}, \qquad f \in L^2 \cap L^p \tag{26}$$

holds, then so would corresponding vector-valued inequalities involving the S_B's and the S^u's.

LEMMA. *Suppose that (26) were valid for some p, $1 \leq p \leq \infty$. If we are given any finite collection of functions $f_1, \dots, f_M \in L^2 \cap L^p$, and any finite collection of unit vectors $u_1, \dots, u_M \in \mathbf{R}^n$, we can then assert that*

$$\left\| \left(\sum_{j=1}^M |S^{u_j}(f_j)|^2 \right)^{1/2} \right\|_{L^p(\mathbf{R}^n)} \leq A_p \left\| \left(\sum_{j=1}^M |f_j|^2 \right)^{1/2} \right\|_{L^p(\mathbf{R}^n)}, \tag{27}$$

where the bound A_p in (27) is the same as that in (26).

To prove the lemma, note that if $B = B_R$ is the ball of radius R centered at the origin, then (26) implies that

$$\|S_B(g)\|_{L^p} \leq A_p \|g\|_{L^p}, \qquad g \in L^2 \cap L^p. \tag{28}$$

Indeed, let δ_R be the scaling operator $\delta_R(g)(x) = g(x/R)$. An easy calculation shows that $\delta_R^{-1} S \delta_R = S_{B_R}$ (since $S = S_{B_1}$), and this gives (28).

Now let $f = (f_1, \dots, f_M)$ be our given M-tuple of functions. If T is an operator, we write $T(f) = (Tf_1, \dots, Tf_M)$. Next, for any unit vector $\omega = (\omega_j) \in \mathbf{C}^M$, let

$$S_\omega(f) = \sum_{j=1}^M \bar{\omega}_j S_B(f_j), \quad \text{and} \quad f_\omega = \sum_{j=1}^M \bar{\omega}_j f_j.$$

Suppose first that $p < \infty$. According to (28), since $S_\omega(f) = S_B(f_\omega)$, we have

$$\int_{\mathbf{R}^n} |S_\omega(f)(x)|^p\, dx \leq A_p^p \int_{\mathbf{R}^n} |f_\omega(x)|^p\, dx. \tag{29}$$

Observe that

$$|S_\omega f(x)| = \left(\sum_{j=1}^M |S_B(f_j)(x)|^2 \right)^{1/2} \cdot |\phi(\omega, S_B f(x))|,$$

where

$$\phi(\omega, S_B f(x)) = \langle S_B f(x)/|S_B f(x)|, \omega \rangle$$

when $S_B f(x) \neq 0$, and is irrelevant when $S_B f(x) = 0$; here $\langle v, z \rangle = v \cdot \bar{z}$ is the usual Hermitian inner product on $\mathbf{C}^M$.

Now integrate both sides of (29) with respect to ω (before integrating in x). The left side of (29) gives

$$\int \left(\sum_j |S_B(f_j)|^2 \right)^{p/2} dx \cdot \gamma_p,$$

where

$$\gamma_p = \int_{|\omega|=1} |\phi(\omega, S_B f(x))|^p\, d\omega.$$

Since the unitary group is transitive and preserves $d\omega$, we see that γ_p is independent of x with

$$\gamma_p = \int_{|\omega|=1} |\phi(\omega, \mathbf{1})|^p\, d\omega \neq 0;$$

here $\mathbf{1} = (1, 0, \dots, 0)$. Similarly, the right side of (29) also contributes a factor of γ_p and therefore‡

$$\left\| \left(\sum_{j=1}^M |S_B(f_j)|^2 \right)^{1/2} \right\|_{L^p(\mathbf{R}^n)} \leq A_p \left\| \left(\sum_{j=1}^M |f_j|^2 \right)^{1/2} \right\|_{L^p(\mathbf{R}^n)}, \tag{30}$$

proving the result for $p < \infty$. A simple variant of this argument gives the corresponding result for $p = \infty$.

‡ Notice that (according to our proof) this inequality actually holds with an arbitrary bounded linear operator T in place of S_B.

Next, let B_R^u be the ball of radius R, centered at uR, where u is a unit vector. Now

$$S_{B_R^u} f(x) = e^{2\pi i uR\cdot x} \cdot S_{B_R}(f \cdot e^{-2\pi i uR\cdot x}),$$

so (30) implies that

$$\left\| \left(\sum_{j=1}^{M} |S_{B_R^{u_j}}(f_j)|^2 \right)^{1/2} \right\|_{L^p(\mathbf{R}^n)} \le A_p \left\| \left(\sum_{j=1}^{M} |f_j|^2 \right)^{1/2} \right\|_{L^p(\mathbf{R}^n)}. \tag{31}$$

Now it is only a matter of letting $R \to \infty$. We note that each ball $B_R^{u_j}$ increases to fill up the half-space $\{\xi : \xi \cdot u_j > 0\}$; hence $S_{B_R^{u_j}}(f_j)$ converges to $S^{u_j}(f_j)$ in L^2, and consequently an appropriate subsequence converges to $S^{u_j}(f_j)$ almost everywhere. Therefore (31) gives us (27), and the lemma is proved.

2.5.2 We now examine further the multiplier operators for the half-space (the S^u), and in particular the case that arises when $n = 1$. Denote by S^+ the multiplier for the half-line $\xi_1 > 0$,

$$(S^+ f)(x) = \int_0^\infty \widehat{f}(\xi) e^{2\pi i x\xi}\, d\xi, \qquad f \in L^2(\mathbf{R}^1).$$

If $f = \chi_{(-1/2,1/2)}$ is the characteristic function of the unit interval $(-1/2, 1/2)$, then observe that

$$|(S^+ f)(x)| \ge c/|x|, \qquad \text{for } |x| > 1/2, \tag{32}$$

for some constant $c > 0$. Indeed,

$$(S^+ f)(x) = \lim_{\varepsilon \searrow 0} \int_0^\infty \widehat{f}(\xi)\, e^{2\pi i (x+i\varepsilon)\xi}\, d\xi$$

in the L^2 sense, by Plancherel's theorem. However, if $\varepsilon > 0$,

$$\begin{aligned}\int_0^\infty \widehat{f}(\xi)\, e^{2\pi i(x+i\varepsilon)\xi}\, d\xi &= \int_{-\infty}^{\infty} \left(\int_0^\infty e^{-2\pi i y\cdot\xi}\, e^{2\pi i(x+i\varepsilon)\cdot\xi}\, d\xi \right) f(y)\, dy \\ &= \frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{f(y)}{(y - x - i\varepsilon)}\, dy,\end{aligned}$$

from which the assertion (32) follows directly.

Next, let R denote the rectangle in the plane whose sides are parallel to the axes and is given as the product of the intervals $(-1/2, 1/2)$ and $(-2^{-N-1}, 2^{-N-1})$. We then have

$$\chi_R(x_1, x_2) = \chi_{(-1/2,1/2)}(x_1) \cdot \chi_{(-2^{-N-1}, 2^{-N-1})}(x_2).$$

If $F(x_1, x_2) = f_1(x_1) \cdot f_2(x_2)$ and u is the unit vector in the x_1-direction, then since

$$(S^u F)(x_1, x_2) = (S^+ f_1)(x_1) \cdot f_2(x_2),$$

we have that

$$(S^u \chi_R)(x_1, x_2) = [S^+ \chi_{(-1/2,1/2)}](x_1) \cdot \chi_{(-2^{-N-1}, 2^{-N-1})}(x_2).$$

So if $\tilde{R}$ is the "reach" rectangle associated to R (as defined at the beginning of §1), we see that, by (32),

$$|S^u(\chi_R)| \ge c' \chi_{\tilde{R}}.$$

A similar inequality holds for any rectangle R_j with sides of length 1 and 2^{-N}. Let u_j denote the unit vector in the (positive) direction of the longest side. Then if we rotate and translate the axes appropriately to transform the rectangle R_j to the previous R, we see that, as above,

$$|S^{u_j}(\chi_{R_j})| \ge c' \chi_{\tilde{R}_j}. \tag{33}$$

2.5.3 We shall now take $R_1, \ldots, R_{2^N}$ to be the collection of rectangles given to us by the Besicovitch construction (Theorem 1 in §1) and obtain a contradiction with (26), first for $p < 2$ and $n = 2$.

As we have seen, the assertion (26) implies (27). If we take $f_j = \chi_{R_j}$, and $M = 2^N$, then (33) shows that the left side of (27) exceeds c', since $|\cup \tilde{R}_j| = 1$. However, by Hölder's inequality, the right side of (27) would then be dominated by

$$A_p \left[\int \left(\sum |\chi_{R_j}|^2 \right) dx \right]^{1/2} \cdot \left(\int_E dx \right)^{1/q};$$

here q is the exponent dual to $2/p$, $1/q = 1 - p/2$ and we have written $E = \cup R_j$, so $(\sum |\chi_{R_j}|^2)^{1/2}$ is supported in E. Moreover

$$\int \sum |\chi_{R_j}|^2\, dx = \sum |R_j| = 1.$$

Taking $|E| < \varepsilon$ as in Theorem 1, we have

$$c' \le A_p \varepsilon^{1-p/2},$$

which is not possible if ε is sufficiently small.

To deal with the case $n > 2$, we split the coordinates in $\mathbf{R}^n$ as $x = (x_1, x_2, x')$ with $x' \in \mathbf{R}^{n-2}$, and take

$$f_j(x) = \chi_{R_j}(x_1, x_2) \cdot f(x'),$$

where f is a fixed function on $\mathbf{R}^{n-2}$. Matters then easily reduce to the case $n = 2$. Finally, the result for $p > 2$ follows from that for $p < 2$ by duality, since

$$\langle Sf, g\rangle = \langle f, Sf\rangle,$$

whenever $f, g \in L^2$.

3. Further results

A. Besicovitch and Nikodym sets

3.1 The Kakeya "needle problem" arises in the question of determining the minimum area of a set K in the plane that contains unit line segments in all possible directions; a more restrictive version of the problem is the construction of a set $\tilde{K}$ of minimal area inside of which a unit line segment can be continuously moved (i.e., translated and rotated) so that it takes on all possible directions. For this second problem, it was originally believed that the minimal area was $\pi/8$. However:

(a) For each $\varepsilon > 0$, there exists a set K, with $|K| < \varepsilon$, that contains unit line segments in all possible orientations.

Indeed, if we refer to §1.1.4 in our construction of the Besicovitch set, we see that all the unit line segments that lie in the triangle ABC and have C as an endpoint have translates that lie in $\Psi_n(ABC)$; a finite union of such figures (with n large) yields K.

(b) For each $\varepsilon > 0$, there is a set $\tilde{K}$, with $|\tilde{K}| < \varepsilon$, inside of which a unit line segment can be moved continuously, changing its orientation by 180 degrees while doing so.

This can be deduced from the construction (a) and the following elementary observations. Suppose ℓ_1, ℓ_2 are two parallel unit line segments in the plane. Then for any $\varepsilon > 0$, there is a set S_ε, $|S_\varepsilon| < \varepsilon$, so that, within S_ε, ℓ_1 can be continuously moved in to the position ℓ_2. To see this, one need merely move ℓ_1 along its initial direction far enough, then return (at a slight angle) to the position of ℓ_2.

(c) Coming back to the first, less restrictive problem, there is a compact set K, with $|K| = 0$, that contains unit line segments in all directions.

This is based on the following proposition. Suppose π is a parallelogram in the (x, y) plane so that two of its sides lie on the lines $y = 0$ and $y = 1$, respectively. Then given any $\varepsilon > 0$, we can find parallelograms $\pi_1, \ldots, \pi_N$, each having two sides lying on the lines $y = 0$ and $y = 1$, with $\pi_i \subset \pi$, $\left|\bigcup_{i=1}^{N} \pi_i\right| < \varepsilon$,

and so that any line segment in π that joins the lines $y = 0$ and $y = 1$ has a translate that is contained in one of the π_i.

This proposition is actually a simple consequence of the construction in §1. Taking it for granted, we prove (c). Let K_1 be a unit square having two sides on the lines $y = 0$ and $y = 1$. Using the proposition, we construct a sequence

$$K_1 \supset K_2 \supset \cdots \supset K_j \supset \cdots$$

of compact sets with $|K_j| \leq 1/j$, and so that K_j contains translates of all the line segments joining the two horizontal sides of K_1. Finally, one takes K to be a finite union of sets of the kind $\bigcap_{j=1}^{\infty} K_j$.

Besicovitch [1928]; see also the accounts in de Guzmán [1975] and Falconer [1985], in which further references to the literature can be found.

3.1[a] A question arises as to the existence of analogues of Besicovitch sets in $\mathbf{R}^n$ (that is, sets of small measure that contain a translate of every unit k-ball in $\mathbf{R}^n$) The construction in §1 obviously gives such sets when $k = 1$, for all $n \geq 2$. On the other hand, it can be shown that when k is sufficiently large (in terms of n), there can be no such sets.

The initial theorems regarding this are in Marstrand [1979] and Falconer [1985]; connections are made with estimates for Radon transforms in Oberlin and Stein [1982]; the most recent results are in Bourgain [1991a], which also contains an extension of §3.12(d) below to certain p, with $p > (n+1)/2$.

3.2 Another paradoxical set in the plane, which for certain problems plays a role of equal significance as do the Besicovitch sets treated in §1 and §3.1, was found by Nikodym [1927]:

There exists a subset N of the unit square, having full measure (i.e., $|N| = 1$), so that for each $x \in N$ there is a line $\ell(x)$ with $N \cap \ell(x) = \{x\}$.

The original proof of the existence of such an N was quite complicated and indirect. It was later shown that the construction of N follows from that given for the Besicovitch set in §1; this also showed that the mapping $x \mapsto \ell(x)$ can be taken to be measurable. See R. Davies [1952], de Guzmán [1975].

3.3 In $\mathbf{R}^2$, differentiability with respect to the collection of centered rectangles with arbitrary orientations fails, even for characteristic functions of measurable sets. That is, the assertion

$$\lim_{\substack{\operatorname{diam}(R)\to 0\\ R\in\mathcal{R}}} |R|^{-1} \int_R f(x-y)\,dy = f(x), \qquad \text{a.e. } x$$

may fail for $f = \chi_E$, where $E \subset \mathbf{R}^2$ is a measurable set.

There are two ways of seeing this. First, given any $\varepsilon > 0$, there is a closed set F in the unit square, with $|F| \geq 1 - \varepsilon$, so that

$$\liminf_{\substack{\operatorname{diam}(R)\to 0\\ R\in\mathcal{R}}} |R|^{-1} \int_R \chi_F(x-y)\,dy = 0 \qquad \text{for every } x.$$

This follows directly from §3.2 by choosing $F \subset N$, $|F| \geq 1-\varepsilon$, as was observed by Zygmund (in Nikodym [1927]).

A second (less elegant) method is to use the Besicovitch set constructed in §1. We find first a set E_ε, $|E_\varepsilon| < \varepsilon$, so that $\mathcal{M}_8(\chi_{E_\varepsilon}) \geq 1/12$ on a set of measure 1 (see §2.2). By re-scaling and using unions of such sets, we can construct $E^{j,\varepsilon}$ so that each $\mathcal{M}_{1/j}(\chi_{E^{j,\varepsilon}}) \geq 1/12$ on a set of measure 1, and $|E^{j,\varepsilon}| < \varepsilon$. Finally, using the lemma in §2.1, we can obtain, for each $\varepsilon > 0$, a set E, with $|E| < \varepsilon$, so that

$$\limsup_{\substack{\operatorname{diam}(R)\to 0\\ R\in\mathcal{R}}} |R|^{-1} \int_R \chi_E(x-y)\,dy \geq 1/12 \qquad \text{for a.e. } x.$$

For the use of the Besicovitch set in this connection, see also Busemann and Feller [1934].

B. The weak-type maximal principle

3.4 In their original formulation, results of the kind given by Proposition 1 in §2 (that almost-everywhere assertions imply weak-type inequalities) apply to operators that are not assumed to preserve positivity. Similar conclusions hold, but only when $p \leq 2$.

More precisely, let $T_j f = f * K_j$ be a sequence of convolution operators whose distribution kernels K_j are supported in a common compact set. Assume that, for each j,

$$T_j : L^p(\mathbf{R}^n) \to \{\text{measurable functions on } \mathbf{R}^n\}$$

is continuous in the following weak sense: If $f_k \to f$ in L^p, then $T_j(f_k) \to T_j(f)$ in measure. If for every $f \in L^p$, we know that $(\mathcal{M}f)(x) = \sup_j |(T_j f)(x)| < \infty$ on a set of positive measure, we can conclude that

$$|\{x : (\mathcal{M}f)(x) > \alpha\}| \leq \frac{A}{\alpha^q}\|f\|^q_{L^p}, \qquad \text{all } \alpha > 0,$$

where $q = \min(p, 2)$.

Moreover, if we drop the assumption that K_j have a common compact support, then we can still conclude that, for each compact set L, there is a bound A_L so that

$$|\{x \in L : (\mathcal{M}f)(x) > \alpha\}| \leq \frac{A_L}{\alpha^q}\|f\|^q_{L^p}, \qquad \text{all } \alpha > 0.$$

The proof of these assertions begins as in §2.1.3. However, once the f_n are constructed (as per (15)), the violating function F can no longer be taken to be $\sup_k \tilde{f}_k$; we instead take

$$F(x,t) = \sum_k r_k(t)\,\tilde{f}_k(x) \qquad (\text{for most } t),$$

where $\{r_k\}$ are the Rademacher functions.† The mutual independence of these functions guarantees that, for those t, the effects

$$(\mathcal{M}\tilde{f}_k)(x) > R_k \qquad \text{on the sets } E_k + x_k$$

are not cancelled out. Moreover, their orthogonality ensures that $F(x,t) \in L^p(\mathbf{R}^n, dx)$ for almost every t, because

$$\int_0^1 \int_{\mathbf{R}^n} |F(x,t)|^p\,dx\,dt \leq \sum_k \|\tilde{f}_k\|^p_{L^p},$$

if $1 \leq p \leq 2$.

For further details, see Stein [1961] where, in addition, an example is given showing that the conclusion with $q = \min(p,2)$ is best possible. See also Burkholder [1964] and S. Sawyer [1966], in which these ideas are adapted to "non-convolution" situations; other generalizations appear below.

3.5 An analysis of the scheme of the proof of the results in §3.4 leads to the following extensions.

(a) We may drop the assumption that our operators are translation-invariant. That is, suppose only that each T_j is continuous from $L^p(\mathbf{R}^n)$ to measurable functions on $\mathbf{R}^n$ (again with the topology of convergence in measure). As before, we set $(\mathcal{M}f)(x) = \sup_j |(T_j f)(x)|$, and now we assume that

$$(\mathcal{M}f)(x) < \infty \qquad \text{for } \textit{almost every } x,$$

whenever $f \in L^p$. Then there is a measure $d\mu = \omega(x)\,dx$, with $\omega(x) > 0$ for a.e. x, so that

$$\mu\{x : (\mathcal{M}f)(x) > \alpha\} \leq \left(\frac{\|f\|_{L^p}}{\alpha}\right)^q, \qquad \text{for all } \alpha > 0,$$

with $q = \min(p, 2)$.

(b) More generally, the L^p space in the domain of the T_j can be replaced by any Banach space B, with norm $\|\cdot\|_B$, which is a space of "type q" in the sense that

$$\int_0^1 \|\sum_k r_k(t) f_k\|^q_B\,dt \leq c_q \sum_k \|f_k\|^q_B,$$

for all $f_k \in B$; here r_k are the Rademacher functions. Each T_j is assumed to be continuous from B to the measurable functions on $\mathbf{R}^n$ (with the topology of convergence in measure), and we suppose that $(\mathcal{M}f)(x) < \infty$ almost everywhere, for every $f \in B$. Then there exists a measure $d\mu = \omega(x)\,dx$, with $\omega(x) > 0$ for a.e. x, so that

$$\mu\{x : (\mathcal{M}f)(x) > \alpha\} \leq \left(\frac{\|f\|_B}{\alpha}\right)^q, \qquad \text{for all } \alpha > 0,$$

again with $q = \min(p, 2)$.

† A definition (and further discussion) of the Rademacher functions may be found in *Singular Integrals*, Chapter 5, §5.2.

Note that by the inequality stated in §3.4, the space L^p is a space of type p, for $1 \le p \le 2$. It should also be remarked that (a) and (b) still hold when the measure space $(\mathbf{R}^n, dx)$ is replaced by any σ-finite measure space.

See Nikishin [1970]; also Maurey [1974] and Gilbert [1979], among the many papers dealing with further generalizations.

3.6 There are other variants of the principle in §3.4, which are not strictly contained in the generalizations §3.5, that are also of interest; the following is an extension in the context of Orlicz spaces.

(a) Let $u \mapsto \Phi(u)$, $u \ge 0$, be a strictly increasing convex function with $\Phi(0) = 0$. Define L^Φ to be the Banach space of functions f for which

$$\int_{\mathbf{R}^n} \Phi\Big(\frac{|f|}{\mathrm{c}}\Big)\, dx \le 1, \qquad \text{for some } \mathrm{c} > 0,$$

and let $\|f\|_{L^\Phi} = \|f\|_\Phi$ be the least c for which the inequality holds. Suppose the T_j are as in §3.4 and assume that $(\mathcal{M}f)(x) < \infty$ on a set of positive measure, whenever $f \in L^\Phi$. Then we can conclude that there is a constant A for which

$$|\{x : (\mathcal{M}f)(x) > \alpha\}| \le \int_{\mathbf{R}^n} \Phi\Big(\frac{A|f|}{\alpha}\Big)\, dx, \qquad \text{all } \alpha > 0,$$

if, in addition, we assume that $u \mapsto \Phi(u^{1/2})$ is *concave*.

(b) As an application of this result, one can show that there is an f that is integrable on $\mathbf{R}^2$, and is also in $L(\log L)^{1-\varepsilon}$,[‡] $\varepsilon > 0$, so that

$$\limsup_{\operatorname{diam}(R)\to 0} |R|^{-1} \int_R f(x-y)\, dy = \infty \qquad \text{a.e. } x,$$

as R ranges over the rectangles centered at the origin whose sides are parallel to the coordinate axes.

The original construction of an example such as in (b) can be found in Saks [1935]; for (a), see Stein [1961]. Note that the condition that $u \mapsto \Phi(u^{1/2})$ be concave generalizes the restriction $p \le 2$.

3.7 Let Γ be the circular cone $\{y : y_3 > (y_1^2+y_2^2)^{1/2}\} \subset \mathbf{R}^3$, and let P_y^Γ be the corresponding Poisson integral (as given in §2.4.1). For any vector $u \in \mathbf{R}^3$, let P^u be the one-dimensional Poisson integral along the direction u:

$$(P^u f)(x) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{f(x-ut)}{1+t^2}\, dt.$$

(a) If u is in the boundary of Γ, then

$$\lim_{\substack{y\in\Gamma\\ y\to u}} (P_y^\Gamma f)(x) = (P^u f)(x)$$

for every continuous, compactly supported function f.

[‡] That is, $f \in L^\Phi$, where $\Phi(u) \approx u$ as $u \to 0$, and $\Phi(u) \approx u(\log u)^{1-\varepsilon}$ as $u \to \infty$.

(b) Related to this is the following inequality. If

$$f(x_1, x_2, x_3) = f'(x_1, x_2)\chi(x_3),$$

where $f' \ge 0$ and χ is the characteristic function of the unit interval, then

$$(P_y^\Gamma f)(0) \ge \frac{c}{|R|} \int_R f'(x_1, x_2)\, dx_1\, dx_2, \qquad y \in \Gamma,$$

where $R \subset \mathbf{R}^2$ is the rectangle whose major axis is in the direction $y' = (y_1, y_2)$, having side lengths $s_1 = |y|$ and $s_2 = y_3 - |y'|$.

(c) As a result, $(P_* f)(x) \ge c(M_{\mathcal{R}} f')(x_1, x_2)\chi(x_3)$, and P_*^Γ is not bounded on any $L^p(\mathbf{R}^3)$, $p < \infty$.

For the above (and some generalizations) see Stein and N. J. Weiss [1969], N. J. Weiss [1972].

3.8 The thrust of §3.7, in connection with the remarks in §2.4.3 above, may be thought of as an illustration of the parallelism that exists between certain results for spheres in $\mathbf{R}^n$ and cones in $\mathbf{R}^{n+1}$. Other examples of this parallelism abound: see, e.g., Chapter 8, §5.16(b) and Chapter 11, §4.12; this kind of analogy is also stressed in Stein [1979].

C. Some positive results for maximal operators

3.9 The discussion in §2 and in §3.3 above shows that positive results cannot be expected when one considers maximal averages with respect to centered rectangles having arbitrary orientations. However, if we restrict appropriately the possible directions of the rectangles, then it turns out that matters can be markedly different. A simple example of this difference arises when we consider the collection of rectangles with sides parallel to the coordinate axes, for which we saw (in Chapter 2, §5.20) that there was an L^p theory when $p > 1$. In what follows, we shall consider the more complex situation when the directions of the rectangles are restricted to a discrete set, determined by angles that approach a limiting direction in a geometric (i.e., "lacunary") manner. The main interest of this example (for us) is that it anticipates one of the principal themes of the following chapter: that ideas of orthogonality (implemented via square functions) can play a decisive role in the study of maximal functions.

For simplicity, we restrict ourselves to the following typical situation. We let $\mathcal{R}_\Delta$ denote the set of all rectangles in $\mathbf{R}^2$ that are centered at the origin and whose major axis makes an angle θ_k with the x-axis, for some $k \in \mathbf{N}$, with $\tan\theta_k = 2^{-k}$. Write

$$(M_{\mathcal{R}_\Delta} f)(x) = \sup_{R\in\mathcal{R}_\Delta} |R|^{-1} \int_R |f(x-y)|\, dy.$$

We assert that

$$\|M_{\mathcal{R}_\Delta}(f)\|_{L^p(\mathbf{R}^2)} \le A_p \|f\|_{L^p(\mathbf{R}^2)}, \qquad \text{for } 1 < p \le \infty.$$

The proof of this result can be given in a series of steps, which we now outline.

(i) We may assume $f \geq 0$. Recall the maximal operator $M^{(u)}$ on $\mathbf{R}^2$, which arises by taking the usual one-dimensional maximal operator M along the vector $u \in \mathbf{R}^2$; that is,[†]

$$(M^{(u)}f)(x) = \sup_{h>0} \frac{1}{2h} \int_{-h}^{h} f(x-ut)\, dt.$$

Our first observation is that it suffices to prove that

$$f \mapsto \sup_k M^{(u_k)}(f)$$

is bounded on L^p, where $u_k = (1, 2^{-k})$, $k = 0, 1, \ldots$. This holds because, if R is a centered rectangle whose major axis makes an angle θ_k with the x-axis, then (as is easily seen)

$$|R|^{-1} \int_R f(x-y)\, dy \leq cM_y(M^{(u_k)}f)(x),$$

where M_y is the one-dimensional maximal operator $M^{(0,1)}$ along the y-axis.

Next, it clearly suffices to replace $M^{(u_k)}(f)$ by $\sup\limits_{h>0} A_h^k(f)$, with

$$(A_h^k f)(x) = \int_{-\infty}^{\infty} f(x-u_k t)\, \psi_h(t)\, dt, \qquad \psi_h(t) = h^{-1}\psi(t/h);$$

here $\psi \in C_0^\infty(\mathbf{R}^1)$ is nonnegative and $\psi \equiv 1$ near the origin. Thus matters are reduced to estimating

$$\sup_{\substack{k\in\mathbf{N}\\ h>0}} A_h^k(f).$$

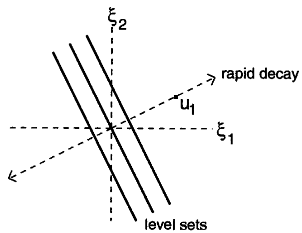

Figure 7. Behavior of $\widehat{A_1^1}$.

[†] See Chapter 2, §4.1; the restriction there that $|u| = 1$ is immaterial here.

(ii) We now restrict our attention to the case $p = 2$ (the situation when $p < 2$ will be treated in §3.10 immediately below), and we use the Fourier transform. Notice that

$$\widehat{A_h^k f}(\xi) = m(\xi \cdot hu_k)\widehat{f}(\xi),$$

where $m(s) = \int_{\mathbf{R}^1} e^{-2\pi ist}\psi(t)\, dt$ is the one-dimensional Fourier transform of ψ. Thus the multiplier $\widehat{A_h^k}$ corresponding to A_h^k is "good" (i.e., is rapidly decreasing) for ξ in the direction u_k, but is "bad" (has no decrease) in the orthogonal direction (see Figure 7). For convenience, we define

$$\Lambda^k(\xi) = \Lambda^k(\xi_1, \xi_2) = (\xi_1, 2^{-k}\xi_2);$$

note that $\widehat{A_h^k}(\xi) = \widehat{A_0^1}(h\Lambda^k(\xi)) = m(h\Lambda^k(\xi)\cdot u_0)$. These observations lead us to decompose the spectrum of $A_h^k(f)$ as follows.

Consider the cones

$$\Gamma = \{\xi \in \mathbf{R}^2 : |\xi \cdot \mathbf{1}| < c|\xi|\} \qquad \text{and} \qquad \Gamma_* = \{\xi \in \mathbf{R}^2 : |\xi\cdot\mathbf{1}| < 2c|\xi|\};$$

here c is a small constant and $\mathbf{1} = (1,1) = u_0 \in \mathbf{R}^2$. Similarly, for any $k \in \mathbf{N}$, define

$$\Gamma^k = \Lambda^{-k}(\Gamma) = \{\xi : (\xi_1, 2^k\xi_2) \in \Gamma\}, \qquad \Gamma_*^k = \Lambda^{-k}(\Gamma_*).$$

There are two key facts about the Γ^k (see Figure 8):

1. Every level set of $\widehat{A_h^k}$ is, except for a bounded interval, contained in Γ^k.
2. The cones $\{\Gamma_*^k\}_{k=0}^\infty$ do not overlap, if c is taken to be sufficiently small.

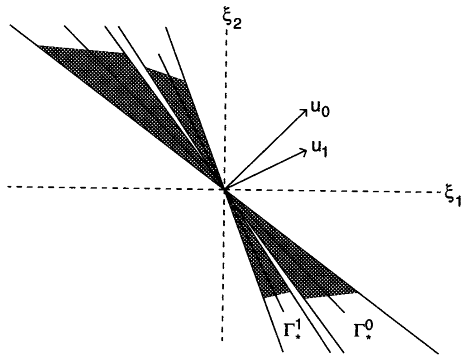

Figure 8. Cones and vectors normal to their axes.

Now let $\omega_\pm(\xi)$ be homogeneous functions of degree 0, smooth away from the origin, with

$$1 \equiv \omega_+(\xi) + \omega_-(\xi), \qquad \operatorname{supp}(\omega_+) \subset \Gamma_*, \qquad \text{and} \quad \operatorname{supp}(\omega_-) \subset {}^c\Gamma.$$

Choose also a cut-off function $\eta \in C_0^\infty(\mathbf{R}^2)$ with $\eta \equiv 1$ near the origin. Then

$$\begin{aligned}\widehat{A_0^1}(\xi) = m(\xi \cdot \mathbf{1}) = m(\xi \cdot \mathbf{1})\,\eta(\xi) + m(\xi\cdot\mathbf{1})\,[1-\eta(\xi)]\,\omega_-(\xi)\\ - m(\xi\cdot\mathbf{1})\,\eta(\xi)\,\omega_+(\xi) + m(\xi\cdot\mathbf{1})\,\omega_+(\xi).\end{aligned}$$

Thus we may write

$$\widehat{A_0^1}(\xi) = \widehat{\Phi}(\xi) + \widehat{\Psi}(\xi)\,\omega_+(\xi) + \widehat{A_0^1}(\xi)\,\omega_+(\xi)$$

with $\Phi, \Psi \in \mathcal{S}$. Precomposing with $h\Lambda^k$ gives

$$A_h^k(f) = (f * \Phi_{h,2^{-k}h}) + (f_k * \Psi_{h,2^{-k}h}) + A_h^k(f_k), \tag{$*$}$$

where $\widehat{f_k}(\xi) = \omega_+(2^k\xi_1, \xi_2)\widehat{f}(\xi)$, and

$$\Phi_{t_1,t_2}(x) = \Phi_{t_1,t_2}(x_1,x_2) = \frac{1}{t_1t_2}\Phi\Big(\frac{x_1}{t_1}, \frac{x_2}{t_2}\Big),$$

with a similar definition for Ψ_{t_1,t_2}.

(iii) The term $f * \Phi_{h2^k,h}$ in $(*)$ is controlled by the strong maximal operator M_S, as is described in Chapter 2, §5.20–§5.22. To take care of the second and third terms of $(*)$, we use orthogonality and square functions, via the observation that

$$\sum_k \|f_k\|_{L^2}^2 \leq c\|f\|_{L^2}^2;$$

this follows from Plancherel's theorem and the disjointness of the cones Γ_*^k.

Now $\sup_h |A_h^k(f)| \leq cM^{(u_k)}(f_k)$, and so

$$[\sup_{h,k} A_h^k(f_k)]^2 \leq c^2 \sum_k (M^{(u_k)} f_k)^2.$$

Therefore $\|M^{(u_k)}(f_k)\|_{L^2} \leq c\|f_k\|_{L^2}$ implies that

$$\int_{\mathbf{R}^2} [\sup_{h,k} A_h^k(f_k)]^2\, dx \leq c^2 \sum_k \|f_k\|_{L^2}^2 \leq c'\|f\|_{L^2}^2.$$

The second term in $(*)$ is handled in the same way, except here one uses M_S rather than $M^{(u_k)}$. This concludes the proof in the case $p = 2$.

3.10 We continue our sketch of the proof that

$$\|M_{\mathcal{R}_\Delta}(f)\|_{L^p(\mathbf{R}^2)} \leq A_p\|f\|_{L^p(\mathbf{R}^2)}, \qquad 1 < p \leq \infty,$$

and now turn to the case $p < 2$.

It will be based on the same analysis as that given above for $p = 2$ (in particular, on the decomposition (ii)), together with the following two additional facts.

(a) First, the square function inequality

$$\Big\|\Big(\sum_k |f_k|^2\Big)^{1/2}\Big\|_{L^p(\mathbf{R}^2)} \leq A_p\|f\|_{L^p(\mathbf{R}^2)}, \qquad 1 < p < \infty,$$

where $\widehat{f_k}(\xi) = \omega_+(2^k\xi_1, \xi_2)\widehat{f}(\xi)$.

(b) A "bootstrap" argument: If we know that

$$f \mapsto \sup_k M^{(u_k)}(f)$$

is bounded on L^q, then

$$\Big\|\Big(\sum_k |M^{(u_k)}(f_k)|^2\Big)^{1/2}\Big\|_{L^p} \leq A_p\Big\|\Big(\sum_k |f_k|^2\Big)^{1/2}\Big\|_{L^p}$$

when $\dfrac{1}{p} < \dfrac{1}{2}\Big(1 + \dfrac{1}{q}\Big)$.

Assuming both (a) and (b), the argument could proceed essentially as for the case $p = 2$ in §3.9. Indeed, the first term in $(*)$ there is controlled by M_S and hence is bounded on L^p, for any $p > 1$. Since we already know the result for $\sup_k M^{(u_k)}(f)$ when $q = 2$, the conclusion (b) (together with (a)) guarantees that

$$\Big\|\Big(\sum_k |M^{(u_k)}(f_k)|^2\Big)^{1/2}\Big\|_{L^p} \leq A_p'\|f\|_{L^p}$$

whenever $p > 4/3$. One also has

$$\Big\|\Big(\sum_k |M_S(f_k)|^2\Big)^{1/2}\Big\|_{L^p} \leq A_p\Big\|\Big(\sum_k |f_k|^2\Big)^{1/2}\Big\|_{L^p},$$

by using the vector-valued maximal inequality (Theorem 1 in Chapter 2, §1.1) first in the x-variable, and then in the y-variable.

In summary we see that, assuming (a) and (b), we can obtain our theorem for $p > 4/3$ from $p = 2$, and then for p from q where $\dfrac{1}{p} < \dfrac{1}{2}\Big(1 + \dfrac{1}{q}\Big)$; that is, successively for

$$p > 4/3,\ p > 8/7, \ldots,\ p > 2^j/(2^j - 1), \ldots$$

and so for all $p > 1$.

It remains to prove (a) and (b). As to (a), consider for each t the multiplier

$$\Omega(\xi_1,\xi_2) = \sum_{k\geq 1} r_k(t)\,\omega_+(2^k\xi_1,\xi_2),$$

where $\{r_k\}$ are the Rademacher functions. It is easy to verify that

$$|\xi_2^{\ell_2}\,\xi_1^{\ell_1}\,\partial_{\xi_2}^{\ell_2}\,\partial_{\xi_1}^{\ell_1}\Omega(\xi_1,\xi_2)| \leq A_\ell$$

for each $\ell = (\ell_1,\ell_2)$. Thus Ω satisfies the conditions of the "two-parameter" Marcinkiewicz multiplier theorem.‡ Hence

$$\left\|\sum r_k(t) f_k\right\|_{L^p(\mathbf{R}^2)} \leq A_p\|f\|_{L^p(\mathbf{R}^2)}$$

for all t, and the standard inequality for Rademacher functions establishes (a).

To prove (b), consider the more general inequality

$$\left\|\Big(\sum_k |M^{(u_k)}(f_k)|^r\Big)^{1/r}\right\|_{L^s(\mathbf{R}^2)} \leq A_s\left\|\Big(\sum_k |f_k|^r\Big)^{1/r}\right\|_{L^s(\mathbf{R}^2)}.$$

The case when $1 < s = r < \infty$ is of course a restatement of the standard inequality (see also (47) of Chapter 2). The case $r = \infty$, $s = q$ is a restatement of the hypothesis of the bootstrap argument. The conclusion with $r = 2$ and $\frac{1}{p} < \frac{1}{2}\left(1 + \frac{1}{q}\right)$ then follows by interpolation, concluding our sketch.

The original theorem involving differentiation in lacunary directions was proved by the use of covering lemmas in Strömberg [1976], and Córdoba and R. Fefferman [1977a], but it was limited to the case $p \geq 2$ (for a related covering lemma, see §3.11(d) below).

The result above, for the full range $p > 1$, as well as some generalizations, may be found in Nagel, Stein, and Wainger [1978]. The bootstrap argument that is used can be applied in some other circumstances; see, e.g., Duoandikoetxea and Rubio de Francia [1986]; Carlsson, Christ, Córdoba, Duoandikoetxea, Rubio de Francia, Vance, Wainger, and Weinberg [1986]. Some further results dealing with lacunary differentiation in higher dimensions are in Carbery [1988].

3.11 The following estimates may be thought of as additional substitute results for the maximal operator $M_{\mathcal{R}}$, when some control is imposed on the rectangles that occur. For each integer N, $N \geq 2$, we consider two sets of centered rectangles in $\mathbf{R}^2$. The first, $\mathcal{R}_N$, consists of all rectangles (of arbitrary orientation) whose "eccentricity" (the length of the long side divided by the length of the short side) is N. The second, $\widehat{\mathcal{R}}_N$, consists of rectangles (having arbitrary side lengths) whose major axis makes an angle $\frac{2\pi j}{N}$, $j = 0,\ldots,N-1$, with the x-axis. Let

$$(M_{\mathcal{R}_N}f)(x) = \sup_{R\in\mathcal{R}_N} |R|^{-1}\int_R |f(x-y)|\,dy,$$

with a similar definition for $M_{\widehat{\mathcal{R}}_N}$. Then

‡ As in *Singular Integrals*, Chapter 4, §6.

(a) $\|M_{\mathcal{R}_N}f\|_{L^2(\mathbf{R}^2)} \leq A\log N\,\|f\|_{L^2(\mathbf{R}^2)}$.

(b) $\|M_{\widehat{\mathcal{R}}_N}f\|_{L^2(\mathbf{R}^2)} \leq A\log N\,\|f\|_{L^2(\mathbf{R}^2)}$.

(c) $|\{x : (M_{\widehat{\mathcal{R}}_N}f)(x) > \alpha\}| \leq A\log N\left(\frac{\|f\|_{L^2(\mathbf{R}^2)}}{\alpha}\right)^2$, all $\alpha > 0$.

(d) The result (c) is a consequence of the following covering lemma. Suppose $\{R_\gamma\}$ is a finite collection of translates of rectangles taken from $\widehat{\mathcal{R}}_N$. Then there is a subcollection $\{R_j\} \subset \{R_\gamma\}$ so that

$$|\bigcup_\gamma R_\gamma| \leq A\log N\cdot|\bigcup_j R_j| \quad\text{and}\quad \int_{\mathbf{R}^2}\Big(\sum_j \chi_{R_j}\Big)^2 dx \leq A|\bigcup_j R_j|.$$

It can be seen that the weak-type inequality (c) implies (b), by a (two-fold) application of the Marcinkiewicz interpolation theorem. The estimate (b) immediately gives (a), since

$$(M_{\mathcal{R}_N}f)(x) \leq c(\mathcal{M}_{\widehat{\mathcal{R}}_N}f)(x).$$

For further details, see Córdoba [1977], Strömberg [1978].

Without using covering lemmas, one can prove (a) and (b) by the use of square functions, in the same spirit as in §3.9. We shall now give a sketchy indication of the argument used for (a) along these lines.

We need only control the averages over those rectangles that have eccentricity N, but whose major axes make angles θ_k with the x-axis, with $\tan\theta_k = k/N$, $k = 0,\ldots,N-1$. For this purpose, let $u_k = (N,k)$, $k = 0,\ldots,N-1$, and let $e_2 = (0,1) \in \mathbf{R}^2$. It suffices to consider the averages

$$(A_h^{k,N}f)(x) = \int\int_{\mathbf{R}^2} f(x-u_kt-e_2s)\,\psi_h(t)\,\psi_h(s)\,ds\,dt, \qquad h > 0,$$

where ψ_h is as in §3.9. For $f \geq 0$,

$$(M_{\mathcal{R}_N}f)(x) \approx \sup_{\substack{0\leq k<n\\ h>0}} (A_h^{k,N}f)(x).$$

Now consider the square functions g_j^N, defined by

$$[(g_0^Nf)(x)]^2 = \int_0^\infty \sum_{k=0}^{N-1} |(A_{2h}^{k,N}f)(x) - (A_k^{2k,2N}f)(x)|^2\frac{dh}{h},$$

$$[(g_1^Nf)(x)]^2 = \int_0^\infty \sum_{k=0}^{N-1} |(A_{2h}^{k,N}f)(x) - (A_k^{2k+1,2N}f)(x)|^2\frac{dh}{h}.$$

The desired result follows from two observations. The first is that when $N = 2^m$,

$$(M_{\widehat{\mathcal{R}}_N}f)(x) \leq c\Big\{(Mf)(x) + \sum_{\ell=0}^m[(g_0^{2^\ell}f)(x) + (g_1^{2^\ell}f)(x)]\Big\};$$

here M is the usual Hardy-Littlewood maximal operator. The second is the estimate

$$\|g_0^N(f)\|_{L^2(\mathbf{R}^2)} \leq A\|f\|_{L^2(\mathbf{R}^2)},$$

with A independent of N, together with a similar assertion for g_1^N.

Conclusion (b) can in fact be deduced from (a) by a decomposition of the spectrum of f into N narrow cones; this is similar to the approach used in §3.9. Arguments of the above kind may be found in Wainger [1979]. Some extensions of (a) are in §3.1^a and §3.12(d).

3.12 As a variant of the results in §3.11, we consider the maximal function

$$(\mathcal{M}^{(q)}f)(x) = \left(\int_{\mathbf{S}^{n-1}} [(M^{(u)}f)(x)]^q \, d\sigma(u)\right)^{1/q}, \quad n \geq 2,$$

where $M^{(u)}$ is the one-dimensional maximal operator in the direction u (as in §3.9(i)). Note that the case $q = 1$ arises in the method of rotations discussed in Chapter 2, §4.1. The inequality

$$\|\mathcal{M}^{(1)}f\|_{L^p(\mathbf{R}^n)} \leq A\|f\|_{L^p(\mathbf{R}^n)}, \quad p > 1,$$

is implicit in that discussion. Here we treat the generalization

$$\|M^{(q)}f\|_{L^p(\mathbf{R}^n)} \leq A_{p,q}\|f\|_{L^p(\mathbf{R}^n)} \tag{$*$}$$

for $q \geq 1$; when $p = \infty$, it holds for all q. One can assert the following.

(a) When $n = 2$ and $q < \infty$, $(*)$ holds for all $p \geq 2$.

(b) When $n \geq 2$ and $q = \infty$, $(*)$ fails for all $p < \infty$.

(c) When $n \geq 3$, $(*)$ holds for $p = (n+1)/2$ and $q < n+1$. Interpolation of this result with the cases $p = \infty$ and $q = 1$ yields a range of (p, q) for which $(*)$ holds.

(d) There is an analogue of §3.11(a) for $n \geq 3$, which holds for $p = (n+1)/2$.

See R. Fefferman [1983], Christ, Duoandikoetxea, and Rubio de Francia [1986]. Assertion (b) is a simple consequence of the Besicovitch construction, and is closely related to the corresponding statements in §2.2. Here the positive results depend in a key way on the case $p = 2$, for which corresponding inequalities are proved by using square functions.

Notes

§1. For the history of the Besicovitch construction and its relation with the Kakeya problem, see Falconer [1985]. Our presentation is based on the standard construction given there (see also de Guzmán [1975]) but, in addition, uses some observations in C. Fefferman [1971b].

§2. The general maximal principle originates in Stein [1961]; see also the further literature cited in §3.4 and §3.5 above. Corollary 1 is implicit in Busemann and Feller [1934]. The counterexample for the Poisson integral for the circular cone in §2.4.2 is (in a slightly different form) in Stein and N. J. Weiss [1969]; the counterexample for the ball multiplier is due to C. Fefferman [1971b].

CHAPTER XI

Maximal Averages and Oscillatory Integrals

The theory of maximal averages, as presented in chapters 1 and 2, depended essentially on the covering lemmas of Vitali type. The further idea we want to develop now is the decisive role that L^2 theory—in particular, orthogonality and oscillatory integrals—can play in this context. Here these notions are best exploited by the use of square functions. In the translation-invariant case, what is involved are certain decay estimates for the Fourier transform, which at bottom reflect some kind of "curvature" for the situation at hand.

From the general point of view of maximal operators and their L^p estimates as set out in the first paragraphs of the previous chapter, the main difficulty occurs if the collection $\mathcal{C}$ (over which we are averaging) consists of "thin" sets. To come to grips with this problem, we consider the limiting situation that arises when we think of $\mathcal{C}$ as consisting of certain subvarieties of $\mathbf{R}^n$. We discuss two instances of this having particular interest.

The first is where $\mathcal{C}$ consists of the "initial segments" of a fixed k-dimensional submanifold of $\mathbf{R}^n$; that is, when we consider the family of averages

$$(A_r f)(x) = \frac{1}{r^k}\int_{|t|\leq r} f(x - \gamma(t))\, dt, \qquad 0 < r \leq 1,$$

where γ is a smooth mapping of (say) the unit ball in $\mathbf{R}^k$ to $\mathbf{R}^n$.

The second instance occurs when the submanifolds in $\mathcal{C}$ are variable, and the special case in which they come about by dilations of a fixed variety. The basic example here is the family of spherical averages in $\mathbf{R}^n$, given by

$$f \mapsto \int_{|y|=1} f(x - ry)\, d\sigma(y),$$

as r varies. What unifies the study of the two general types of averages we are considering is the crucial role of curvature, which results in certain similarities in some of the techniques used. However, there are also significant differences; it is these similarities and differences we now explain.

In the first case, our assumption is that the underlying manifold is "curved", in the sense that it is of finite type at the point corresponding to $t = 0$. As we have seen in Chapter 8, this implies that the Fourier transform of a suitable measure carried on this submanifold has a certain (power) decay. This decay allows one to give an L^2 estimate for an appropriate square function, which in turn yields the case $p = 2$ of the estimate for the maximal operator in question. The L^p theory, $p \neq 2$, is actually implicit in this approach. Indeed, an examination of the L^2 argument shows that there is some "slack" in it, and that a "rougher" version of our maximal operator is still bounded on L^2. One can then combine this with a "smoother" variant, based on the more standard maximal theory of Chapter 2, which is bounded on all L^p, $1 < p \leq \infty$, and produce the desired result.

We gave a schematic description of the proof, but have left out an important consideration: an inherent nonisotropic homogeneity. This feature is most easily understood when we consider a model problem, in which $\gamma(t)$ is a (vector-valued) polynomial function of t; it is best exploited through the use of a "lifting" technique. The latter embeds our $\mathbf{R}^n$ in an $\mathbf{R}^N$ of higher dimension, for which the different monomials we consider are "free".

Turning to the second problem (that of spherical averages and their generalizations), we may say that what is determining here is the *degree* of decay of the Fourier transform (which is $O(|\xi|^{(1-n)/2})$ in the case of the sphere $\mathbf{S}^{n-1} \subset \mathbf{R}^n$). It turns out that a desired L^2 estimate holds whenever the exponent of $|\xi|$ is $< -1/2$ (e.g., for spheres, when $n \geq 3$). So, following an analogous route, one gets the maximal inequality for $p > n/(n-1)$, $n \geq 3$. Further, more refined arguments allow one to obtain the result for $n = 2$, and $p > 2$.

Instead of limiting our exposition for spherical averages to the above context, we present its natural extension in a diffeomorphism-invariant setting where translation invariance plays no role. The formulation of this generalization requires that the assumption of nonvanishing Gaussian curvature (the key property used in the case of the sphere) be replaced by the notion of nonvanishing *rotational curvature*. This concept, which is inherent in the problem, is also related to the Fourier integral operators treated in Chapter 9. Incidentally, exploiting this connection is one approach to our problem but we do not pursue it, using instead a method that admits generalization to the case in which the curvature is allowed to vanish (to finite order) at any point.

From this brief description, the reader can see some of the differences between the two types of maximal operators we consider. Another difference is that the diffeomorphism-invariant version of the averages when the variety is "fixed" (i.e., the analogue of A_r) does not seem to allow a treatment by oscillatory integrals. Instead, considerations related

to approximation by homogeneous groups have to be used. For further discussion of this point, see §4.7 below.

We make a final remark about the role of curvature in these problems. In the first setting, there is a submanifold, which is not of finite type at the origin, for which the maximal operator fashioned from the A_r is not bounded on any L^p, $p < \infty$. A similar assertion holds for the maximal operator associated to the dilations of a hypersurface whose curvature is allowed to vanish to infinite order at some point.

1. *Maximal averages and square functions*

In this section we shall introduce the general point of view by describing two simple examples of the approach which already contain some of the essential aspects. The first illustration deals with spherical averages.

1.1 Spherical averages when $n \geq 4$. Let $d\sigma$ denote the measure on the sphere $\mathbf{S}^{n-1} \subset \mathbf{R}^n$ induced by Lebesgue measure on $\mathbf{R}^n$, normalized to have total mass one.† We then define the average value of f over the sphere of radius t, centered at x, by

$$(A_t f)(x) = \int_{|y|=1} f(x - ty)\, d\sigma(y) = f * d\sigma_t, \tag{1}$$

for any continuous function f; here we have taken

$$\int g(x)\, d\sigma_t(x) = \int g(tx)\, d\sigma(x).$$

For more general f, we will ultimately be concerned with the question whether

$$(A_t f)(x) \to f(x) \quad \text{as } t \to 0, \qquad \text{for a.e. } x.$$

To begin with, it is far from obvious that $(A_t f)(x)$ is continuous in t, $0 < t < \infty$, for *any* x; or that

$$\sup_{t_1 < t < t_2} |A_t f(x)|$$

is even measurable in x.

For the moment we bypass these problems by considering *a priori* estimates.

PROPOSITION 1. *Suppose $n \geq 4$. Then we have the estimate*

$$\| \sup_{t>0} |(A_t f)(x)| \, \|_{L^2(\mathbf{R}^n)} \leq A \|f\|_{L^2(\mathbf{R}^n)} \tag{2}$$

for all continuous, compactly supported functions f.

Of course, the bound A is independent of f.

† This is a slight change from the notation used in Chapter 8, §3, where $d\sigma$ was not normalized. In the present context, the formulas are simplified if $\int_{\mathbf{S}^{n-1}} d\sigma = 1$.

1.1.1 The square function that intervenes here is given by

$$(Sf)(x) = \left(\int_0^\infty \left| \frac{\partial (A_t f)(x)}{\partial t} \right|^2 t \, dt \right)^{1/2}, \tag{3}$$

which is certainly well-defined if f is smooth (say C^1) and has compact support.

If M denotes the standard maximal operator, given in terms of averages over centered balls, the proof of the proposition (for smooth functions) follows from two easily established assertions: first, the pointwise estimate

$$\sup_{t>0} |(A_t f)(x)| \le (Mf)(x) + c(Sf)(x), \tag{4}$$

where c is an appropriate constant; and second, the fact that when $n \ge 4$,

$$\|Sf\|_{L^2} \le A\|f\|_{L^2}. \tag{5}$$

To prove (4), note that when $t > 0$,

$$\begin{aligned}(A_t f)(x) &= t^{-n} \int_0^t \frac{\partial}{\partial s}[s^n (A_s f)(x)] \, ds \\ &= t^{-n} \cdot n \int_0^t s^{n-1} (A_s f)(x) \, ds + t^{-n} \int_0^t s^n \frac{\partial}{\partial s}[(A_s f)(x)] \, ds \\ &= I_1 + I_2.\end{aligned}$$

Observe that I_1 equals the average of f over the ball, centered at x, of radius t. We estimate the second term by Schwarz's inequality. Indeed,

$$|I_2| \le \left(\int_0^\infty s \left| \frac{\partial (A_s f)(x)}{\partial s} \right|^2 ds \right)^{1/2} \cdot t^{-n} \left(\int_0^t s^{2n-1} \, ds \right)^{1/2} \le cSf(x),$$

with $c = (2n)^{-1/2}$. Taking the supremum over t gives us (4).

1.1.2 To prove (5), note that $A_s f$ is the convolution of f with $d\sigma_s$. In terms of the Fourier transform, this becomes

$$\widehat{A_s f}(\xi) = \widehat{f}(\xi) \cdot \widehat{\sigma}(s\xi).$$

Therefore

$$\left(\frac{\partial A_s f}{\partial s} \right)^{\wedge}(\xi) = \frac{\mu(s\xi)}{s} \cdot \widehat{f}(\xi), \qquad \text{where} \quad \mu(\xi) = \sum_{j=1}^n \xi_j \frac{\partial \widehat{\sigma}}{\partial \xi_j}.$$

The crucial fact is that

$$|\mu(\xi)| \le \min\{A'|\xi|, A'|\xi|^{-(n-3)/2}\} \tag{6}$$

for an appropriate constant A'.

Indeed, since

$$\frac{\partial \widehat{\sigma}(\xi)}{\partial \xi_j} = -2\pi i \int_{\mathbf{S}^{n-1}} e^{-2\pi i x \cdot \xi} x_j \, d\sigma(x),$$

we have that $|\partial \widehat{\sigma}(\xi)/\partial \xi_j| \le A'(1 + |\xi|)^{-(n-1)/2}$ by the Bessel function representation given at the beginning of §3 in Chapter 8, together with the asymptotics in §1.4; or, more directly, we can appeal to Theorem 1 in §3.1 of that chapter. In either case, (6) is established.

Next, by Plancherel's theorem,

$$\int_{\mathbf{R}^n} \left| \frac{\partial (A_s f)(x)}{\partial s} \right|^2 dx = \int_{\mathbf{R}^n} |\widehat{f}(\xi)|^2 \frac{|\mu(s\xi)|^2}{s^2} \, d\xi,$$

so an integration in s gives

$$\int_{\mathbf{R}^n} |(Sf)(x)|^2 \, dx = \int_{\mathbf{R}^n} |\widehat{f}(\xi)|^2 \left(\int_0^\infty |\mu(s\xi)|^2 \frac{ds}{s} \right) d\xi. \tag{7}$$

However

$$\int_0^\infty |\mu(s\xi)|^2 \frac{ds}{s} \le A \left(|\xi|^2 \int_0^{|\xi|^{-1}} s \, ds + |\xi|^{-(n-3)} \int_{|\xi|^{-1}}^\infty s^{-(n-2)} \, ds \right)$$

because of (6). The second integral on the right converges when $n-2 > 1$, and is then bounded by a constant multiple of $|\xi|^{n-3}$. Altogether then

$$\int_0^\infty |\mu(s\xi)|^2 \frac{ds}{s} \le A,$$

and inserting this in (7) yields the desired estimate for the square function.

Finally note that once the *a priori* inequality (2) has been established for compactly supported smooth functions, it holds as well for functions that are merely continuous, as a simple limiting argument shows. The proposition is therefore proved.

1.1.3 Remarks. (a) As we will see below, analogous results hold for $n \ge 2$; we then have the boundedness of the maximal operator $f \mapsto \sup_{t>0} |A_t f|$ on L^p, whenever $p > n/(n-1)$.

(b) On the other hand, it is easy to see that nothing of this kind can hold for $n = 1$ and $p < \infty$, or for $n \ge 2$ and $p \le n/(n-1)$. In fact, note that in the one-dimensional case

$$A_t f(x) = \frac{1}{2}[f(x+t) + f(x-t)];$$

thus if f is positive and unbounded (near the origin, say) then

$$\sup_t A_t f(x) = \infty$$

everywhere, and one can choose such an f so that it belongs to all L^p, $p < \infty$. Now in $\mathbf{R}^n$, $n \geq 2$, choose f so that

$$f(y) = \frac{|y|^{1-n}}{\log(1/|y|)} \qquad \text{near the origin,}$$

and f vanishes for large y. Then, for any x, $A_t f(x)$ is unbounded for t near $|x|$, so $\sup_t A_t f(x) = \infty$ everywhere, while clearly $f \in L^p(\mathbf{R}^n)$ if $p \leq n/(n-1)$.

(c) In reality, all of the above results are closely related to the curvature of the sphere. Indeed, if S is a smooth hypersurface with nonvanishing Gaussian curvature, then estimates like (6) are valid because of Theorem 1 in Chapter 8, and all of the conclusions above continue to hold in this more general setting. For this see §3 below. However, if the curvature of S vanishes, these assertions may fail utterly; see §4.2.

1.2 An example of averages with respect to a (fixed) curve. In the situation of the spherical maximal function just considered, we averaged our function f over a continuous family of submanifolds approaching the point x. In the next example, given a point x, the manifold will be fixed, but the portion over which we integrate will be allowed to vary.

The simplest case of this occurs in $\mathbf{R}^2$, when we are dealing with a smooth curve, which is given by a mapping $\gamma : \mathbf{R}^1 \to \mathbf{R}^2$. We then consider the averages

$$\frac{1}{2h}\int_{-h}^{h} f(x-\gamma(t))\,dt, \qquad h > 0,$$

and the corresponding maximal function

$$\mathcal{M}f(x) = \sup_{h>0} \frac{1}{2h}\left|\int_{-h}^{h} f(x-\gamma(t))\,dt\right|. \tag{8}$$

The example we consider here is that of the parabola $\gamma(t) = (t, t^2)$. Again, to obviate technical difficulties, we turn our attention to (8) and make the *a priori* assumption that f is continuous and has compact support.

PROPOSITION 2. *Let γ be the parabola above and let $\mathcal{M}$ be the maximal operator given by* (8). *Then there is a bound A so that*

$$\|\mathcal{M}f\|_{L^2(\mathbf{R}^2)} \leq A\|f\|_{L^2(\mathbf{R}^2)}, \tag{9}$$

whenever f is continuous and has compact support.

1.2.1 In proving (9), it is clear that we may assume that f is nonnegative; also, in contrast to the example of the spherical maximal function, we can reduce matters to a discrete (dyadic) analogue. Indeed, let η be a fixed smooth, nonnegative function on $\mathbf{R}^1$ with $\eta(t) = 1$ for $|t| \leq 1$, and $\eta(t) = 0$ for $|t| \geq 2$. For each integer j, set

$$A_j f(x) = 2^j \int f(x-\gamma(t))\,\eta(2^j t)\,dt = \int f(x_1 - 2^{-j}t, x_2 - 2^{-2j}t^2)\,\eta(t)\,dt. \tag{10}$$

Then, obviously $A_j f(x) \geq 2^j \int_{|t|\leq 2^j} f(x_1 - t, x_2 - t^2)\,dt$, so

$$\mathcal{M}f(x) \leq \sup_j A_j f(x). \tag{11}$$

The one-dimensional averages $\{A_j\}$ will be compared with certain two-dimensional averages $\{B_j\}$, the latter bearing some structural resemblance to the former. In fact, the scaling property

$$\gamma(2^{-j}t) = (2^{-j}t, 2^{-2j}t^2)$$

leads us to define

$$B_j f(x) = \int\int_{\mathbf{R}^2} f(x_1 - 2^{-j}y_1, x_2 - 2^{-2j}y_2)\,\psi(y_1, y_2)\,dy_1\,dy_2,$$

where ψ is a fixed smooth function on $\mathbf{R}^2$ with compact support, normalized so that $\int_{\mathbf{R}^2} \psi\,dy = \int_{\mathbf{R}^1} \eta\,dt$. Also, note that

$$B_j f(x) = f * \psi_j, \qquad \text{where} \quad \psi_j(x_1, x_2) = 2^{3j}\psi(2^j x_1, 2^{2j} x_2).$$

Underlying these definitions is the one-parameter family of scalings

$$x \mapsto \delta \circ x = (\delta x_1, \delta^2 x_2) \qquad \text{for } \delta > 0.$$

Connected with this homogeneity is the real-variable structure described in Chapter 1, §2.3 and §2.4. In the present instance, the quasi-distance is given by

$$\rho(x,y) = \max\{|x_1 - y_1|, |x_2 - y_2|^{1/2}\},$$

so $\rho(x,y) = \rho(x-y)$, with $\rho(x) = \max\{|x_1|, |x_2|^{1/2}\}$. The balls are given by

$$B(x,\delta) = \{y : |y_1 - x_1| < \delta, |y_2 - x_2| < \delta^2\}.$$

All the postulates set forth in Chapter 1, §1 are satisfied by this collection of balls, if we take $d\mu$ to be the Lebesgue measure dx.

Let $\tilde{M}$ denote the maximal operator with respect to these balls, i.e.,

$$\tilde{M}f(x) = \sup_{\delta>0} \frac{1}{|B(\delta)|} \int_{B(\delta)} |f(x-y)|\,dy, \tag{12}$$

where $B(\delta) = B(0,\delta)$. Then Theorem 1 of Chapter 1 guarantees that

$$\|\tilde{M}f\|_{L^2(\mathbf{R}^2)} \leq A\|f\|_{L^2(\mathbf{R}^2)}. \tag{13}$$

Now suppose that ψ is supported in the unit "ball"

$$\{x : |x_1| \leq 1,\ |x_2| \leq 1\},$$

then

$$\begin{aligned} B_j f(x) &= 2^{3j} \int f(x_1 - y_1, x_2 - y_2)\,\psi(2^j y_1, 2^{2j} y_2)\,dy_1\,dy_2 \\ &\leq \frac{c}{|B(2^j)|} \int_{B(2^j)} f(x-y)\,dy \leq c\tilde{M}f(x). \end{aligned}$$

Therefore

$$\sup_j B_j f(x) \leq c\tilde{M}f(x). \tag{14}$$

1.2.2 The square function which is relevant to the problem at hand is given by

$$Sf(x) = \left(\sum_{j=-\infty}^{\infty} |A_j f(x) - B_j f(x)|^2 \right)^{1/2}. \tag{15}$$

Notice that because of (11), (14), and the fact that

$$\sup_j |A_j f(x) - B_j f(x)| \leq Sf(x),$$

we have that

$$\mathcal{M}f(x) \leq c\tilde{M}f(x) + Sf(x).$$

Thus if we apply (13), our proposition will be proved once we show that

$$\|Sf\|_{L^2(\mathbf{R}^2)} \leq A\|f\|_{L^2(\mathbf{R}^2)}. \tag{16}$$

To prove (16), let $d\mu$ be the measure on $\mathbf{R}^2$ defined by

$$\int f\,d\mu = \int f(\gamma(t))\,\eta(t)\,dt;$$

similarly define

$$\int f\,d\mu_j = \int f(\gamma(2^{-j}t))\,\eta(t)\,dt = \int f(2^{-j}t, 2^{-2j}t^2)\,\eta(t)\,dt.$$

Then $A_j f = f * d\mu_j$, while the Fourier transform of $d\mu_j$ is given by

$$\begin{aligned} \widehat{d\mu_j}(\xi) &= \int e^{-2\pi i(\xi_1 2^{-j}t + \xi_2 2^{-2j}t^2)}\eta(t)\,dt \\ &= \widehat{d\mu}(2^{-j}\xi_i, 2^{-2j}\xi_2) = \widehat{d\mu}(2^{-j} \circ \xi). \end{aligned}$$

Thus

$$\widehat{A_j f}(\xi) = \hat{f}(\xi) \cdot \widehat{d\mu}(2^{-j} \circ \xi). \tag{17}$$

Similarly, since

$$B_j f = f * \psi_j, \qquad \text{with} \quad \psi_j(x_1, x_2) = 2^{3j}\psi(2^j x_1, 2^{2j} x_2),$$

we have

$$\widehat{B_j f}(\xi) = \hat{f}(\xi) \cdot \hat{\psi}(2^{-j} \circ \xi). \tag{18}$$

Because ψ was normalized by the condition

$$\int_{\mathbf{R}^2} \psi(x)\,dx = \int_{\mathbf{R}^1} \eta(t)\,dt,$$

we see that $\widehat{d\mu}(0) = \hat{\psi}(0)$; moreover, both $d\mu$ and ψ have compact support, so it follows that $\widehat{d\mu}(\xi) - \hat{\psi}(\xi)$ is smooth. As a result

$$|\widehat{d\mu}(\xi) - \hat{\psi}(\xi)| \leq A|\xi| \leq A'\rho(\xi), \tag{19}$$

if $\rho(\xi) = \max\{|\xi_1|, |\xi_2|^{1/2}\} \leq 1$.

The parabola (t, t^2) is a submanifold of finite type, in the sense of Chapter 8, §3.2. In fact, since it has nonvanishing curvature at every point, its type is exactly 2. Thus, by Theorem 2 of that chapter, we have that

$$|\widehat{d\mu}(\xi)| \leq A|\xi|^{-1/2} \leq A'(\rho(\xi))^{-1/2}, \qquad \text{if } \rho(\xi) \geq 1.$$

Moreover, $\hat{\psi}(\xi)$ is rapidly decreasing at infinity, so

$$|\widehat{d\mu}(\xi) - \hat{\psi}(\xi)| \leq A(\rho(\xi))^{-1/2}, \qquad \text{if } \rho(\xi) \geq 1. \tag{20}$$

Finally,

$$\begin{aligned} \|Sf\|_{L^2}^2 &= \sum_j \|A_j f - B_j f\|_{L^2}^2 \\ &= \sum_j \int_{\mathbf{R}^2} |\hat{f}(\xi)|^2 \cdot |\widehat{d\mu}(2^{-j} \circ \xi) - \hat{\psi}(2^{-j} \circ \xi)|^2\,d\xi \end{aligned} \tag{21}$$

because of (17), (18), and Plancherel's theorem. However

$$\sum_j |\widehat{d\mu}(2^{-j}\circ\xi)-\widehat{\psi}(2^{-j}\circ\xi)|^2=\sum_{\rho(2^{-j}\circ\xi)\le 1}+\sum_{\rho(2^{-j}\circ\xi)>1}$$

For the first sum, we use (19), and note that $\rho(2^{-j}\circ\xi)=2^{-j}\rho(\xi)$. Therefore that sum is dominated by

$$A^2[\rho(\xi)]^2\sum_{\rho(\xi)\le 2^j}2^{-2j}\le A'.$$

Similarly, by (20) the second sum is majorized by a multiple of

$$[\rho(\xi)]^{-1}\sum_{2^j<\rho(\xi)}2^j,$$

which is again bounded.

If we insert this in (21), we obtain that $\|Sf\|_{L^2}^2$ is majorized by a constant multiple of $\int_{\mathbf{R}^2}|\widehat{f}(\xi)|^2\,d\xi$, which proves (16) and concludes the proof of the proposition.

2. Averages over a k-dimensional submanifold of finite type

In this section we want to present a wide-ranging generalization of the example for parabolas treated in §1.2 above. The setting is as follows: Suppose we are given a portion of a k-dimensional submanifold S in $\mathbf{R}^n$, with a distinguished point x^0 in S; we may assume that S is the image of a smooth mapping $t\mapsto\gamma(t)$, where t ranges over the unit ball in $\mathbf{R}^k$, and $x^0=\gamma(0)$.

We study the averages of functions f taken over the translates of neighborhoods of x^0 in S by forming the corresponding maximal function

$$\mathcal{M}f(x)=\sup_{0<r\le 1}\frac{1}{r^k}\left|\int_{|t|\le r}f(x-\gamma(t))\,dt\right|. \tag{22}$$

The key geometric assumption we make is that S is of finite type at the point $x^0=\gamma(0)\in S$; that is, at x^0, S has at most a finite order of contact with any hyperplane in $\mathbf{R}^n$ (as in Chapter 8, §3.2).

THEOREM 1. *If S and γ are as above, then for each p, $1<p\le\infty$, there is a bound A_p so that*

$$\|\mathcal{M}f\|_{L^p(\mathbf{R}^n)}\le A_p\|f\|_{L^p(\mathbf{R}^n)} \tag{23}$$

whenever f is continuous and has compact support.

2.1 Model cases. In order to clarify the proof of this theorem, we formulate two model problems. In the first, we restrict our attention to $\gamma(t)$ that are given as *polynomial* functions in t, i.e., where

$$\gamma(t)=P(t)=(P_1(t),\dots,P_n(t))$$

and the $P_j(t)$ are real-valued polynomials of $t\in\mathbf{R}^k$. We define $\mathcal{M}_P$ as in (22), except that now we take the supremum over *all* positive values of r:

$$(\mathcal{M}_P f)(x)=\sup_{r>0}\frac{1}{r^k}\left|\int_{|t|\le r}f(x-P(t))\,dt\right|.$$

The result for $\mathcal{M}_P$ is then

PROPOSITION 1. *The estimate (23) holds for $\mathcal{M}_P$ in place of $\mathcal{M}$.*

The proof of the proposition already contains the main features of the proof of Theorem 1. This is because the crucial condition imposed on γ (that it be of finite type at $t=0$) bears on finitely many derivatives of γ at the origin, and this property is captured by a Taylor polynomial of γ at $t=0$ of sufficiently high degree.

However, in order to utilize the scaling arguments of the kind that entered in the case of the parabola, we are required to simplify the situation further. In effect, we need to "free" the various monomials that appear in the polynomials $P_1,\dots,P_n$, so that these monomials stand by themselves. One can do this by passing to a higher-dimensional situation. When properly formulated, this higher-dimensional variant is paradoxically "universal"—it allows us to recover the particular polynomial manifold we started with. We accomplish this as follows.

Let d denote the maximum degree of the $P_j(t)$. We consider the collection of all monomials t^α with $1\le|\alpha|\le d$. Here $t^\alpha=t_1^{\alpha_1}\cdots t_k^{\alpha_k}$ has degree $|\alpha|=\alpha_1+\cdots+\alpha_n$. We let N denote the number of these monomials, and we shall work in the space $\mathbf{R}^N$, whose coordinates are labeled by the multi-indices α with $1\le|\alpha|\le d$, i.e., $\mathbf{R}^N=\{(x_\alpha)\}_{1\le|\alpha|\le d}$.

For functions on $\mathbf{R}^N$ we define the analogous maximal operator $\mathcal{M}_{\mathfrak{p}}$ corresponding to the polynomial map

$$\mathfrak{p}:\mathbf{R}^k\to\mathbf{R}^N,\qquad \mathfrak{p}(t)=(t^\alpha)_{1\le|\alpha|\le d}.$$

We have

$$\begin{aligned}\mathcal{M}_{\mathfrak{p}}f(x)&=\sup_{r>0}\frac{1}{r^k}\left|\int_{|t|\le r}f(x-\mathfrak{p}(t))\,dt\right|\\&=\sup_{r>0}\frac{1}{r^k}\left|\int_{|t|\le r}f((x_\alpha-t^\alpha)_{1\le|\alpha|\le d})\,dt\right|.\end{aligned} \tag{24}$$

PROPOSITION 2. *For each p with $1<p\le\infty$, there is a constant A_p so that the estimate*

$$\|\mathcal{M}_{\mathfrak{p}}f\|_{L^p(\mathbf{R}^N)}\le A_p\|f\|_{L^p(\mathbf{R}^N)} \tag{25}$$

holds for all continuous functions f having compact support.

We shall see (in §2.4 below) that, on the basis of general principles, Proposition 2 actually implies Proposition 1, and furthermore that the bound on the operator $\mathcal{M}_P$ can be taken to depend only on the degree of the polynomial P, and not to depend otherwise on its coefficients.

2.2 L^2 boundedness of $\mathcal{M}_\mathbf{p}$. The proof of (25) when $p = 2$ is, except for small changes of notation, identical with that given in §1.2 for the special case of the parabola.

Indeed, the one-parameter family of dilations relevant here is given by‡

$$x \mapsto \delta \circ x = (\delta^{|\alpha|} x_\alpha)$$

for $x = (x_\alpha) \in \mathbf{R}^N$. We alse have the corresponding norm

$$\rho(x) = \max_{1 \le |\alpha| \le d} \{|x_\alpha|^{1/|\alpha|}\}$$

and quasi-distance $\rho(x, y) = \rho(x - y)$. Note that $\rho(\delta \circ x) = \delta\rho(x)$. If we define our family of balls by

$$B(x, \delta) = \{y : \rho(x, y) < \delta\},$$

then the axioms of Chapter 1, §1 are clearly satisfied. We denote by $\tilde{M}$ the maximal operator given in terms of these balls, as in (12); of course, $\tilde{M}$ is bounded on L^2 (as in (13)).

Next, fix a smooth nonnegative function η on $\mathbf{R}^k$ with $\eta(t) = 1$ for $|t| \le 1$, and $\eta(t) = 0$ for $|t| \ge 2$. Define

$$\begin{aligned} A_j f(x) &= 2^{kj} \int_{\mathbf{R}^k} f[x - \mathbf{p}(t)]\, \eta(2^j t)\, dt \\ &= \int_{\mathbf{R}^k} f[x - 2^{-j} \circ \mathbf{p}(t)]\, \eta(t)\, dt \\ &= \int_{\mathbf{R}^k} f[(x_\alpha - 2^{-j|\alpha|} t^\alpha)_\alpha]\, \eta(t)\, dt. \end{aligned}$$

Again, taking f to be nonnegative, we have that (11) holds, up to a fixed multiplicative constant.

Now choose a fixed $\psi \in C_0^\infty(\mathbf{R}^N)$ so that

$$\int_{\mathbf{R}^N} \psi\, dx = \int_{\mathbf{R}^k} \eta\, dt,$$

and set $\psi_j(x) = 2^{\Delta j}\psi(2^j \circ x)$, where Δ denotes the "homogeneous dimension" of $\mathbf{R}^N$,

$$\Delta = \sum_{1 \le |\alpha| \le d} |\alpha|.$$

Let $B_j f = f * \psi_j$; one has again the analogue of (14). The appropriate square function is defined according to (15).

‡ We remark that the argument given here actually works for any $\mathcal{M}_P$, provided each of the polynomials P_j is *homogeneous*. The "universality" of $\mathbf{p}$ (and the treatment of nonhomogeneous polynomials) will appear in §2.4.1 below.

We now take $d\mu$ to be the measure on $\mathbf{R}^N$ defined by

$$\int_{\mathbf{R}^N} f\, d\mu = \int_{\mathbf{R}^k} f[\mathbf{p}(t)]\, \eta(t)\, dt$$

and similarly define $d\mu_j$ by

$$\int_{\mathbf{R}^N} f\, d\mu_j = \int_{\mathbf{R}^k} f[2^{-j} \circ \mathbf{p}(t)]\, \eta(t)\, dt = \int_{\mathbf{R}^k} f[\mathbf{p}(2^{-j}t)]\, \eta(t)\, dt.$$

Since $A_j f = f * d\mu_j$ and $\widehat{d\mu_j}(\xi) = \widehat{d\mu}(2^{-j} \circ \xi)$, the identity (17) continues to hold. Similarly, (18) is still valid. Again, one also has the estimate

$$|\widehat{d\mu}(\xi) - \widehat{\psi}(\xi)| \le A\rho(\xi), \qquad \text{if } \rho(\xi) \le 1.$$

Next, one verifies that $\mathbf{p}(t)$ is of finite type at each point, and is indeed of type at most d. To see this, according to the definition in Chapter 8, §3.2, we need to verify that for each $t = t_0$, and each vector $\eta \in \mathbf{R}^N$, not all of the derivatives

$$\left(\frac{\partial}{\partial t}\right)^\beta [\mathbf{p}(t) \cdot \eta], \qquad 1 \le |\beta| \le d$$

can vanish at $t = t_0$. Were this to happen, $\mathbf{p}(t) \cdot \eta$ would be constant in t, since $\mathbf{p}(t) \cdot \eta$ is a polynomial of degree at most d in t. The independence of the monomials $\{t^\alpha\}$ defining $\mathbf{p}(t)$ gives a contradiction and proves the assertion.

Theorem 2 of Chapter 8 then allows us to conclude that, in analogy with (20), we have

$$|\widehat{d\mu}(\xi) - \widehat{\psi}(\xi)| \le A\rho(\xi)^{-1/d}, \qquad \text{if } \rho(\xi) \ge 1, \tag{26}$$

and the proof of the L^2 estimate for $M_\mathbf{p}$ is finished, as before.

2.2.1 A rougher version. When $p \ne 2$, the key idea for proving L^p estimates for $M_\mathbf{p}$ is that the above argument has some slack in it; this is due to the exponent $1/d$ appearing in (26). Thus, in fact, a stronger version of the L^2 result (involving a "more singular" maximal operator) can be proved. This is then combined, via an interpolation argument, with a "less singular" maximal operator—one that falls under the scope of the standard theory presented in chapters 1 and 2—to give our desired result. The singular variant we need comes about as follows.

For each complex s, we denote by ν^s the distribution on $\mathbf{R}^N$ with Fourier transform

$$\widehat{\nu^s}(\xi) = \widehat{d\mu}(\xi) \cdot (1 + |\xi|^2)^{s/2}. \tag{27}$$

Similarly, we let

$$(\widehat{\nu_j^s})(\xi) = \widehat{\nu^s}(2^{-j} \circ \xi).$$

We then define $A_j^s(f) = f * \nu_j^s$, so that

$$\widehat{A_j^s f}(\xi) = \widehat{f}(\xi) \cdot \widehat{\nu_j^s}(\xi) = \widehat{f}(\xi) \cdot \widehat{\nu^s}(2^{-j} \circ \xi).$$

Note that the following simple variants of (19), (20), and (26) hold:

$$\begin{cases} |\widehat{\nu^s}(\xi) - \widehat{\psi}(\xi)| \leq A_s \rho(\xi), & \text{if } \rho(\xi) \leq 1 \\ |\widehat{\nu^s}(\xi) - \widehat{\psi}(\xi)| \leq A_s \rho(\xi)^{-1/d + \mathrm{Re}(s)}, & \text{if } \rho(\xi) \geq 1. \end{cases} \tag{28}$$

As is easily seen, we have $A_s \leq A(1 + |\mathrm{Im}(s)|)$, as long as $\mathrm{Re}(s)$ remains in a compact subinterval of $\mathrm{Re}(s) < d^{-1}$. The other bounds A_s that appear below also grow at most polynomially in $(1 + |\mathrm{Im}(s)|)$.

Now define

$$\mathcal{N}^s f(x) = \sup_j |A_j^s f(x)|. \tag{29}$$

We can now invoke the square function

$$\Big(\sum_j |A_j^s f - B_j f|^2\Big)^{1/2},$$

which is bounded on L^2 as long as the exponent of ρ in the first of the inequalities (28) is strictly positive, while the exponent of ρ in the second inequality is strictly negative. The argument beginning with (21) then shows that

$$\|\mathcal{N}^s f\|_{L^2(\mathbf{R}^N)} \leq A_s \|f\|_{L^2(\mathbf{R}^N)}, \qquad \text{if } \mathrm{Re}(s) < 1/d. \tag{30}$$

Observe that $\mathcal{N}^0(f) = \sup_j A_j(f) \geq \mathcal{M}(f)$, if $f \geq 0$.

2.3 L^p boundedness of $\mathcal{M}_p$. Our next goal will be to prove that

$$\|\mathcal{N}^s(f)\|_{L^p(\mathbf{R}^N)} \leq A_{p,s} \|f\|_{L^p(\mathbf{R}^N)}, \text{ if } 1 < p \leq \infty \text{ and } \mathrm{Re}(s) < 0, \tag{31}$$

where $\mathcal{N}^s(f)$ is given by (29), and $A_j^s(f) = f * \nu_j^s$.

The argument is based on the following observation. Recall that ν^s is the distribution given by

$$\widehat{\nu^s}(\xi) = \widehat{d\mu}(\xi) \cdot (1 + |\xi|^2)^{s/2}.$$

LEMMA. *When* $\mathrm{Re}(s) < 0$, *then* ν^s *is a locally integrable function on* $\mathbf{R}^N$ *that satisfies*

$$\int_{\mathbf{R}^N} |\nu^s(x-y) - \nu^s(x)|\, dx \leq A_s |y|^{\varepsilon}, \quad \text{where } \varepsilon = \min\{-\mathrm{Re}(s), 1\},$$

$$\int_{|x| \geq R} |\nu^s(x)|\, dx \leq A_s R^{-M}, \quad \text{for } R \geq 1, \text{ and all } M \geq 0. \tag{32}$$

Indeed, if G_s is given by $\widehat{G_s}(\xi) = (1 + |\xi|^2)^{s/2}$, then since this function of ξ is a symbol in the standard class $S^{\mathrm{Re}(s)}$, it follows by Proposition 4 in Chapter 6, §4 that G_s is a locally integrable function on $\mathbf{R}^N$ that satisfies the inequalities

$$|G_s(x)| \leq A_s |x|^{-N - \mathrm{Re}(s)}, \qquad |\nabla_x G_s(x)| \leq A_s |x|^{-N - \mathrm{Re}(s) - 1},$$

while $|G_s(x)|$ and $|\nabla_x G_s(x)|$ are rapidly decreasing as $|x| \to \infty$. It is then immediate that the analogues of estimates (32) hold for G_s in place of ν^s. Finally, since

$$\widehat{\nu^s}(\xi) = \widehat{\nu}(\xi) \cdot \widehat{G_s}(\xi),$$

we have that $\nu^s = \nu * G_s$. Now ν is a finite measure on $\mathbf{R}^N$ with bounded support. Hence (32) also holds for ν^s, and the lemma is proved.

Next observe that, since $\widehat{\nu_j^s}(\xi) = \widehat{\nu^s}(2^{-j} \circ \xi)$, we have that

$$\nu_j^s(x) = 2^{j\Delta} \cdot \nu(2^j \circ x),$$

where $\Delta = \sum_{1 \leq |\alpha| \leq d} |\alpha|$ is as in §2.2. So we can now apply the maximal argument formulated in Chapter 2, §4.2.1. Here $N = n$ and $\Phi(x) = \nu(x)$. The only difference is that the isotropic dilations that give

$$\Phi_{2^j}(x) = 2^{-nj} \Phi(2^{-j} x)$$

are replaced by nonisotropic dilations giving

$$\nu_j^s(x) = 2^{\Delta j} \nu^s(2^j \circ x).$$

To see that the argument given there goes through essentially unchanged, one notes that equation (16) of Chapter 2 has to be replaced by the parallel fact that

$$\int_{\rho(x) \geq c\rho(y)} \sup_j |\nu_j^s(x-y) - \nu_j^s(x)|\, dx \leq A_s, \qquad \text{when } \mathrm{Re}(s) > 0. \tag{33}$$

The proof of (33) is deduced (as before) from the Dini regularity and decay of ν, which can be stated as

$$\begin{cases} \int |\nu^s(x-y) - \nu^s(x)|\, dx \leq A_s \rho(y)^{\varepsilon}, & \text{if } \rho(y) \leq 1 \\ \int_{\rho(x) > \rho(y)} |\nu^s(x)|\, dx \leq A_s \rho(y)^{-\varepsilon}, & \text{if } \rho(y) \geq 1, \end{cases}$$

and is an immediate consequence of (32). This establishes the L^p estimate (31) for $\mathcal{N}^s$.

2.3.1 Having in our possession the bounds for $\mathcal{N}^s$ on L^2 when $\mathrm{Re}(s) < 1/d$, and the bounds on L^p when $\mathrm{Re}(s) < 0$, we can apply complex interpolation to obtain the L^p bounds for $\mathcal{N}^0$. The methods used are very similar to those described in Chapter 4, §5.2 and in Chapter 9, §1.2.5; we shall therefore be brief.

We let $x \mapsto j(x)$ be an arbitrary integer-valued function on $\mathbf{R}^N$. With it we define the analytic family of operators U^s given by

$$U^s(f)(x) = e^{s^2}(A^s_{j(x)}f)(x). \tag{34}$$

We remark that the factor e^{s^2} is inserted to compensate for the (linear) growth of the bound A_s appearing in (30) and (31). In view of (29), we have

$$\begin{cases} \|U^s(f)\|_{L^2} \leq A\|f\|_{L^2} & \text{if } \mathrm{Re}(s) = \sigma_0 < 1/d \\ \|U^s(f)\|_{L^q} \leq A\|f\|_{L^q} & \text{if } \mathrm{Re}(s) = \sigma_1 < 0,\ 1 < q \leq \infty. \end{cases} \tag{35}$$

The bounds A in (35) are independent of the choice function $j(x)$. The interpolation then gives

$$\|U^0(f)\|_{L^p} \leq A\|f\|_{L^p} \tag{36}$$

where

$$\frac{1}{p} = \frac{(1-\theta)}{2} + \frac{\theta}{q}, \qquad 0 = (1-\theta)\sigma_0 + \theta\sigma_1. \tag{37}$$

Since the bound A in (36) is again independent of $j(x)$, the result is that, for those p allowed by the restriction (37),

$$\|\sup_j |A_j(f)|\|_{L^p} \leq A\|f\|_{L^p}.$$

Moreover, when $f \geq 0$, $\sup_j A_j(f) \geq \mathcal{M}_{\mathrm{p}}(f)$, and we get that

$$\|\mathcal{M}_{\mathrm{p}} f\|_{L^p} \leq A\|f\|_{L^p}.$$

The proof of Proposition 2 (i.e., estimate (25)) will be complete as soon as we see that all p, $1 < p \leq \infty$, fall under the scope of (37). Indeed, with p given, fix θ and q, with $0 < \theta < 1$ and $1 < q < \infty$, so that

$$\frac{1}{p} = \frac{(1-\theta)}{2} + \frac{\theta}{q};$$

for example, when $1 < p \leq 2$, take

$$\theta = \frac{1}{p}, \qquad \text{and} \qquad \frac{1}{q} = \frac{3}{2} - \frac{p}{2}.$$

With θ and p chosen, take σ_0 to be $d/2$ (for instance). Then σ_1 can be determined by the relation

$$0 = (1-\theta)\sigma_0 + \theta\sigma_1,$$

and we have $\sigma_1 < 0$. This establishes Proposition 2.

2.4 Method of descent. We shall now describe the method that allows us to pass from results about (translation-invariant) operators in one Euclidean space ($\mathbf{R}^N$) to corresponding operators on another Euclidean space ($\mathbf{R}^n$).

We begin by considering a fixed linear mapping L from $\mathbf{R}^N$ to $\mathbf{R}^n$. With L given, suppose T is a translation-invariant operator acting on functions on $\mathbf{R}^n$. Then we can "induce", in a natural way, an operator T^L, acting on functions on $\mathbf{R}^n$. In order not to involve irrelevant technicalities, we limit ourselves to the situation where T is given by convolution with a finite measure $d\mu$,

$$T(f) = f * d\mu. \tag{38}$$

When dealing with estimates of norms (in (39) and other identities below), it will suffice to restrict our attention to continuous functions with compact support. The operator T^L is then defined by

$$T^L f(x) = \int_{\mathbf{R}^N} f(x - L(z))\, d\mu(z), \qquad x \in \mathbf{R}^n, \tag{39}$$

this time for f given on $\mathbf{R}^n$.

Two quick remarks may be in order. First, the operator (39) is in reality a convolution operator on $\mathbf{R}^n$ with the measure $d\mu^L$, where $d\mu^L$ is the "push-forward" of $d\mu$ to $\mathbf{R}^n$; that is, with $d\mu^L$ defined by

$$\int_{\mathbf{R}^n} f(x)\, d\mu^L(x) = \int_{\mathbf{R}^N} f(Lz)\, d\mu(z).$$

Second, there is the important special case that arises when $N \geq n$, and $\mathbf{R}^n$ is the subspace of $\mathbf{R}^N$ determined by the first n coordinates. In that case, consider the corresponding direct sum decomposition $\mathbf{R}^N = \mathbf{R}^n \oplus \mathbf{R}^{N-n}$. If L denotes the projection operator from $\mathbf{R}^N$ to $\mathbf{R}^n$, i.e., $L(z) = x$, where $z = (x, x')$, $x \in \mathbf{R}^n$, $x' \in \mathbf{R}^{N-n}$, then as is directly verified from (39)

$$T^L f(x) = f * d\mu',$$

where $d\mu'$ is obtained by integrating $d\mu$ on $\mathbf{R}^{N-n}$,

$$d\mu'(x) = \int_{\mathbf{R}^{N-n}} d\mu(x, x').$$

We shall be dealing with one operator, $Tf = f * d\mu$, and its induced operator T^L, as well as a countable family $T_j f = f * d\mu_j$ of such operators and the corresponding induced operators T_j^L. To simplify the proof below, we assume that $d\mu$ and the $d\mu_j$ all have fixed compact support. The support restriction can be removed by an additional argument, but it is in any case satisfied in the applications below.

LEMMA. *Suppose $L : \mathbf{R}^N \to \mathbf{R}^n$ is a fixed linear mapping as above, T is as in (38), and $1 \le p \le \infty$.*

(a) *The norm of the operator T^L acting on $L^p(\mathbf{R}^n)$ does not exceed the norm of T acting on $L^p(\mathbf{R}^N)$.*

(b) *Let $T_j(f) = f * d\mu_j$ be as above. Assume additionally that the $d\mu_j$ are nonnegative, and set*

$$T_*f(x) = \sup_j |T_jf(x)|, \quad \text{and} \quad T_*^L f(x) = \sup_j |T_j^L f(x)|.$$

If T_* is bounded on $L^p(\mathbf{R}^N)$, then T_*^L is bounded on $L^p(\mathbf{R}^n)$, and its norm does not exceed that of T_*.

Proof. Suppose that $d\mu$ is supported in the ball $|z| \le M$. We may also assume that $p < \infty$. For $u \in \mathbf{R}^N$, we let τ_u denote the translation operator acting on functions on $\mathbf{R}^n$ by

$$(\tau_u f)(x) = f(x + L(u)).$$

Then

$$\|T^L f\|^p_{L^p(\mathbf{R}^n)} = \frac{1}{c_N R^N} \int_{|u|\le R} \|\tau_u T^L f\|^p_{L^p(\mathbf{R}^n)}\, du,$$

where c_N is the volume of the unit ball in $\mathbf{R}^N$. Now $\tau_u T^L = T^L \tau_u$ and

$$\begin{aligned}\tau_u T^L f(x) &= \int_{\mathbf{R}^N} f(x - L(z-u))\, d\mu(z) \\ &= \int_{\mathbf{R}^N} f(x - L(z-u))\, \chi_{R+M}(z-u)\, d\mu(z) \\ &= \int_{\mathbf{R}^N} F_x(u-z)\, d\mu(z) = T(F_x)(u),\end{aligned}$$

where χ_{R+M} is the characteristic function of the ball (about the origin) of radius $R+M$; this is because $|u| \le R$ and $|z| \le M$, hence $|z-u| \le R+M$. Here $F_x(z) = f(x+L(z))\,\chi_{R+M}(z)$ is a function on $\mathbf{R}^N$, for each $x \in \mathbf{R}^n$. However,

$$\int_{|u|\le R} |TF_x(u)|^p\, du \le \|T\|^p_{p,p} \int_{\mathbf{R}^N} |F_x(u)|^p\, du. \tag{40}$$

Note that

$$\int_{\mathbf{R}^n}\int_{\mathbf{R}^N} |F_x(u)|^p\, du\, dx = \|f\|^p_{L^p} c_N (R+M)^N,$$

so that integrating (40) over $x \in \mathbf{R}^n$ gives

$$\int_{\mathbf{R}^n}\int_{|u|\le R} |TF_x(u)|^p\, du\, dx \le \|T\|^p_{p,p} \cdot \|f\|^p_{L^p} \cdot c_N (R+M)^N.$$

Altogether then,

$$\|T^L f\|^p_{L^p} \le \frac{(R+M)^N}{R^N} \cdot \|T\|^p_{p,p} \cdot \|f\|^p_{L^p}. \tag{41}$$

If we let $R \to \infty$, we get conclusion (a). The same argument also shows that

$$\|T_*^L\|^p_{L^p} \le \frac{(R+M)^N}{R^N} \cdot \|T_*\|^p_{p,p} \cdot \|f\|^p_{L^p},$$

and (b) is also established.

2.4.1 L^p boundedness of $\mathcal{M}_P$. Looking back to §2.1, we can now deduce Proposition 1 from Proposition 2. In the second proposition we are concerned with the averages on $\mathbf{R}^N$

$$\frac{1}{r^k}\int_{|t|\le r} f(x - \mathfrak{p}(t))\, dt. \tag{42}$$

For each r, we define a corresponding measure $d\mu_r$ on $\mathbf{R}^N$ given by

$$\int_{\mathbf{R}^N} f(x)\, d\mu_r(x) = \frac{1}{r^k}\int_{|t|\le r} f(\mathfrak{p}(t))\, dt.$$

Then $f * d\mu_r = T_r(f)$ is the average (42).

Similarly, in the first proposition we considered the averages on $\mathbf{R}^n$ given by

$$\frac{1}{r^k}\int_{|t|\le r} f(x - P(t))\, dt, \tag{43}$$

where $P = (P_1, \dots, P_n)$ is an n-tuple of real-valued polynomials whose degrees are all at most d. We may assume that the P_j vanish at the origin; to ensure this involves merely a translation. We write

$$P_j(t) = \sum_{1\le|\alpha|\le d} a_{j\alpha} t^\alpha.$$

Using the coefficients $a_{j\alpha}$, we define a linear mapping $L : \mathbf{R}^N \to \mathbf{R}^n$, where the coordinates of $L(x) \in \mathbf{R}^n$ are given by

$$L(x)_j = x'_j = \sum_\alpha a_{j\alpha} x_\alpha;$$

here $x = (x_\alpha) \in \mathbf{R}^N$. Then clearly $d\mu_r^L$ is given by

$$\int_{\mathbf{R}^n} f(x)\, d\mu_r^L(x) = \frac{1}{r^k}\int_{|t|\le r} f(L(\mathfrak{p}(t)))\, dt = \frac{1}{r^k}\int_{|t|\le r} f(P(t))\, dt,$$

since, by the definition of L, we have $L(\mathfrak{p}(t)) = P(t)$. As a result, the averages (43) equal

$$f * d\mu_r^L = T_r^L(f).$$

We can now invoke part (b) of the lemma just proved, restricting ourselves to a countable dense set of r, for the interval $0 < r < R$, and then letting $R \to \infty$. The conclusion is that the estimate

$$\|\mathcal{M}_{\mathfrak{p}}\|_{p,p} \leq A_p$$

implies

$$\|\mathcal{M}_P\|_{p,p} \leq A_p,$$

and therefore Proposition 1 is established.

2.5 A general version. Having described in detail the proofs of the two model cases, we return to the general theorem stated at the beginning of §2. To see how all the ideas needed for it have already arisen in the proof of the model cases, we abstract these arguments in the following general proposition.

Suppose we are given a sequence $\{dm^j\}$ of measures on $\mathbf{R}^N$, which for simplicity we assume to be nonnegative, have a common compact support, and are normalized by the condition that $\int dm^j = c$, with c independent of j.

The crucial assumption we make on the dm_j is the uniform decay of their Fourier transforms; that is, the existence of a $\delta > 0$ so that

$$|\widehat{dm^j}(\xi)| \leq A|\xi|^{-\delta}, \tag{44}$$

with A independent of j.†

Recalling the nonisotropic dilations $x \mapsto \delta \circ x$ on $\mathbf{R}^N$ (described in §2.2), we can define the dilated measures $d\mu_j$ by

$$\int_{\mathbf{R}^N} f(x)\, d\mu_j(x) = \int_{\mathbf{R}^N} f(2^{-j} \circ x)\, dm^j(x),$$

which means that

$$\widehat{d\mu_j}(\xi) = \widehat{dm^j}(2^{-j} \circ \xi). \tag{45}$$

Now let $A_j(f) = f * d\mu_j$, and

$$\mathcal{N} f(x) = \sup_j |A_j f(x)|.$$

PROPOSITION. *Let $1 < p \leq \infty$. Under the above assumptions on the measures $\{dm^j\}$, there is a bound A_p so that*

$$\|\mathcal{N}(f)\|_{L^p(\mathbf{R}^N)} \leq A_p \|f\|_{L^p(\mathbf{R}^N)} \tag{46}$$

whenever f is continuous and has compact support.

† Of course, this condition is of interest only for large ξ; at any rate, it is obvious that the $\widehat{dm^j}(\xi)$ are uniformly bounded.

Before we turn to the proof, the following remarks may be helpful to the reader. The conclusions we have previously established in the model case correspond here to the situation where there is one measure, i.e., where the measure dm^j is *independent* of j. The general case of Theorem 1 is approximated by the model case in the sense that the measures dm^j that arise in the theorem converge in a suitable sense (as $j \to \infty$) to *the* measure for the model case.

2.5.1 L^2 estimates for $\mathcal{N}$. For each complex s, let M_j^s be the distribution defined by

$$M_j^s = dm^j * G_s,$$

where $\widehat{G_s}(\xi) = (1 + |\xi|^2)^{s/2}$ is as in §2.3; thus

$$\widehat{M_j^s}(\xi) = \widehat{dm^j}(\xi) \cdot (1 + |\xi|^2)^{s/2}.$$

Also, define ν_j^s by

$$\widehat{\nu_j^s}(\xi) = \widehat{M_j^s}(2^{-j} \circ \xi),$$

and write, using the same notation as above,

$$A_j^s(f) = f * \nu_j^s, \qquad \mathcal{N}^s f(x) = \sup_j |A_j^s f(x)|.$$

Next, fix a smooth function ψ with compact support in $\mathbf{R}^N$ so that it has the same normalization as the dm^j, i.e.,

$$\int \psi\, dx = c.$$

Let $\psi_j(x) = 2^{j\Delta}\psi(2^j \circ x)$, and set

$$B_j(f) = f * \psi_j.$$

Define the square function

$$S_s f(x) = \Big(\sum_j |A_j^s f(x) - B_j f(x)|^2\Big)^{1/2}.$$

The first point is that

$$\|S_s(f)\|_{L^2(\mathbf{R}^N)} \leq A_s \|f\|_{L^2(\mathbf{R}^N)}, \qquad \text{if } \mathrm{Re}(s) \leq \delta; \tag{47}$$

as in §2.2.1, the bound A_s can be estimated by $A_s \leq A(1 + |\mathrm{Im}(s)|)$, so long as $\mathrm{Re}(s)$ is restricted to a compact subinterval of $\mathrm{Re}(s) < \delta$.

Indeed, this is equivalent by Plancherel's theorem to

$$\left(\sum_j |\widehat{M_j^s}(2^{-j}\circ\xi) - \widehat{\psi}(2^{-j}\circ\xi)|^2\right)^{1/2} \leq A_s. \tag{48}$$

To prove (48), recall that

$$\widehat{M_j^s}(0) = \widehat{dm^j}(0) = \widehat{\psi}(0),$$

and that the measures dm^j have uniform norms and a common compact support. Thus

$$|\widehat{M_j^s}(\xi) - \widehat{\psi}(\xi)| \leq A_s|\xi| \leq A_s\rho(\xi), \qquad \text{if } \rho(\xi) \leq 1. \tag{49}$$

Also, because $\widehat{M_j^s}(\xi) = \widehat{dm^j}(\xi)\cdot(1+|\xi|^2)^{s/2}$, our assumption (44) implies that

$$|\widehat{M_j^s}(\xi) - \widehat{\psi}(\xi)| \leq A|\xi|^{-\delta+\mathrm{Re}(s)} \leq A\rho(\xi)^{-\delta+\mathrm{Re}(s)}, \tag{50}$$

if $\rho(\xi) \geq 1$ and $\mathrm{Re}(s) < \delta$.

In estimating the left side of (48), we repeat the argument of §1.2.2, using (49) for those terms where

$$\rho(2^{-j}\circ\xi) = 2^{-j}\rho(\xi) \leq 1,$$

and using (50) where

$$\rho(2^{-j}\circ\xi) > 1.$$

This establishes the desired inequality, and with it the estimate (47) for the square function S_s. Since

$$\mathcal{N}^s(f) \leq \sup_j |B_j(f)| + S_s(f),$$

the result is

$$\|\mathcal{N}^s(f)\|_{L^2} \leq A_s\|f\|_{L^2}, \qquad \text{for } \mathrm{Re}(s) < \delta. \tag{51}$$

2.5.2 The L^p estimates depend on the following variant of the argument used in §2.3. Suppose M_j is a sequence of integrable functions on $\mathbf{R}^N$ that satisfy the following conditions for some $\varepsilon > 0$, uniformly in j:

$$\begin{cases} \displaystyle\int_{\mathbf{R}^N} |M_j(x-y) - M_j(x)|\,dx & \leq A[\rho(y)]^{\varepsilon} \\ \displaystyle\int_{\rho(x)\geq c\rho(y)} |M_j(x)|\,dx & \leq A[\rho(y)]^{-\varepsilon}. \end{cases} \tag{52}$$

LEMMA. *Under these assumptions, if*

$$\nu_j(x) = 2^{j\Delta}M_j(2^j\circ x),$$

then

$$f \mapsto \sup_j |(f*\nu_j)(x)|$$

is of weak-type $(1,1)$ *and is bounded from* $L^p(\mathbf{R}^N)$ *to itself, when* $p>1$.

This is a reprise of the situation covered in Chapter 2, §4.2. Following the argument there, it suffices to verify that

$$\int_{\rho(x)\geq c\rho(y)} \sup_j |\nu_j(x-y) - \nu_j(x)|\,dx \leq A. \tag{53}$$

To check (53), we observe that

$$\sup_j |\nu_j(x-y) - \nu_j(x)| \leq \sum_j |\nu_j(x-y) - \nu_j(x)|.$$

In the sum above we use the first estimate in (52) when $\rho(y) \leq 2^{-j}$:

$$\begin{aligned}\int |\nu_j(x-y) - \nu_j(x)| &\leq \int |M_j(x - 2^j\circ y) - M_j(x)|\,dx \\ &\leq A2^{\varepsilon j}[\rho(y)]^{\varepsilon}.\end{aligned}$$

Similarly, when $\rho(y) > 2^{-j}$,

$$\int_{\rho(x)\geq c\rho(y)} |\nu_j(x-y)-\nu_j(x)|\,dx \leq \int_{\rho(x)\geq c'\rho(y)} |\nu_j(x)|\,dx \leq A2^{-\varepsilon j}[\rho(y)]^{-\varepsilon},$$

by the second estimate in (52).

Next we apply this lemma to the situation where

$$M_j = M_j^s = (dm^j)*G_s, \qquad \text{for } \mathrm{Re}(s) < 0.$$

Note that here $\nu_j = \nu_j^s$ (as defined in §2.5.1), so that

$$\sup_j |f*\nu_j| = \sup_j |f*\nu_j^s| = \mathcal{N}^s(f).$$

The fact that the M_j satisfy the conditions (52) is proved exactly as (32) was (in §2.3), since

$$M_j = M_j^s = dm^j * G_s,$$

and since we supposed that the dm^j have uniform norm and a common bounded support. Our conclusion is then (like (31))

$$\|\mathcal{N}^s(f)\|_{L^p(\mathbf{R}^N)} \leq A_{p,s}\|f\|_{L^p(\mathbf{R}^N)}, \qquad \text{for } 1 < p \leq \infty \text{ and } \mathrm{Re}(s) < 0.$$

By use of the interpolation argument in §2.3.1, we combine this with the L^2 estimate (51), giving us (46). The proposition is therefore proved.

2.6 Completion of the proof for submanifolds. We can now finish the proof of the theorem stated at the beginning of §2.

Suppose that $t \mapsto \gamma(t)$ is of type d at $t=0$; we may assume, after a possible translation, that $\gamma(0)=0$. By a Taylor expansion,

$$\gamma(t) = P(t) + R(t),$$

where P is a polynomial (without constant term) of degree d, and R vanishes to order $d+1$ at the origin; in particular

$$\left(\frac{\partial}{\partial t}\right)^{\alpha} R(t) = O(|t|^{d-|\alpha|+1}) \qquad \text{as } t \to 0,$$

for every $|\alpha| \le d$. The fact that γ is of type d at the origin implies that P is also of type d, because this property is determined by the derivatives of order $\le d$ at the origin.

Write $P = (P_1, \dots, P_n)$,

$$P_j(t) = \sum_{1 \le |\alpha| \le d} a_{j\alpha} t^{\alpha},$$

and define the linear mapping $L : \mathbf{R}^N \to \mathbf{R}^n$ by

$$L(x)_j = x'_j = \sum_{\alpha} a_{j\alpha} x_{\alpha}$$

where $x = (x_\alpha) \in \mathbf{R}^N$, as in §2.4.1. Now the assumption that P is of type d implies, among other things, that the polynomials

$$P_1(t), \dots, P_n(t)$$

are linearly independent (otherwise S would lie in a hyperplane). As a result, the range of L is all of $\mathbf{R}^n$, and there is a linear mapping $M : \mathbf{R}^n \to \mathbf{R}^N$ so that $L \cdot M$ is the identity mapping on $\mathbf{R}^n$.

Next, recall the definition of $\mathfrak{p}$, the "free" polynomial of degree d mapping $\mathbf{R}^n$ to $\mathbf{R}^N$:

$$\mathfrak{p}(t) = (t^{\alpha})_{1 \le |\alpha| \le d},$$

as in §2.1. Then of course $P(t) = L(\mathfrak{p}(t))$. Set $\mathfrak{r}(t) = M(R(t))$.

If we define $\Gamma(t) = \mathfrak{p}(t) + \mathfrak{r}(t)$, we see that

$$L(\Gamma) = L(\mathfrak{p}) + L(\mathfrak{r}) = P + R = \gamma.$$

Therefore, if we make use of the descent procedure described in §2.4, it will suffice to control the corresponding maximal function (acting on functions on $\mathbf{R}^N$) given by

$$\sup_{0<r<1} \frac{1}{r^k} \left| \int_{|t| \le r} f(x - \Gamma(t))\, dt \right|. \tag{54}$$

To deal with (54), we first define the measures $d\mu_j$ on $\mathbf{R}^n$ by

$$\begin{aligned} \int_{\mathbf{R}^N} f(x)\, d\mu_j(x) &= 2^{jk} \int_{\mathbf{R}^k} f[\Gamma(t)]\, \eta(2^j t)\, dt \\ &= \int_{\mathbf{R}^k} f[\Gamma(2^{-j}t)]\, \eta(t)\, dt. \end{aligned} \tag{55}$$

Here η is a fixed smooth nonnegative function, supported in the ball $|t| \le 2$, with $\eta(t) = 1$ for $|t| \le 1$. To be precise, we take the above as our definition when $j \ge j_0$, for $j < j_0$, we set $d\mu_j \equiv 0$; here j_0 is a fixed positive integer that will be determined later.

Write $A_j(f) = f * d\mu_j$. Then it is clear that the desired estimates for the maximal function (54) follow from those for

$$\mathcal{N}(f) = \sup_j |A_j(f)|.$$

Indeed,

$$\sup_{0<r\le 2^{-j_0}} \frac{1}{r^k} \left| \int_{|t| \le r} f(x - \Gamma(t))\, dt \right| \le c \sup_j |A_j f(x)|,$$

while

$$\sup_{2^{-j_0} \le r \le 1} \frac{1}{r^k} \left| \int_{|t| \le r} f(x - \Gamma(t))\, dt \right| \le c \int_{|t| \le 1} |f(x - \Gamma(t))|\, dt.$$

If $1 \le p \le \infty$, the L^p norm of the last integral is bounded by a constant multiple of the L^p norm of f, as Minkowski's inequality shows.

To treat $\mathcal{N}$, we invoke the proposition formulated at the beginning of §2.5. We define the measures dm^j by

$$\int_{\mathbf{R}^N} f(x)\, dm^j(x) = \int_{\mathbf{R}^k} f(2^j \circ \Gamma(2^{-j}t)) \eta(t)\, dt. \tag{56}$$

Note that dm^j is nonnegative, and

$$\int dm^j = \int \eta(t)\, dt = c.$$

Moreover, the support of dm^j is contained in the image of the ball $|t| \le 2$ under the mapping

$$t \mapsto 2^j \circ \Gamma(2^{-j}t).$$

However, $\Gamma(2^{-j}t) = \mathfrak{p}(2^{-j}t) + \mathfrak{r}(2^{-j}t) = 2^{-j} \circ \mathfrak{p}(t) + O(2^{-j(d+1)})$; thus the values of this mapping remain bounded, and the dm^j have a common compact support.

Next, observe that the $d\mu_j$ are the dilates of the dm^j, i.e.,

$$\int f(x)\,d\mu_j(x) = \int f(2^{-j} \circ x)\,dm^j(x),$$

because of (55) and (56). What remains to be done, then, is to check the decay property (44) of the Fourier transforms of the dm^j.

Now

$$\int_{\mathbf{R}^N} e^{-2\pi i x\cdot\xi}\,dm^j(x) = \int_{\mathbf{R}^k} \exp\{-2\pi i[2^j \circ \Gamma(2^{-j}t)]\cdot\xi\}\,\eta(t)\,dt.$$

However, $2^j \circ \Gamma(2^{-j}t) = \mathfrak{p}(t) + 2^j \circ \mathfrak{r}(2^{-j}t)$ and, moreover,

$$\left|\left(\frac{\partial}{\partial t}\right)^{\alpha} 2^j \circ \mathfrak{r}(2^{-j}t)\right| \le c2^{-j}, \qquad \text{whenever } |\alpha| \le d,$$

since $\mathfrak{r}(t) = M(R(t))$, while $|(d/dt)^\alpha r(t)| \le A|t|^{-d-|\alpha|+1}$.

Write $\xi = \lambda\zeta$, where ζ is a unit vector. We already saw that for each t and ζ, there is an α, $1 \le |\alpha| \le d$, so that

$$\left(\frac{\partial}{\partial t}\right)^{\alpha}[\mathfrak{p}(t)\cdot\zeta] \ne 0.$$

Thus if j_0 is sufficiently large, and $j \ge j_0$

$$\left(\frac{\partial}{\partial t}\right)^{\alpha}[2^j \circ \Gamma(2^{-j}t)\cdot\zeta] \ne 0.$$

We can therefore invoke the decay estimates in §2.2 and §3.2 of Chapter 8 (and, in particular, the uniformities of Proposition 5 there) to conclude that

$$|\widehat{dm^j}(\xi)| \le A|\xi|^{-1/d}, \qquad \text{if } j \ge j_0.$$

This is assumption (44); therefore the hypotheses of the proposition are verified, and the proof of the theorem is complete.

2.7 Further remarks. 1. Suppose first that f is locally integrable on $\mathbf{R}^n$. Then, for almost every x, the integral

$$\frac{1}{c_k r^k}\int_{|t|\le r} f(x-\gamma(t))\,dt \tag{57}$$

converges absolutely for all r, $0 < r \le 1$, and the resulting function of r is continuous for these r.‡ Indeed, if we set

$$F(x,t) = f(x-\gamma(t)),$$

then by Fubini's theorem, F is locally integrable on $\mathbf{R}^n \times \mathbf{R}^k$. Using Fubini's theorem again, we see that $F(x,t)$ is integrable in t for almost every x, from which our assertion follows. The result is that, in forming the maximal function (22), it suffices to take the supremum over a countable dense set of r, $0 < r \le 1$. Thus $\mathcal{M}f$ is well defined, whenever f is locally integrable.

‡ For the purposes of the discussion below, the normalization used in (57) (c_k is the volume of the unit ball in $\mathbf{R}^k$) will prove to be convenient.

2. Next, if we make the finite-type assumption of Theorem 1, the L^p inequality (23), initially proved for continuous functions of compact support, extends to all $f \in L^p(\mathbf{R}^n)$, as a simple limiting argument shows.

3. As a consequence, we have that the averages (57) converge to $f(x)$ as $r \to 0$, for almost every x, whenever f is locally in some L^p, $p > 1$.

4. Assertions of this type continue to hold for certain classes "close" to L^1. Indeed, the almost everywhere conclusions are still valid when f belongs locally to the space $L \log L$ (see §4.18). What happens for L^1 is still unsettled.

5. All conclusions of this kind may fail if we drop the finite-type assumption. In fact, there is a smooth curve $t \mapsto \gamma(t)$ in $\mathbf{R}^2$, which is infinitely flat near the origin, and so that the maximal operator (22) is unbounded for every $p < \infty$. Moreover, the limit of (57), as $r \to 0$, may fail to exist almost everywhere, even when f is the characteristic function of a suitable set. For this, see §4.3 below.

6. It should be noted that our assumptions stipulated that γ be a smooth mapping, but did not require that the submanifold $S = \{\gamma(t)\}$ itself be smooth. In particular, varieties S with various singularities can be realized in this way. Simple examples are curves with cusps, given by

$$\gamma(t) = (t^k, t^\ell) \qquad \text{where } k \text{ and } \ell \text{ are positive integers,}$$

and cones in $\mathbf{R}^n$, $n \ge 3$, given by

$$\begin{aligned} x_j &= 2t_j t_{n-1}, & 1 \le j \le n-2,\\ x_{n-1} &= t_1^2 + \cdots + t_{n-2}^2 - t_{n-1}^2, & \\ x_n &= t_1^2 + \cdots + t_{n-2}^2 + t_{n-1}^2. & \end{aligned}$$

3. Averages on variable hypersurfaces

We take up again the spherical averages dealt with in a preliminary way in §1.1. Our purpose will be to pass from that setting to one that allows a great deal more flexibility, and in particular is not bound by the rather rigid requirements of translation invariance.

What we have in mind can be described roughly as follows. For each $x \in \mathbf{R}^n$, we will be given a family of hypersurfaces $S_{x,t}$, where the parameter t ranges over (say) $0 < t \le 1$, so that as $t \to 0$, the hypersurfaces $S_{x,t}$ shrink to x.

We then define the averages

$$(A_t f)(x) = \int_{S_{x,t}} f(y)\,d\sigma_{x,t}(y), \tag{58}$$

where $d\sigma_{x,t}$ is a suitably normalized measure on $S_{x,t}$. We will be concerned with

$$\lim_{t\to 0}(A_t f)(x)$$

and the corresponding maximal function

$$(\mathcal{A}f)(x) = \sup_{0<t\leq 1} |A_t f(x)|.$$

In dealing with these averages, we shall have to face two related (but not identical) issues: first, the behavior of $A_t f(x)$ as a function of t, for t in a range bounded away from zero; and second, the limiting situation as $t \to 0$. For the behavior of A_t when t is near zero, what will turn out to be crucial is the "derived hypersurface" given in effect by

$$\lim_{t\to 0} \frac{S_{x,t} - x}{t},$$

and the condition that this hypersurface has nonzero Gaussian curvature, for each x. However, when t is not small, these considerations are no longer relevant. The appropriate notion that takes both ranges of t into account is that of "rotational curvature", which we now take up.

3.1 Rotational curvature. To explain this notion, we momentarily fix the parameter t, and consider a mapping $x \mapsto S_x$ that assigns to each x (in some region of $\mathbf{R}^n$), a hypersurface S_x. This mapping will be determined by a smooth real-valued defining function $\Phi(x, y)$ given in some region of $\mathbf{R}^n \times \mathbf{R}^n$ by the condition

$$S_x = \{y : \Phi(x, y) = 0\} \tag{59}$$

We now define the *rotational curvature* of Φ, which we abbreviate as $\operatorname{rot\,curv}(\Phi)$, by

$$\operatorname{rot\,curv}(\Phi)(x,y) = \det \begin{pmatrix} \Phi & \frac{\partial\Phi}{\partial y_1} & \cdots & \frac{\partial\Phi}{\partial y_n} \\ \frac{\partial\Phi}{\partial x_1} & \frac{\partial^2\Phi}{\partial x_1\partial y_1} & \cdots & \frac{\partial^2\Phi}{\partial x_1\partial y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial\Phi}{\partial x_n} & \frac{\partial^2\Phi}{\partial x_n\partial y_1} & \cdots & \frac{\partial^2\Phi}{\partial x_n\partial y_n} \end{pmatrix} \tag{60}$$

Our basic condition is that

$$\operatorname{rot\,curv}(\Phi) \neq 0, \qquad \text{where } \Phi = 0. \tag{61}$$

Let us remark that when this is satisfied, we have automatically that $\nabla_y \Phi \neq 0$ where $\Phi = 0$. Thus the surfaces S_x, implicitly defined by (59), are (locally) smooth submanifolds and vary smoothly with x.

We illustrate the notion of rotational curvature by mentioning three examples:

(i) The situation when $x \mapsto S_x$ is translation-invariant. This means that $S_x = x + S_0$, where S_0 is the hypersurface associated to the origin. In terms of our defining function, this arises if

$$\Phi(x, y) = \Psi(y - x), \qquad \text{and} \quad S_0 = \{y : \Psi(y) = 0\}.$$

When $y \in S_0$, $\nabla_y \Psi(y)$ gives a normal vector to the hypersurface S_0 at y. The condition (61) is then the same as the nondegeneracy of the $(n-1) \times (n-1)$ symmetric matrix obtained by restricting

$$\left(\frac{\partial^2 \Psi}{\partial y_i \partial y_j}\right)_{1\leq i,j\leq n}$$

to the hyperplane perpendicular to this normal vector (i.e., the tangent plane to S_0 at y). Hence, nonvanishing rotational curvature is equivalent to the nonvanishing of the Gaussian curvature of S_0. In view of this, it will be convenient to define for later purposes

$$\operatorname{Gauss\,curv}(\Psi)(y - x) = \operatorname{rot\,curv}(\Phi)(x, y), \tag{62}$$

where $\Phi(x, y) = \Psi(y - x)$.

(ii) Consider $\Phi(x, y) = x \cdot y + t$, where t is a nonzero constant. Note that

$$\operatorname{rot\,curv}(\Phi)(x, y) = -x \cdot y.$$

Here each S_x is an affine hyperplane, and of course has zero Gaussian curvature. The condition (61) is satisfied, essentially because of the turning of these hyperplanes as x varies.

(iii) Another example of "rotating" hyperplanes is as follows. For $x \in \mathbf{R}^n$, write $x = (x', x_n)$, where $x' \in \mathbf{R}^{n-1}$, $x_n \in \mathbf{R}^1$, and let $B(x', y')$ be a nondegenerate bilinear form on $\mathbf{R}^{n-1} \times \mathbf{R}^{n-1}$. Set

$$\Phi(x, y) = (x_n - y_n) + B(x', y').$$

The special case when B is antisymmetric arises in the Heisenberg group (see (6) in Chapter 12 below); then $B(x', x') = 0$ and S_x is a hyperplane passing through x.

We now describe the invariance properties of rotational curvature.

(a) Suppose the family $x \mapsto S_x$ is given by another defining function Φ', where $\Phi' = 0$ when $\Phi = 0$. Then, if we assume that $\nabla_y \Phi' \neq 0$ when $\Phi' = 0$, we can assert the existence of a smooth function ψ, with $\psi \neq 0$ when $\Phi = 0$, so that

$$\Phi'(x, y) = \psi(x, y) \cdot \Phi(x, y).$$

Note that as a result, $\operatorname{rot\,curv}(\Phi') = \psi^{n+1} \operatorname{rot\,curv}(\Phi)$ when $\Phi = 0$. Thus the condition (61) on Φ is equivalent to the same condition on Φ'.

(b) The condition (61) is unchanged if the underlying spaces undergo (local) diffeomorphisms. Indeed, let $x \mapsto \psi_1(x)$ and $y \mapsto \psi_2(y)$ be diffeomorphisms of the x and y spaces. Then if

$$\Phi'(x,y) = \Phi(\psi_1(x), \psi_2(y)),$$

we have that

$$\operatorname{rot\,curv}(\Phi') = J_1(x)\cdot J_2(y)\cdot \operatorname{rot\,curv}(\Phi), \qquad \text{wherever } \Phi' = 0;$$

here J_1 and J_2 are the determinants of the Jacobian matrices of the transformations ψ_1 and ψ_2, respectively.

(c) There is a simple relation between the condition (61) and the nonvanishing of a Hessian in $\mathbf{R}^{n+1}\times\mathbf{R}^{n+1}$ that is worth noting. Define $\tilde{\Phi}$ on $\mathbf{R}^{n+1}\times\mathbf{R}^{n+1}$ by

$$\tilde{\Phi} = x_0\cdot y_0\cdot \Phi(x,y),$$

whenever $(x_0, x)\in\mathbf{R}^{n+1}$, $(y_0,y)\in\mathbf{R}^{n+1}$. Then, as is easily verified,

$$\det\left(\frac{\partial^2\tilde{\Phi}}{\partial x_i\partial y_j}\right)_{0\le i,j\le n} = (x_0y_0)^n\cdot \operatorname{rot\,curv}(\Phi). \tag{63}$$

(d) The rotational curvature condition is essentially equivalent to the property that the averaging operators (58) are actually Fourier integral operators of order $(1-n)/2$. For more about this, see §4.8 below; a special case is indicated in Chapter 8, §5.7.

3.1.1 Scaled family. To deal with the behavior for variable t, we need to consider the scalings about x; these form a one-parameter family of transformations defined by

$$x\mapsto x, \qquad y\mapsto x + t(y-x);$$

the inverse scalings are given by

$$x\mapsto x, \qquad y\mapsto x + \frac{y-x}{t}.$$

Whenever Φ^t is a parametrized family, depending smoothly on t in the closed interval $0\le t\le 1$, the corresponding *scaled family*, denoted by Φ_t is defined by

$$\Phi_t(x,y) = \Phi^t(x, x+(y-x)/t),$$

when $t>0$.†

† In what follows, we shall use the convention of attaching the superscript t to a parametrized family and the subscript t to the corresponding scaled family.

Note that if S_x^t are the hypersurfaces determined by $\Phi^t = 0$, where Φ^t is a parametrized family, then the hypersurfaces $S_{x,t}$ arising from the scaled family Φ_t are given by

$$S_{x,t} = x + t(S_x^t - x). \tag{64}$$

Moreover, a simple calculation shows that

$$t^{2n}\cdot \operatorname{rot\,curv}(\Phi_t)\big|_{\substack{x=x^0\\ y=tu+x^0}} \to \operatorname{Gauss\,curv}(\Psi)(u) \qquad \text{as } t\to 0, \tag{65}$$

where $\Psi(u) = \Phi^0(x^0, u)$.

In view of this, the natural condition to impose on the scaled family Φ_t is that

$$t^{2n}\cdot\operatorname{rot\,curv}(\Phi_t)\ge c>0 \quad \text{where } \Phi_t = 0, \tag{66}$$

for $0<t\le 1$.

We make the following observations about a scaled family that satisfies the curvature condition (66).

(i) In any t-interval bounded away from zero, Φ_t is uniformly smooth in all variables, and of course satisfies the condition (61) uniformly in t.

(ii) The condition (66) is equivalent for small t, because of (65), to the nonvanishing of the Gaussian curvature of S_x^0, for each x. Note that $S_x^0 - x$ is the limiting hypersurface, obtained as

$$\lim_{t\to 0}\frac{S_{x,t}-x}{t};$$

see (64).

(iii) The basic property (66) is invariant under diffeomorphisms. Indeed, let $x\mapsto\psi(x)$ be a (local) diffeomorphism of $\mathbf{R}^n$, and define

$$\Phi'_t(x,y) = \Phi_t(\psi(x),\psi(y)).$$

Then the invariance for t bounded away from zero follows because of remark (b) above. For small t, we observe that if $S'_{x,t}$ is the family of hypersurfaces defined by Φ'_t, then

$$\psi(S'_{x,t}) = S_{\psi(x),t}\,,$$

and thus for the derived hypersurfaces we have

$$L(S'^0_x) = S^0_{\psi(x)}\,,$$

where L is the (linear) Jacobian derivative of the mapping $x\mapsto\psi(x)$. As a result, S'^0_x has nonzero Gaussian curvature if and only if the same is true for S^0_x.

3.1.2 Before coming to the statement of our main theorem, we need to dispose of one further preliminary matter: the meaning to be given to integrals taken over a hypersurface S determined by a defining function Φ. We first note two common ways of assigning measures to surfaces $S \subset \mathbf{R}^n$.

1. The first may be defined for any surface $S \subset \mathbf{R}^n$, regardless of its dimension $m < n$. It is the *induced Lebesgue measure*, usually called $d\sigma$, and defined by

$$\int_S f\, d\sigma = \lim_{\varepsilon \to 0} \frac{1}{\nu(B_\varepsilon)} \int_{S_\varepsilon} f(x)\, dx;$$

here S_ε is the "tubular neighborhood" consisting of all points $x \in \mathbf{R}^n$ with $\operatorname{dist}(x, S) < \varepsilon$ and $\nu(B_\varepsilon) = c_r \varepsilon^r$ is the (r-dimensional) volume of the r-dimensional ball B_ε of radius ε, where r is the codimension $n - m$.

2. The second may be defined only when S is a *hypersurface*, i.e., when the dimension of S is $n-1$. We suppose that Φ is a *defining function* for S, that is, $S = \{x : \Phi(x) = 0\}$, and $\nabla\Phi(x) \neq 0$ whenever $x \in S$. We then define the "Dirac measure" $\delta(\Phi)\, dy$ associated to Φ by

$$\int_{\mathbf{R}^n} f(y)\, \delta(\Phi)\, dy = \lim_{\varepsilon \to 0} \frac{1}{2\varepsilon} \int_{-\varepsilon < \Phi(x) < \varepsilon} f(x)\, dx. \tag{67}$$

Note that $\delta(\Phi)|_S = \dfrac{d\sigma}{|\nabla\Phi|}$.

In the case at hand, we will use the measure $d\sigma$ on S, where $d\sigma(y) = \delta(\Phi)\, dy$[‡] is defined by (67) for every continuous function f having compact support. If the support of f is sufficiently small, and if (after a translation and rotation) we choose a coordinate system so that $\partial\Phi/\partial y_n \neq 0$ in the support of f, then (in the support of f) we can represent S by

$$y_n = \phi(y'), \qquad y' = (y_1, \ldots, y_{n-1});$$

a straightforward calculation shows that the definition (67) implies

$$\int f(y)\, \delta(\Phi)\, dy = \int f(y', \phi(y')) \cdot \left| \frac{\partial \Phi}{\partial y_n} \right|^{-1} dy'. \tag{68}$$

With this in mind, given the scaled family $\Phi_t = \Phi_t(x, y)$, we define the averages A_t by

$$(A_t f)(x) = \int f(y)\, \psi_t\, \delta(\Phi_t)\, dy. \tag{69}$$

[‡] Note that now $d\sigma$ is *not* the induced Lebesgue measure, as it was in the first definition above and in other parts of the book. Also, we sometimes consider $\delta(\Phi)\, dy$ to be a measure on S when, strictly speaking, it is a measure on $\mathbf{R}^n$.

Here $\psi_t(x, y) = t^{-n}\psi(t, x, (x - y)/t)$, where $\psi(t, x, y)$ is a fixed smooth function of compact support in all the variables. The factor t^{-n} is introduced to normalize the total mass of the integral in (69), in view of the fact that the $n - 1$ dimensional measure of $S_{x,t}$ (in the support of ψ_t) is of the order t^n.

THEOREM 2. *Suppose that Φ_t is a scaled family, given for $0 < t \leq 1$, that satisfies the curvature condition* (66). *If*

$$\mathcal{A} f(x) = \sup_{0 < t \leq 1} |A_t f(x)|,$$

then we have the a priori *inequality*

$$\|\mathcal{A} f\|_{L^p(\mathbf{R}^n)} \leq A_p \|f\|_{L^p(\mathbf{R}^n)}, \tag{70}$$

for all continuous functions f having compact support, provided that

$$n \geq 3 \qquad \text{and} \qquad p > \frac{n}{n-1}.$$

3.2 Two lemmas. The proof of the theorem will require two lemmas. The first is entirely elementary. It makes precise the idea that, for functions of one variable, the supremum norm is controlled by the geometric mean[†] of the L^2 norms of the function and its first derivative.

LEMMA 1. *Suppose $F(t)$ is a smooth function for t in an interval I. Then for each ℓ with $\ell \leq |I|$, we have*

$$\sup_{t \in I} |F(t)| \leq \ell^{-1/2} \left(\int_I |F(t)|^2\, dt \right)^{1/2} + \ell^{1/2} \left(\int_I |F'(t)|^2\, dt \right)^{1/2}. \tag{71}$$

Proof. Each $t \in I$ belongs to an interval $I_0 \subset I$ with $|I_0| = \ell$. Observe that

$$F(t) - F(s) = \int_s^t F'(u)\, du,$$

so by Schwarz's inequality

$$|F(t)| \leq |F(s)| + \ell^{1/2} \left(\int_I |F'(t)|^2\, dt \right)^{1/2},$$

if $s \in I_0$. Take the mean value of the above inequality for s ranging through I_0. Then

$$|F(t)| \leq \ell^{-1} \int_{I_0} |F(s)|\, ds + \ell^{1/2} \left(\int_I |F'(t)|^2\, dt \right)^{1/2}.$$

Use Schwarz's inequality again, this time for the first integral above. The result is

$$|F(t)| \leq \ell^{-1/2} \left(\int_I |F(t)|^2\, dt \right)^{1/2} + \ell^{1/2} \left(\int_I |F'(t)|^2\, dt \right)^{1/2},$$

and (71) is established.

[†] At least when $\ell = (\int_I |F|^2 / \int_I |F'|^2)^{1/2} \leq |I|$.

3.2.1 The second lemma connects rotational curvature and oscillatory integrals.

LEMMA 2. *Let ψ be a fixed function of $(x, y, y_0) \in \mathbf{R}^n \times \mathbf{R}^n \times \mathbf{R}^1$ having compact support, whose projection on $\mathbf{R}^1$ does not contain the point $y_0 = 0$. Suppose $\Phi(x, y)$ satisfies*

$$\operatorname{rot\,curv}(\Phi) \neq 0$$

for all (x, y) in the support of ψ. If we define

$$(T_\lambda f)(x) = \int_{\mathbf{R}^{n+1}} e^{i\lambda y_0 \Phi(x,y)} f(y)\, \psi(x, y, y_0)\, dy\, dy_0,$$

then

$$\|T_\lambda(f)\|_{L^2(\mathbf{R}^n)} \leq A\lambda^{-(n+1)/2} \|f\|_{L^2(\mathbf{R}^n)}. \tag{72}$$

The main point of this lemma is that T_λ is an oscillatory integral, mapping functions on $\mathbf{R}^n$ to functions on $\mathbf{R}^n$, whose norm as an operator on L^2 is essentially $\lambda^{-(n+1)/2}$; this is in distinction to the factor $\lambda^{-n/2}$ which appeared in the proposition in Chapter 9, §1.1.

Proof. We define a new operator $\tilde{T}_\lambda$, mapping functions on $\mathbf{R}^n$ to functions on $\mathbf{R}^{n+1}$, by

$$(\tilde{T}_\lambda f)(x, x_0) = \int_{\mathbf{R}^{n+1}} e^{i\lambda x_0 y_0 \Phi(x,y)} f(y)\, \tilde{\psi}(x, y, x_0, y_0)\, dy\, dy_0. \tag{73}$$

Here $\tilde{\psi}(x, y, x_0, y_0) = \eta(x_0) \cdot \psi(x, y, y_0)$, with η a smooth function of x_0, supported in $1/2 \leq x_0 \leq 3/2$, and so that $\eta(x_0) = 1$. Now according to our assumptions, the phase function $\tilde{\Phi} = x_0 y_0 \Phi(x, y)$ has a nonvanishing Hessian in the support of $\tilde{\psi}$ (see (63)). Thus it follows easily from the proposition in Chapter 9, §1.1, when applied in $n + 1$ dimensions, that

$$\|\tilde{T}_\lambda(f)\|_{L^2(\mathbf{R}^{n+1})} \leq A\lambda^{-(n+1)/2} \|f\|_{L^2(\mathbf{R}^n)}. \tag{74}$$

We next apply the same reasoning to $(\partial/\partial x_0)\tilde{T}_\lambda f(x, x_0)$. Indeed, carrying out the differentiation with respect to x_0 in the integral in (73) merely brings down a factor of $i\lambda y_0 \Phi(x, y)$. However,

$$i\lambda y_0 \Phi(x, y) e^{i\lambda x_0 y_0 \Phi(x,y)} = \frac{y_0}{x_0} \frac{\partial}{\partial y_0} (e^{i\lambda x_0 y_0 \Phi(x,y)}).$$

Inserting this in (73) and carrying out the indicated integration by parts, noting that $f(y)$ is independent of y_0, gives us an integral of the same kind as $\tilde{T}_\lambda$, and then again

$$\left\| \frac{\partial}{\partial x_0} \tilde{T}_\lambda(f) \right\|_{L^2(\mathbf{R}^{n+1})} \leq A\lambda^{-(n+1)/2} \|f\|_{L^2(\mathbf{R}^n)}. \tag{75}$$

Now $(\tilde{T}_\lambda f)(x, x_0)|_{x_0=1} = (T_\lambda f)(x)$. Hence if we apply the previous lemma with $\ell = 1$, but for $1/2 \leq x_0 \leq 3/2$ instead of $0 \leq t \leq 1$, then (74) and (75) imply

$$\|T_\lambda(f)\|_{L^2(\mathbf{R}^n)} \leq A\lambda^{-(n+1)/2} \|f\|_{L^2(\mathbf{R}^n)},$$

and Lemma 2 is established.

Remark. For later purposes it is useful to point out that an examination of the proof of the lemma (as well as the proposition in §1.1 of Chapter 9) shows that the bound A in (72) depends only on the following: the size of the support of ψ, the distance of that support from the hyperplane $y_0 = 0$, upper bounds for the derivatives of Φ and ψ of order not exceeding $n + 2$, and a positive lower bound for $\operatorname{rot\,curv}(\Phi)$.

3.3 Proof when t is not small. We now describe how to control the averages A_t when t is bounded away from zero and the exponent p is restricted to the range $n/(n-1) < p \leq 2$. This is only part of what we are required to do, but it already contains the essence of the ideas needed. Thus, for a fixed $t_0 > 0$, we define

$$(\mathcal{A}_1 f)(x) = \sup_{t_0 \leq t \leq 1} |(A_t f)(x)|,$$

and shall show how to estimate $\mathcal{A}_1$ when $n/(n-1) < p \leq 2$.

First we take the precaution of modifying the cut-off function ψ_t appearing in the definition (69) so that ψ_t is supported only on the set where $\operatorname{rot\,curv}(\Phi_t) \neq 0$. Indeed, because of the delta function, what matters in (69) depends only on the immediate neighborhood of $\Phi_t = 0$, so the hypothesis (66) allows us to change ψ_t accordingly.

3.3.1 We next use the idea that the delta function can be written as

$$\delta(u) = \int_{-\infty}^{\infty} e^{2\pi i u\mu}\, d\mu,$$

which we will interpret in terms of the dyadic decomposition already used in Chapter 6, §4.1. If we choose α to be a smooth function on $\mathbf{R}^1$ so that (say) $\alpha(\mu) = 1$ for $|\mu| \leq 1$, and $\alpha(\mu) = 0$ for $|\mu| \geq 2$, then

$$\begin{aligned} \delta(u) &= \lim_{j\to\infty} \int_{-\infty}^{\infty} \alpha(2^{-j}\mu)\, e^{2\pi i u\mu}\, d\mu \\ &= \int_{-\infty}^{\infty} \alpha(\mu)\, e^{2\pi i u\mu}\, d\mu + \sum_{j=1}^{\infty} \int_{-\infty}^{\infty} \beta(2^{-j}\mu)\, e^{2\pi i u\mu}\, d\mu, \end{aligned}$$

where $\beta(\mu) = \alpha(\mu) - \alpha(2\mu)$ is supported in $1/2 \leq |\mu| \leq 2$.

With

$$\gamma(u) = \int \alpha(\mu)\, e^{2\pi i u\mu}\, d\mu,$$

this means

$$\delta(u) = \lim_{j\to\infty} 2^j \gamma(2^j u).$$

As in (67) and (68) we have, then,

$$\int_S f(y)\, \delta(\Phi)\, dy = \lim_{j\to\infty} 2^j \int \gamma(2^j \Phi)\, f(y)\, dy,$$

and, by (69), whenever f is continuous and has compact support, we can write (for each x and t)

$$(A_t f)(x) = \sum_{j=1}^{\infty} (A_t^j f)(x),$$

where

$$(A_t^j f)(x) = \int_{\mathbf{R}^1}\int_{\mathbf{R}^n} \beta(2^{-j}\mu)\, e^{2\pi i\mu\Phi_t(x,y)}\, f(y)\, \psi_t\, dy\, d\mu, \qquad (76)$$

for $j \geq 1$, with a similar expression for $j = 0$. If we make the change of variables $2^{-j}\mu = y_0$, this becomes

$$(A_t^j f)(x) = 2^j \int_{\mathbf{R}^{n+1}} e^{i\lambda y_0 \Phi_t(x,y)}\, f(y)\, \psi_t^*\, dy\, dy_0, \qquad \lambda = 2\pi\cdot 2^j, \quad (77)$$

with $\psi_t^*(x, y, y_0) = t^{-n}\psi(t, x, (x-y)/t)\beta(y_0)$ (see (69)).

Therefore Lemma 2 and the remarks that follow its proof guarantee that

$$\|A_t^j(f)\|_{L^2(\mathbf{R}^n)} \leq A 2^j \cdot 2^{-j(n+1)/2} \|f\|_{L^2(\mathbf{R}^n)}, \qquad t_0 \leq t \leq 1. \qquad (78)$$

A similar estimate applies to $(\partial/\partial t)A_t^j(f)$, except that now an additional factor of $\lambda = 2\pi\cdot 2^j$ is brought down by the differentiation applied to $e^{i\lambda y_0\Phi_t}$. This gives us

$$\left\|\frac{\partial}{\partial t}A_t^j(f)\right\|_{L^2(\mathbf{R}^n)} \leq A\cdot 2^{2j}\cdot 2^{-j(n+1)/2}\|f\|_{L^2(\mathbf{R}^n)}, \qquad t_0 \leq t \leq t_1. \quad (79)$$

We now apply Lemma 1 in the form

$$\sup_{t_0\leq t\leq 1} |A_t^j f(x)|^2 \leq A\left(\ell^{-1}\int_{t_0}^1 |A_t^j f(x)|^2\, dt + \ell\int_{t_0}^1 \left|\frac{\partial}{\partial t}A_t^j f(x)\right|^2 dt\right). \quad (80)$$

We take $\ell = (1-t_0)2^{-j}$, integrate both sides of (80) with respect to x, and apply (78) and (79). After simplification the result is

$$\|\sup_{t_0\leq t\leq 1} |A_t^j f|\,\|_{L^2(\mathbf{R}^n)} \leq A 2^{-j(n-2)/2}\|f\|_{L^2(\mathbf{R}^n)}; \qquad (81)$$

this is the key L^2 estimate. We will end up summing such estimates in j. For this to converge, we need that $(n-2) > 0$; this is where the dimension restriction $n \geq 3$ comes in.

There is also a very crude L^1 estimate that comes about by writing

$$A_t^j f(x) = 2^j \int \gamma(2^j\Phi_t)\, f(y)\, \psi_t^*\, dy - 2^{j-1}\int \gamma(2^{j-1}\Phi_t)\, f(y)\, \psi_t^*\, dy, \quad (82)$$

which follows from (76) and the fact that

$$2^j\gamma(2^j u) - 2^{j-1}\gamma(2^{j-1}u) = \int_{-\infty}^{\infty} \beta(2^{-j}\mu)\, e^{2\pi i u\mu}\, d\mu.$$

This last expression for A_t^j makes it clear that

$$\|\sup_{t_0\leq t\leq 1} |A_t^j f|\,\|_{L^1(\mathbf{R}^n)} \leq A 2^j \|f\|_{L^1(\mathbf{R}^n)}.$$

3.3.2 We now combine the L^2 and L^1 estimates by complex interpolation. For this purpose, let $t(x)$ be an arbitrary measurable function of $x \in \mathbf{R}^n$, with values in $[t_0, 1]$, and set

$$(Uf)(x) = (A_{t(x)}^j f)(x).$$

We then apply the interpolation lemma given in Chapter 9, §1.2.5, with $U = U^s$ independent of s. This yields

$$\|Uf\|_{L^p} \leq A 2^{-(1-\theta)j(n-2)/2}\cdot 2^{j\theta}\|f\|_{L^p},$$

where $1/p = (1-\theta)/2 + \theta/2$. We next take the supremum over all functions of the form $t(x)$ described above. The result is that

$$\sup \|Uf\|_{L^p} = \|\sup_{t_0\leq t\leq 1} |A_t^j(f)|\,\|_{L^p} \leq A 2^{j\delta}\|f\|_{L^p}, \qquad (83)$$

where $\delta = -(1-\theta)(n-2)/2 + \theta = -n + 1 + n/p$. We see from this that $\delta < 0$ if $p > n/(n-1)$ (and of course $p \leq 2$).

Therefore, since

$$\sup_{t_0\leq t\leq 1} |A_t f| \leq \sum_j \sup_{t_0\leq t\leq 1} |A_t^j f|,$$

we have by (83)

$$\|\sup_{t_0\leq t\leq 1} |A_t^j f|\,\|_{L^p} \leq A\|f\|_{L^p},$$

and so

$$\|\mathcal{A}_1 f\|_{L^p} \leq A\|f\|_{L^p}$$

whenever $p > n/(n-1)$ and $p \leq 2$.

3.4 Proof when t is near zero. We next turn to the study of $A_t f$ for t small. Our first task will be to see that (78) continues to be valid in this range; that is

$$\|A_t^j(f)\|_{L^2} \le A \cdot 2^{-j(n-1)/2}\|f\|_{L^2}, \qquad \text{for } 0 < t \le t_0, \tag{84}$$

when t_0 is sufficiently small.

The proof of this is as follows. Let $K(x,y)$ be the kernel of the operator A_t^j. Then the norm of A_t^j is the same as that of the operator whose kernel is $t^n K(tx,ty)$.‡

Now looking back at (77), recalling that the scaled family Φ_t is given (in terms of the parametrized family Φ^t) by

$$\Phi_t(x,y) = \Phi^t(x, x + (y-x)/t)$$

and that the cut-off function entering in A_t^j is

$$t^{-n}\psi(t,x,(y-x)/t),$$

we see that the operator we need to consider has kernel

$$2^j \int e^{i\lambda y_0 \Phi^t(tx, y-x+tx)}\, \psi(t,tx,x-y)\,\beta(y_0)\,dy_0. \tag{85}$$

In view of the assumptions (65), (66) near $t=0$, we can again invoke Lemma 2 and see that the operator with kernel (85) has a bound not exceeding

$$A2^j \cdot \lambda^{-(n+1)/2} \qquad \text{with } \lambda = 2\pi 2^j,$$

and this proves (84).

Again, inspection of (77) reveals that $(\partial/\partial t)A_t^j(f)$ is of the same form as A_t^j, but the differentiation has the effect of bringing down a factor of $2^j/t$. Thus one has

$$\left\|\frac{\partial}{\partial t}A_t^j f\right\|_{L^2} \le A(2^j/t)2^{-j(n-1)/2}\|f\|_{L^2}, \qquad 0 < t \le t_0. \tag{86}$$

3.4.1 There is a minor difficulty with the argument up to this point. When we apply Lemma 2 (in §3.2.1), its hypotheses require that the cut-off function ψ have compact support (in x and y). However, in (85), the relevant part of the cut-off function is $\psi(t,tx,x-y)$, which does have compact support in $x-y$, but not a fixed compact support in x, as $t \to 0$.

To get around this point, we pre-multiply the operator (given by kernel (85)) by cut-off functions in x, whose supports are of uniform size as $t \to 0$, obtaining thereby localized versions of the desired estimates. Since the kernel has uniform compact support in $x-y$, we can then complete the argument, obtaining the global result by appealing to an integration argument that we have used several times before, most recently at the end of Chapter 9, §2.2.2; see also §3.5.5 below. We leave the details to the interested reader.

‡ The reader should observe that the dilation in the (x,y) space used here is not the same as that used in the definition of the scaled family Φ_t.

3.4.2 Another parametrization. The reason we cannot proceed directly with (86) and (84) (as we did with (79) and (78)) is the occurrence of the additional factor t^{-1} in (86). One gets around this by an analysis that reveals that the main contribution to $A_t^j(f)$ comes essentially from certain "frequencies" of f of size roughly $2^j/t$. This suggests that we should use an orthogonal decomposition of f, in terms of the Fourier transform.

These considerations require that we choose an equivalent family $\tilde{\Phi}_t$ representing our hypersurfaces, and we then need to modify our operators accordingly. The choice of $\tilde{\Phi}_t$ will be made so that, in an appropriate coordinate system, writing $y = (y', y_n)$ with $y' \in \mathbf{R}^{n-1}$, we have that $\tilde{\Phi}(x,y)$ depends linearly on $y_n \in \mathbf{R}^1$; this will allow us to deal with the analogues of the A_t^j by using the Fourier transform in the y_n variable.

We proceed as follows. First, by restricting ourselves to a sufficiently small compact set, we can assume that our "parametrized" hypersurfaces S_x^t given by $\Phi^t(x,y) = 0$, are represented by

$$\tilde{\Phi}^t(x,y) \equiv \phi_t(x,y') - y_n = 0,$$

where $\phi_t(x,y')$ is smooth for $(t,x,y') \in [0,t_0] \times \mathbf{R}^n \times \mathbf{R}^{n-1}$ with t_0 sufficiently small, once we have chosen the coordinate system $y = (y', y_n)$ appropriately. This means that, in our case, we take

$$\tilde{\Phi}_t(x,y) = \phi_t'\left(x, \frac{y'-x'}{t}\right) - \frac{y_n - x_n}{t}. \tag{87}$$

We define A_t as before (see (69)), except we modify the cut-off function ψ_t to be independent of $x_n - y_n$, i.e., we replace it with a $\tilde{\psi}_t$ of the form $t^{-n}\tilde{\psi}(t,x,(x'-y')/t)$. Thus we write

$$(\tilde{A}_t f)(x) = \int f(y)\tilde{\psi}_t\,\delta(\tilde{\Phi}_t)\,dy. \tag{88}$$

Since the δ-function limits the above integration to the set where $\tilde{\Phi}_t = 0$, the fact that no restriction on the size of $x_n - y_n$ is made by $\tilde{\psi}$ is immaterial. In fact, for a suitable choice of $\tilde{\psi}_t$, the operator $\tilde{A}_t$ is actually identical to A_t.

With the realization (88) in hand, we define the operators $\tilde{A}_t^j$ in the same way as those defined earlier (by (76)); that is

$$(\tilde{A}_t^j f)(x) = \int_{\mathbf{R}^n \times \mathbf{R}^1} \beta(2^{-j}\mu)\,e^{2\pi i\mu\tilde{\Phi}_t(x,y)}\,f(y)\,\tilde{\psi}_t\,dy\,d\mu. \tag{89}$$

However, now $\tilde{A}_t^j$ does differ from A_t^j, because in principle, the integration is over the whole space. We adjust the cut-off function determining A_t^j by taking

$$\psi_t = \tilde{\psi}_t \cdot \psi'\left(\frac{x_n - y_n}{t}\right),$$

where ψ' is a $C_0^\infty(\mathbf{R}^1)$ function with $\psi'(u) = 1$ for $|u| \le c$.

In addition, observe that $A_t^j - \tilde{A}_t^j$ has a kernel majorized by

$$A 2^{-jN} \chi\left(\frac{x-y}{t}\right) t^{-n},$$

where χ is the characteristic function of a ball of bounded radius, and N is an arbitrary positive integer. This comes about by observing that the Fourier transform of β is rapidly decreasing; also, in the support of $1 - \psi'((x_n - y_n)/t)$, we have $|(x_n - y_n)/t| \ge c$, and thus

$$|\tilde{\Phi}_t| \approx \left|\frac{x_n - y_n}{t}\right|,$$

if c is chosen sufficiently large. It follows from this, and the standard maximal theorem, that

$$\left\| \sup_{0<t\le t_0} |A_t^j f - \tilde{A}_t^j f| \right\|_{L^2} \le A 2^{-jN} \|f\|_{L^2}. \tag{90}$$

Similarly, but more directly,

$$\left\| \frac{\partial}{\partial t}(A_t^j f - \tilde{A}_t^j f) \right\|_{L^2} \le \frac{A}{t} \cdot 2^{-jN} \|f\|_{L^2}. \tag{91}$$

3.4.3 Dyadic decomposition. Next, we invoke a variant of the Littlewood-Paley "dyadic projection" operators, acting in the y_n direction only. If $f(y) = f(y', y_n)$, we define $\hat{f}(y', \xi_n)$ to be the Fourier transform in the y_n variable, and set

$$(\Delta_k f)(x', x_n) = \int_{\mathbf{R}^1} \tilde{\beta}(2^{-k}\xi_n)\, e^{2\pi i x_n \xi_n}\, \hat{f}(x', \xi_n)\, d\xi_n, \tag{92}$$

where $\tilde{\beta} \in C_0^\infty(\mathbf{R}^1)$ is chosen so that $\tilde{\beta}(\mu) = 1$ for $1/4 \le |\mu| \le 2$, and $\tilde{\beta}(\mu) = 0$ for $|\mu| \le 1/8$ or $|\mu| \ge 4$.

Notice that, in (89), the cut-off function is independent of y_n. Because of this, and by the form of (87), it is easy to see that

$$\tilde{A}_t^j \Delta_k = \tilde{A}_t^j \tag{93}$$

whenever $\tilde{\beta}(2^{-k}\mu) = 1$ on the support of $\beta(2^{-j}\mu t)$. In other words, (93) holds for $2^{-m} \le t \le 2^{-m+1}$ when $k = m + j$.

In view of (90) and (91), we have the same estimates (84) and (86) for $\tilde{A}_t^j$ instead of A_t^j. Next, we see that by Lemma 1 in §3.2,

$$\sup_{2^{-m}\le t\le 2^{-m+1}} |(\tilde{A}_t^j f)(x)|^2 \le A\Big(\ell^{-1}\int_{2^{-m}}^{2^{-m+1}} |(\tilde{A}_t^j f)(x,t)|^2\, dt + \ell \int_{2^{-m}}^{2^{-m+1}} \left|\frac{\partial}{\partial t}(\tilde{A}_t^j f)(x,t)\right|^2 dt\Big),$$

whenever $\ell \le 2^{-m-1}$.

Now integrate both sides with respect to x, and use (93) for $2^{-m} \le t \le 2^{-m+1}$, together with the L^2 estimates for $\tilde{A}_t^j$ and $(\partial/\partial t)\tilde{A}_t^j$ just mentioned. The result, using that $\tilde{A}_t^j = \tilde{A}_t^j \Delta_k$, is

$$\left\| \sup_{2^{-m}\le t\le 2^{-m+1}} |\tilde{A}_t^j(f)| \right\|_{L^2}^2 \le A(\ell^{-1}\, 2^{-m}\, 2^{-j(n-1)} + \ell\, 2^{-m}\, 2^{2j+2m}\, 2^{-j(n-1)}) \cdot \|\Delta_k(f)\|_{L^2}^2.$$

If we now choose $\ell = 2^{-j-m}$, this yields

$$\left\| \sup_{2^{-m}\le t\le 2^{-m+1}} |\tilde{A}_t^j(f)| \right\|_{L^2}^2 \le A 2^{-j(n-2)} \|\Delta_k(f)\|_{L^2}^2.$$

With j fixed, we sum over m, noting that

$$\sum_k \|\Delta_k(f)\|_{L^2}^2 \le A\|f\|_{L^2}^2.$$

Hence

$$\left\| \sup_{0<t\le t_0} |\tilde{A}_t^j(f)| \right\|_{L^2} \le A 2^{-j(n-2)/2} \|f\|_{L^2}. \tag{94}$$

3.4.4 There remain only a few small steps in the proof of Theorem 2. From the realization (82), we see that the kernel of A_t^j is dominated by $2^j \chi[(x-y)/t]t^{-n}$; here χ is the characteristic function of a fixed ball in $\mathbf{R}^n$. We have already noted a similar (in fact better) estimate for the kernel of $A_t^j - \tilde{A}_t^j$. The result of this is that

$$\sup_{0<t\le t_0} |(\tilde{A}_t^j f)(x)| \le A 2^j (Mf)(x),$$

where M is the standard maximal operator, and so

$$\left\| \sup_{0<t\le t_0} |\tilde{A}_t^j(f)| \right\|_{L^{p_1}(\mathbf{R}^n)} \le A_{p_1} 2^j \|f\|_{L^{p_1}(\mathbf{R}^n)}, \qquad p_1 > 1.$$

We combine this with (94) via interpolation, in the same way as the argument of §3.3.2. As a consequence, we have

$$\| \sup_{0<t\le t_0} |\tilde{A}_t^j(f)| \|_{L^p} \le 2^{j\delta'} \|f\|_{L^p},$$

where $\delta' < 0$, if $n/(n-1) < p \le 2$.

Summing in j then gives us

$$\| \sup_{0<t\le t_0} |A_t^j(f)| \|_{L^p} \le A\|f\|_{L^p}, \qquad \text{when} \quad \frac{n}{n-1} < p \le 2.$$

Together with what we have already proved in §3.3, this yields the required conclusion for $\mathcal{A}$, when $n/(n-1) < p \le 2$. To obtain the full range $n/(n-1) < p \le \infty$, we simply interpolate between the above result and the trivial case at $p = \infty$. The proof of Theorem 2 is thereby concluded.

3.5 Corollaries and some further remarks. Theorem 2 was formulated in its *a priori* form: as a majorization of the p-norms of the maximal function $\mathcal{A}f$, where f is assumed to be a continuous function of compact support. We now want to extend this operator to all of L^p; however, this is not entirely straightforward. We first rewrite (69) as

$$(A_t f)(x) = \int_{S_{x,t}} f(y)\, d\sigma_{x,t}(y),$$

where $d\sigma_{x,t}$ is the measure on $S_{x,t}$ given by $\delta(\Phi_t)\psi_t\, dy$; it will be convenient here to make the harmless assumption that $d\sigma_{x,t}$ is nonnegative, i.e., that $\psi_t \geq 0$.

Suppose $f \in L^p(\mathbf{R}^n)$. The first question that arises is whether $A_t f(x)$ is well-defined for a given (x,t). Put another way, we ask whether $f|_{S_{x,t}}$ is integrable (in particular, measurable) with respect to the measure $d\sigma_{x,t}$. When this is so, $(A_t f)(x)$ is defined as the value of the corresponding integral.

COROLLARY 1. *Suppose $f \in L^p(\mathbf{R}^n)$, $p > n/(n-1)$, $n \geq 3$. Then, for almost every x, $(A_t f)(x)$ is well-defined for* all t, $0 < t \leq 1$; *also, the resulting function of t is continuous. Moreover, the inequality* (70) *holds for*

$$(\mathcal{A}f)(x) = \sup_{0<t\leq 1} |(A_t f)(x)|.$$

3.5.1 A key point in the proof of the corollary is the following (nonobvious) special case.

LEMMA. *The assertions of the corollary hold when $f = \chi_E$, where E is a set of measure zero. Equivalently, for almost every x, the set $E \cap S_{x,t}$ has measure zero (with respect to $d\sigma_{x,t}$) for* all $t > 0$.

To begin with, let us limit ourselves to *Borel* measurable functions and sets. Notice that if f is bounded and Borel-measurable then $f|_{S_{x,t}}$ is automatically Borel-measurable on $S_{x,t}$, hence $(A_t f)(x)$ is well-defined for every (x,t). Moreover, if $0 \leq f_n \uparrow f$, then clearly $\mathcal{A}(f_n) \uparrow \mathcal{A}(f)$.

Now if $f = \chi_O$ is the characteristic function of an open set, we can find f_n that are continuous and have compact support with $f_n \uparrow f$; therefore, by (70), we have

$$\|\mathcal{A}(\chi_O)\|_{L^p} \leq A_p |O|^{1/p}, \quad \text{for any fixed } p, \quad \frac{n}{n-1} < p < \infty.$$

However, if $|E| = 0$, there is a sequence of open sets $O_n \supset E$ with $|O_n| \to 0$. Since $\mathcal{A}(\chi_E) \leq \mathcal{A}(\chi_{O_n})$ for all n and $\|\chi_{O_n}\|_{L^p} \to 0$,

$$\|\mathcal{A}(\chi_E)\|_{L^p} = 0;$$

hence $\mathcal{A}(\chi_E) = 0$ almost everywhere, establishing the desired result when E is Borel-measurable. Since every Lebesgue-measurable set of measure zero is contained in a Borel set of measure zero, our lemma is proved.

3.5.2 Next let f be a bounded (Borel-measurable) function with compact support. Then there is a sequence of continuous functions f_k, having a common compact support, so that $f_k(x) \to f(x)$ almost everywhere with $\|f - f_k\|_{L^p} \leq 2^{-k}$. The maximal inequality (70) appled to $f_k - f_j$ shows that, for almost every x,

$$\lim_{k\to\infty} (A_t f_k)(x)$$

exists, uniformly in t, $0 < t \leq 1$.

Clearly, for each k and x, $(A_t f_k)(x)$ is continuous in t; hence

$$\lim_{k\to\infty} (A_t f_k)(x)$$

is also continuous in t, for almost every x. Since $f_k(x) \to f(x)$ a.e., the lemma, together with the dominated convergence theorem, gives

$$\lim_{k\to\infty} (A_t f_k)(x) = (A_t f)(x), \qquad \text{a.e. } x.$$

Thus we have proven the corollary for Borel-measurable f that are bounded and have compact support. An easy limiting argument, using the lemma, extends these conclusions to general f and completes the proof of Corollary 1.

3.5.3 Now let us assume that, for a given compact set K, we have normalized the measure $d\sigma_{x,t}$ so that

$$\int_{S_{x,t}} d\sigma_{x,t} = 1, \qquad \text{for all } 0 < t \leq 1,\ x \in K.$$

COROLLARY 2. *Suppose that f is locally in $L^p(\mathbf{R}^n)$, for some $p > n/n-1$. Then*

$$\lim_{t\to 0}(A_t f)(x) = f(x) \tag{95}$$

for almost every $x \in K$.

This follows in the usual way from the maximal inequality, as guaranteed by Corollary 1 and the fact that (95) holds (everywhere in K) for compactly supported continuous functions.

3.5.4 As we remarked in §1.1.3, the main inequality (70), and the conclusions deduced from it, may fail when $p \leq n/(n-1)$. Here we want to point out that our results in their general formulation, which we established when $n \geq 3$, may cease to be valid when $n < 3$. Indeed, given any $\delta > 0$, it is not difficult to construct a scaled family Φ_t, with the following properties: It is defined for both x and y outside the unit ball and, on any compact set there, it satisfies the basic curvature assumption (66) for $0 < t \le 1$. Moreover, when $\delta \le t \le 1$, we have that

$$\Phi_t(x,y) = x \cdot y + t;$$

see also the example (ii) in §3.1.

Note that when $n = 2$, the manifolds $S_{x,t}$ are segments of lines (if $\delta \le t \le 1$) whose direction is fixed by x and whose distance from the origin varies with t. As a result, no inequality of the type (70) can hold, as is seen by testing it against the characteristic function of the Besicovitch set treated in Chapter 10, §1 and §3.1(c).

3.5.5 Dilations of a fixed hypersurface. We return to the special case of the spherical maximal function first treated in §1.1 under the limitation $n \ge 4$, $p = 2$. We now deal with it in a context of intermediate generality: our setting will not be limited to spheres but will be translation-invariant.

The starting point will be a fixed smooth hypersurface S in $\mathbf{R}^n$ whose Gaussian curvature is nowhere zero. We shall denote by $d\sigma$ the induced Lebesgue measure on S, but modified by multiplication with a fixed (but arbitrary) smooth function of compact support. We let $d\sigma_t$ denote the dilated measure, i.e.,

$$\int f(x)\, d\sigma_t(x) = \int f(tx)\, d\sigma(x),$$

for continuous f having compact support, and set

$$(\mathcal{A}f)(x) = \sup_{0<t<\infty} |(f * d\sigma_t)(x)|. \tag{96}$$

COROLLARY 3. *The operator $\mathcal{A}$, given by* (96), *extends to a bounded operator from $L^p(\mathbf{R}^n)$ to itself when $p > n/(n-1)$ and $n \ge 3$, and the analogues of corollaries 1 and 2 also hold, with $A_t(f) = f * d\sigma_t$.*

Remark. The corollary follows easily from our considerations above, as we will see momentarily. However, as opposed to the general result, the corollary is still valid when $n = 2$. This special case requires additional ideas; see §4.11–§4.13 below.

To prove the corollary, fix η to be a smooth function of compact support that equals 1 on the unit ball, and set

$$A_t f(x) = \eta(x) \cdot (f * d\sigma_t)(x).$$

Then these averages, which are now appropriately localized, are of the type treated by Theorem 2 when (say) $0 < t \le 1$, and thus

$$\int_{|x|\le 1} \sup_{0<t\le 1} |f * d\sigma_t|^p\, dx \le A_p \int_{|y|\le r+1} |f(y)|^p\, dy,$$

if $d\sigma$ is supported on the ball of radius r centered at the origin. By translation invariance, we have, for all $x_0 \in \mathbf{R}^n$,

$$\int_{|x-x_0|\le 1} \sup_{0<t\le 1} |f * d\sigma_t|^p\, dx \le A_p \int_{|y-x_0|\le r+1} |f(y)|^p\, dy,$$

and an integration in x_0 gives

$$\|\mathcal{A}_1 f\|_{L^p} \le A'_p \|f\|_{L^p}, \tag{97}$$

where $(\mathcal{A}_R f)(x) = \sup\limits_{0<t<R} |(f * d\sigma_t)(x)|$.

A simple scaling argument shows that (97) holds (with the same bound) if $\mathcal{A}_1$ is replaced by $\mathcal{A}_R$. Letting $R \to \infty$ then gives us the desired inequality for $\mathcal{A}$. The other assertions of the corollary follow in the same way.

3.5.6 As was remarked above, the kinds of conclusions we have obtained can fail completely if all curvature conditions are dropped. However, in the context of §3.5.5 above, when the Gaussian curvature of S does not vanish to infinite order at any point, then there is a $p_0 = p_0(S) < \infty$ so that the L^p inequalities for $\mathcal{A}$ still hold, if $p_0 < p \le \infty$.

A more general result of this type is also valid in the setting of Theorem 2: If we are given a scaled family Φ_t whose rotational curvature does not vanish to infinite order in an appropriate uniform sense, then there are L^p inequalities for $\mathcal{A}$ when $p > p_0$. Further details are in §4.9.

4. Further results

A. Curvature and square functions

4.1 A modification of the argument in §1.1 yields the following more general maximal theorem for L^2.

Suppose $m(\xi)$ is given on $\mathbf{R}^n$ and satisfies

$$|m(\xi)| \le A(1+|\xi|)^{-\frac{1}{2}-\varepsilon}, \qquad |\nabla m(\xi)| \le A(1+|\xi|)^{-\frac{1}{2}-\varepsilon},$$

for some $\varepsilon > 0$. Define A_t by $\widehat{A_t f}(\xi) = m(t\xi)\widehat{f}(\xi)$, whenever $f \in \mathcal{S}$. If $(\mathcal{A}f)(x) = \sup\limits_{t>0} |(A_t f)(x)|$, then we have the *a priori* inequality

$$\|\mathcal{A}f\|_{L^2} \le c\|f\|_{L^2}.$$

Even more generally, the same conclusion holds if $\widehat{A_t f}(\xi) = m(t,\xi)\widehat{f}(\xi)$ with

$$|m(t,\xi)| \le A(1+t|\xi|)^{-\frac{1}{2}-\varepsilon}, \qquad \text{and} \qquad \left|\frac{\partial}{\partial t} m(t,\xi)\right| \le A|\xi|(1+t|\xi|)^{-\frac{1}{2}-\varepsilon}.$$

See Stein [1976a], Sogge and Stein [1985], Rubio de Francia [1986].

4.2 We saw in §1.1.3 that, for all $p \leq n/(n-1)$ (or $p < \infty$ when $n = 1$), there is an $f \in L^p(\mathbf{R}^n)$ so that $\sup_{t>0}(A_t f)(x) = \infty$ everywhere, where A_t is the mean-value operator given by

$$(A_t f)(x) = \int_S f(x - ty)\, d\sigma(y);$$

here S is the unit sphere in $\mathbf{R}^n$.

A somewhat similar example indicates what happens when we replace the sphere by a hypersurface S whose curvature is allowed to vanish, and how this affects the possible range of p in the corresponding maximal estimates. We consider the surface $S \subset \mathbf{R}^n$ given by

$$x_n^2 + (x_1^2 + \cdots + x_{n-1}^2)^k = 1,$$

where $k \geq 2$ is an integer; the curvature of S then vanishes at the points $(0, \ldots, 0, \pm 1)$.

Suppose

$$f(x) = \frac{|x_n|^{(1-n)/2k}}{\log(1/|x_n|)} \qquad \text{for } |x| < 1/2,$$

and $f(x) = 0$ when $|x| \geq 1/2$. Then $\sup_{t>0}(A_t f)(x) = \infty$ for all x. Note that $f \in L^p(\mathbf{R}^n)$ if $p \leq 2k/(n-1)$.

An easy modification of this example shows that we can choose a hypersurface $S \subset \mathbf{R}^n$ whose Gaussian curvature vanishes at only one point (but of infinite order there) so that no L^p estimates hold when $p < \infty$. To see this we take, for instance, S to be given near the point $(0, \ldots, 0, 1)$ as the graph

$$x_n = 1 - \exp(-[x_1^2 + \cdots + x_{n-1}^2]^{-1}),$$

and test the averages A_t against an f with $f(x) = |x_n|^{-\varepsilon}$, for $|x| \leq 1$.

4.3 We turn to the maximal averages over a fixed submanifold (treated in §2) and consider the situation of a curve in $\mathbf{R}^2$ given by $t \mapsto \gamma(t)$, where γ is a smooth function of $t \in [0,1]$. The maximal operator in question is

$$(\mathcal{M}f)(x) = \sup_{0 \leq h \leq 1} h^{-1} \left| \int_0^h f(x - \gamma(t))\, dt \right|.$$

We saw that when the curvature at $t = 0$ did not vanish to infinite order (and hence the curve was of finite type there), the maximal inequalities in this setting hold for all p, $1 < p \leq \infty$. However, when the curve is allowed to be infinitely flat at the origin, all these results may fail.

To see this, consider the "staircase" curve $\gamma(t)$ which for t in the middle half of the interval $[2^{-k}, 2^{1-k}]$ is the line segment $\{(t, 2^{-2^k})\}$ above that portion of the x-axis, $k = 1, 2, \ldots$ and extend γ so that it is smooth for all $t \in [0,1]$; it is possible to complete the given "steps" to a C^∞ curve because $2^{-2^k}/2^{-k} \to 0$ rapidly, as $k \to \infty$ (see Figure 1).

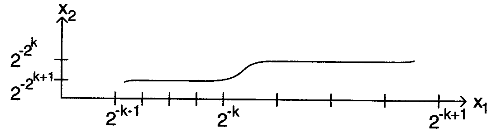

Figure 1. A section of the curve γ.

Let χ_{R_ε} denote the characteristic function of the thin rectangle

$$R_\varepsilon = \{(x_1, x_2) : |x_1| < 1, |x_2| < \varepsilon\}.$$

It is easy to see that $(Mf)(x) > 1/4$ whenever x belongs to the rectangle

$$[2^{-k}, 2^{1-k}] \times [2^{-2^k}, 2^{-2^k} + \varepsilon].$$

Since the first $c \log\log \varepsilon^{-1}$ such rectangles are disjoint, we see that there is a set E_ε, with $|E_\varepsilon| \geq c \log\log \varepsilon^{-1}$, so that $(\mathcal{M}\chi_{R_\varepsilon})(x) > 1/4$ for $x \in E_\varepsilon$.

This shows that $\mathcal{M}$ is not of weak-type (p,p) for any $p < \infty$, and thus, by §2 of Chapter 10, there is an $f \in L^p$ with $(\mathcal{M}f)(x) = \infty$ almost everywhere. Moreover, a modification of the argument in §3.3 of that chapter allows us to construct a set F_ε with $|F_\varepsilon| < \varepsilon$ and

$$\limsup_{h \to 0} h^{-1} \int_0^h \chi_{F_\varepsilon}(x - \gamma(t))\, dt \geq \frac{1}{4} \qquad \text{for a.e. } x.$$

Stein and Wainger [1978].

B. Singular Radon transforms

4.4 There is a wide-ranging parallelism between maximal functions and singular integrals; this was already stressed in Chapter 1. Here we want to describe singular integral versions of the maximal theorems considered in §2. We shall first consider the analogue of the main technical proposition formulated in §2.5.

As before, we are given a sequence $\{dm^j\}$ of measures on $\mathbf{R}^N$ which we assume have uniformly bounded norms, $\int |dm^j| \leq A$, and are supported in a common compact set. Again, the crucial assumption is on the decay of their Fourier transforms; we suppose that

$$|\widehat{dm^j}(\xi)| \leq A|\xi|^{-\delta},$$

where $\delta > 0$ and A are independent of j. In addition, we require here the cancellation condition

$$\int_{\mathbf{R}^N} dm^j = 0, \qquad \text{for all } j.$$

Using the nonisotropic dilations $x \mapsto \delta \circ x$ on $\mathbf{R}^N$, we let $d\mu_j$ denote the dilated measures given by

$$\widehat{d\mu^j} = \widehat{dm^j}(2^{-j} \circ \xi),$$

or equivalently, $\int_{\mathbf{R}^N} F(x)\, d\mu_j(x) = \int_{\mathbf{R}^N} F(2^j \circ x)\, dm_j(x)$.

Then we can assert that the sum $\sum d\mu_j$ converges to a tempered distribution K, and that the operator T, given by $Tf = f * K$, extends to a bounded operator from $L^p(\mathbf{R}^N)$ to itself, for $1 < p < \infty$.

The pattern of proof of this assertion is the same as that in §2.5. We imbed the operator T in an analytic family T_s, with $T_s f = f * K_s$ and $T_0 = T$.

If $\mathrm{Re}(s) < \delta$, then T_s is bounded on $L^2(\mathbf{R}^N)$ because $\widehat{K_s}(\xi) \in L^\infty(\mathbf{R}^N)$; for $\mathrm{Re}(s) < 0$, one has that T_s is bounded on $L^p(\mathbf{R}^N)$, $1 < p < \infty$, by virtue of the more standard theory of singular integrals.

The definition of T_s is as follows. Let the distributions M_j^s, ν_j^s be given by

$$\widehat{M_j^s}(\xi) = \widehat{dm^j}(\xi) \cdot (1 + |\xi|^2)^{s/2}, \qquad \widehat{\nu_j^s}(\xi) = \widehat{M_j^s}(2^{-j} \circ \xi),$$

and set $K_s = \sum_j \nu_j^s$. The convergence of this sum (in the sense of distributions) and the bound $|\widehat{K_s}(\xi)| \le A_s$, for $\mathrm{Re}(s) < \delta$, follow easily from the estimates

$$|\widehat{M_j^s}(\xi)| \le A|\xi|, \qquad \text{and} \qquad |\widehat{M_j^s}(\xi)| \le A|\xi|^{\mathrm{Re}(s) - \delta}.$$

The estimates of T_s for $\mathrm{Re}(s) < 0$ are carried out as in §2.5.2, but in the setting of singular integrals, for which see also Chapter 2, §5.15(b).

4.5 We formulate the singular integral analogue of the principal results, namely that of Theorem 1 when S is of finite type, and that of the model case for polynomials, given in Proposition 1 of §2.

Let $K(t)$ be a Calderón-Zygmund kernel in $\mathbf{R}^k$ of the following kind: K is homogeneous of degree $-k$, smooth away from the origin, and

$$\int_{|t|=1} K(t)\, d\sigma(t) = 0.$$

Suppose first that γ is a smooth mapping of the unit ball in $\mathbf{R}^k$ to $\mathbf{R}^n$ that is of finite type at the origin (in the sense of Chapter 8, §3.2). Define the operator

$$(T_\gamma f)(x) = \text{p.v.} \int_{|t|\le 1} f(x - \gamma(t))\, K(t)\, dt,$$

for $f \in \mathcal{S}$.

Next let $P : \mathbf{R}^k \to \mathbf{R}^n$ be any polynomial map and define

$$(T_P f)(x) = \text{p.v.} \int_{\mathbf{R}^k} f(x - P(t))\, K(t)\, dt,$$

for $f \in \mathcal{S}$. We can then assert:

T_γ and T_P extend to bounded operators from $L^p(\mathbf{R}^n)$ to itself, for $1 < p < \infty$.

To prove the statement for T_γ, we begin with the lifting technique used in §2.6. We write the Taylor expansion $\gamma(t) = P(t) + R(t)$ and construct a corresponding $\Gamma(t) = \mathfrak{p}(t) + \mathfrak{r}(t)$ on $\mathbf{R}^N$. It then suffices to prove a similar result for

$$(T_\Gamma f)(x) = \text{p.v.} \int_{|t|\le 1} f(x - \Gamma(t))\, K(t)\, dt.$$

Now decompose $K(t) = \sum_{j=0}^{\infty} K_j(t)$, $K_j(t) = K(t)\psi(2^j t)$, where $\psi \in C_0^\infty(\mathbf{R}^k)$ is *radial* and vanishes near the origin. Then in accordance with §4.4 above, we consider the measures dm^j determined by

$$\int_{\mathbf{R}^N} f(x)\, dm^j(x) = \int_{\mathbf{R}^k} f(2^j \circ \Gamma(2^{-j} t))\, K(t)\, \psi(t)\, dt;$$

this implies that the $d\mu_j$ are given by

$$\int_{\mathbf{R}^n} f(x)\, d\mu_j(x) = \int_{\mathbf{R}^k} f(\Gamma(t))\, K_j(t)\, dt,$$

and that $T_\Gamma(f) = f * K$, with $K = \sum_j d\mu_j$.

One then observes that the $\{dm^j\}$ satisfy all the hypotheses required in §4.4.

This result was proved first for the case of curves γ in $\mathbf{R}^n$, in Stein and Wainger [1978]; see also the earlier work in Nagel, Rivière, and Wainger [1974]. A somewhat different approach is in Duoandikoetxea and Rubio de Francia [1986], and is based in part on the "bootstrap" argument occurring in §3.10 of the previous chapter. The case of curves already represents the main ideas, but is somewhat simpler because the lifting technique (i.e., the passage to $\mathbf{R}^N$) is not needed.

4.6 The idea of "descent" and the technique of addition of variables, which arose in §2.4, can also be used in modified form in several other contexts (such as in §4.7 and Chapter 13, §7.14 and §7.15). In the present context, where we deal with translation-invariant operators on Euclidean spaces, then in one of its forms (part (a) of the lemma in §2.4) it can essentially be re-interpreted in terms of *multipliers*.

Recall that a function m on $\mathbf{R}^n$ is said to be a (Fourier) multiplier on $L^p(\mathbf{R}^n)$ if the operator T, initially defined on $\mathcal{S}$ by $\widehat{Tf}(\xi) = m(\xi)\hat{f}(\xi)$, extends to a bounded operator from $L^p(\mathbf{R}^n)$ to itself. The (L^p) bound of the multiplier is then the norm of the corresponding operator. We can then assert the following.

Suppose m is a multiplier on $L^p(\mathbf{R}^N)$, and $L_1 : \mathbf{R}^n \to \mathbf{R}^N$ is a linear transformation. Then the function m_1, defined on $\mathbf{R}^n$ by $m_1(\xi) = m(L_1(\xi))$, is a multiplier on $L^p(\mathbf{R}^n)$ if m is continuous at each ξ in the range of L_1, in the sense that

$$m(\xi) = \lim_{r\to 0} |B(\xi, r)|^{-1} \int_{B(\xi,r)} m(\eta)\, d\eta;$$

here $B(\xi, r) \subset \mathbf{R}^N$ is the ball with center ξ and radius r. Moreover, the bound of m_1 does not exceed that of m.

The result goes back to de Leeuw [1956], in the case where $L_1 : \mathbf{R}^n \to \mathbf{R}^N$ is the usual inclusion map.† The general assertion made above can be obtained by similar methods, and in fact can be deduced from this special case. For this one proceeds by writing $L_1 = i\pi_2 \circ L_0 \circ \pi_1$, where $\pi_1 : \mathbf{R}^n \to V_1$ is a projection onto a complement V_1 of the null space of L_1, and $\pi_2 : \mathbf{R}^N \to V_2$ is a projection onto the image V_2 of L_1; $L_0 : V_1 \to V_2$ is invertible, and $i : V_2 \to \mathbf{R}^N$ is the inclusion.

Note that in the lemma in §2.4, we have $Tf = f * d\mu$; as a result, $m(\xi) = \widehat{d\mu}(\xi)$ is the corresponding multiplier on $\mathbf{R}^N$. Moreover, if

$$(T^L f)(x) = \int_{\mathbf{R}^N} f(x - L(z))\, d\mu(z),$$

then T^L corresponds to the multiplier $m(L_1(\xi))$ on $\mathbf{R}^n$, where L_1 is the transpose of the mapping $L : \mathbf{R}^N \to \mathbf{R}^n$. Note also in this case that m is continuous everywhere.

4.7 The maximal theorems in §2, and their singular integral analogues described in §4.5, have been extended to the variable coefficient (i.e., non-translation-invariant) case to give a theory which, in particular, is invariant under diffeomorphisms. The generalized singular integrals are referred to as *singular Radon transforms*, because they display features of both Radon transforms and singular integrals.

We consider a generalization $\gamma(x,t)$ of the basic function $x - \gamma(t)$ which occurs above. We require γ to be a C_0^∞ mapping of $(x,t) \in \mathbf{R}^n \times \mathbf{R}^k$ to $\mathbf{R}^n$, for x and t restricted to their respective unit balls. We also assume that $\gamma(x,0) \equiv x$.

It can be proved that such mappings can be asymptotically described by

$$\gamma(x,t) \sim \exp\Big(\sum_{|\alpha|>0} \frac{t^\alpha}{\alpha!} X_\alpha\Big)(x),$$

in the sense that there exist real, smooth vector fields $\{X_\alpha\}$ on $\mathbf{R}^n$, determined by γ, so that

$$\gamma(x,t) - \exp\Big(\sum_{0<|\alpha|<N} \frac{t^\alpha}{\alpha!} X_\alpha\Big)(x) = O(|t|^N)$$

as $t \to 0$, for every $N > 0$.

The analogue of the finite-type condition is the requirement that the vector fields X_α, together with their commutators, span $\mathbf{R}^n$, for each x with $|x| < 1$. Note that in the translation-invariant case (when $\gamma(x,t) = x - \gamma(t)$), the X_α are constant coefficient vector fields: $X_\alpha = -a_\alpha \cdot \nabla_x$, where $\sum_\alpha \frac{a_\alpha t^\alpha}{\alpha!}$ is the Taylor development of $\gamma(t)$ at $t = 0$; our assumption on the X_α is then equivalent to the finite-type assumption on $\gamma(t)$ at $t = 0$.

† See also the related results in *Fourier Analysis*, Chapter 7, §3.

To formulate the result, we take a smooth cut-off function η supported in the unit ball, and define

$$(\mathcal{M}f)(x) = \sup_{0<r\le 1} r^{-k} \left| \int_{|t|\le r} f(\gamma(x,t))\, \eta(x)\, dt \right|,$$

$$(Tf)(x) = \text{p.v.} \int_{|t|\le 1} f(\gamma(x,t))\, \eta(x)\, K(t)\, dt,$$

where K is a Calderón-Zygmund kernel, as in §4.5. The general theorem then asserts that $\mathcal{M}$ is bounded from $L^p(\mathbf{R}^n)$ to itself, for $1 < p < \infty$, and that T is bounded from $L^p(\mathbf{R}^n)$ to itself, for $1 < p < \infty$.

Besides the translation-invariant case for $\mathbf{R}^n$ dealt with in §2 and §4.5, the following represent further stages in the development of the above general theorem: The result in Nagel, Stein, and Wainger [1979] for $n = 2$, in which $\gamma(x,t)$ is a variable family of curves; the situation on the Heisenberg group (where $\gamma(x,t) = x \cdot \gamma(t)$ is determined by group multiplication) treated by Geller and Stein [1984]; its generalization in terms of "rotational curvature" in Phong and Stein [1983], [1986a]; finally, extensions to other nilpotent groups (besides the Heisenberg group) in D. Müller [1983], [1985], Christ [1985a], Ricci and Stein [1988]. The general result above is in Christ, Nagel, Stein, and Wainger [1993]. The proof given there is based in part on the technique of modeling on nilpotent groups, as in Chapter 13, §7.14 below.

C. Averages over hypersurfaces when $n \ge 3$

4.8 The notion of rotational curvature for hypersurfaces treated in §3.1, and its use for the averaging operators (69) on these hypersurfaces, has an interpretation in terms of the Fourier integral operators treated in Chapter 9.

(a) Suppose $\Phi(x,y)$ is given on a region of $\mathbf{R}^n \times \mathbf{R}^n$. Then it satisfies the rotational curvature condition (61) exactly when the phase function $\phi(x,y,\tau) = \tau\Phi(x,y)$, $\tau \in \mathbf{R}^1$, is nondegenerate in the sense of Chapter 9, §6.11. Thus the averaging operator

$$(Af)(x) = \int_{\mathbf{R}^n} \delta(\Phi)\, f(y)\, \psi(x,y)\, dy,$$

when expressed as the operator $\tilde{T}$, with the above phase and with symbol $b(x,y,\tau) = \psi(x,y)$, is equivalent to a Fourier integral operator of order $(1-n)/2$.

(b) As a consequence of Chapter 9, §6.17, if Φ_t is a scaled family that satisfies the rotational curvature condition (66), and $\eta \in C_0^\infty$ is a cut-off function vanishing near $t = 0$, then the operator

$$f \mapsto (A_t f)(x) = \eta(t) \int_{\mathbf{R}^n} \delta(\Phi_t)\, f(y)\, \psi_t\, dy$$

maps $L^2(\mathbf{R}^n) \to L^2_k(\mathbf{R}^{n+1})$, where $k = (n-1)/2$ and $(x,t) \in \mathbf{R}^{n+1}$; here ψ_t is a suitable family of cut-off functions. This fact, in effect, is behind the proof of the easier case of Theorem 2 (in §3.3), that is, the case when t is not small.

Additional discussion of the relation between rotational curvature and Fourier integral operators may be found in Phong and Stein [1986a], Sogge and Stein [1990]; this connection already occurs in Guillemin and Sternberg [1977].

4.9 Theorem 2 for maximal averages, when $p > n/(n-1)$ and $n \geq 3$, and its corollaries were proved under the hypothesis that an appropriate curvature is nowhere zero. Similar results hold, when $n \geq 3$, if we assume that the curvature in question does not vanish to infinite order at any point; then, however, a more restricted range of p is dictated, as indicated by the examples in §4.2.

(a) Consider first the translation-invariant case. Let $S \subset \mathbf{R}^n$ be a smooth, compact hypersurface and assume that its Gaussian curvature does not vanish to infinite order at any point of S. Then there is a $p_0 = p_0(S) < \infty$ so that

$$\|\mathcal{A}(f)\|_{L^p} \leq A_p \|f\|_{L^p}, \qquad \text{when } p_0 < p \leq \infty,$$

where $(\mathcal{A}f)(x) = \sup\limits_{t>0} |\int_S f(x - ty)\, d\sigma(y)|$.

(b) The "variable coefficient" version of this result can be formulated as follows. Suppose that the scaled family $\Phi_t(x, y)$ satisfies the following two conditions, instead of (66). First, for each t, $t > 0$, the quantity $\operatorname{rot\,curv}(\Phi_t)$ does not vanish to infinite order at any point of $S_{x,t}$. Second, a corresponding limiting condition, that the Gaussian curvature of S_x^0 does not vanish to infinite order at any point of S_x^0. Under these assumptions, we can assert that there is a $p_0 = p_0(\Phi_t) < \infty$, so that $\|\mathcal{A}(f)\|_{L^p} \leq A_p\|f\|_{L^p}$, for $p_0 < p \leq \infty$, when $n \geq 3$, if $\mathcal{A}(f) = \sup\limits_{0<t\leq 1} |A_t(f)|$, with $A_t(f)$ given by (69).

An illustrative special case of (a) arose first in the work of Cowling and Mauceri [1987]. The general conclusions above are in Sogge and Stein [1985] and [1990].

4.10 The averaging operator $(A_t f)(x) = \int_S f(x - ty)\, d\sigma(y)$ can be embedded in an analytic family (as in Chapter 8, §5.21), for which there are also versions of the maximal theorem. We illustrate this by considering the case when S is the unit sphere. This case has additional interest arising from its relation to the solution of the wave equation.

For $\mathrm{Re}(\alpha) > 0$, we define $(A_t^\alpha f)(x) = (f * m_t^\alpha)(x)$, with $m_t^\alpha(x) = t^{-n} m^\alpha(x/t)$,

$$m^\alpha(x) = \frac{(1 - |x|^2)^{\alpha-1}}{\Gamma(\alpha)} \quad \text{for } |x| < 1, \text{ and} \quad m^\alpha(x) = 0 \quad \text{for } |x| > 1.$$

For all complex α, and $f \in \mathcal{S}$, the function $A_t^\alpha(f)$ is defined by analytic continuation in α. Note that $A_t^0(f) = A_t(f)$. We set

$$(\mathcal{A}^\alpha f)(x) = \sup_{t>0} |(A_t^\alpha f)(x)|.$$

(a) We have the *a priori* inequality

$$\|\mathcal{A}^\alpha(f)\|_{L^p(\mathbf{R}^n)} \leq A_{p,\alpha}\|f\|_{L^p(\mathbf{R}^n)}, \qquad f \in \mathcal{S},$$

when $\alpha > 1 - n + \dfrac{n}{p}$ if $1 < p \leq 2$, and $\alpha > \dfrac{2-n}{p}$ if $2 \leq p \leq \infty$.

(b) Let $\alpha = (3 - n)/2$. For an appropriate constant c_n, we have that $c_n t(A_t^\alpha f)(x) = u(x, t)$, where u is the solution of the wave equation

$$\frac{\partial^2 u}{\partial t^2} = \Delta_x u, \qquad u(x, 0) \equiv 0, \qquad \frac{\partial}{\partial t} u(x, 0) = f(x).$$

As a consequence, one can assert that

$$\left\| \sup_{t>0} \left| \frac{u(x,t)}{t} \right| \right\|_{L^p(\mathbf{R}^n)} \leq A_p \|f\|_{L^p(\mathbf{R}^n)},$$

if $\dfrac{2n}{n+1} < p \leq \infty$ for $n \leq 3$, and $\dfrac{2n}{n+1} < p < \dfrac{2(n-2)}{n-3}$ for $n \geq 4$.

Also, $\lim\limits_{t\to 0} \dfrac{u(x,t)}{t} = f(x)$, a.e. x, if $f \in L^p(\mathbf{R}^n)$ and $\dfrac{2n}{n+1} < p$.

Stein [1976a]. "Variable coefficient" analogues may be found in Greenleaf [1981], Ruiz [1985], Sogge [1987c].

The proof of (a) is reduced, via interpolation, to the case $p = 2$ with $\alpha = 1 - (n/2)$; in this connection, see §4.1.

The inequalities in (a) and (b) are sharp for $1 < p \leq 2$, as examples similar to that in §1.1.3 show, but the optimal results of this kind, when $2 < p$ and $n \geq 2$, are still a mystery; see, however, Mockenhaupt, Seeger, and Sogge [1992b]. For what is known when $n = 2$, see §4.12 below.

D. Averages over hypersurfaces when $n = 2$

4.11 The results in §3, given by Theorem 2 and its corollaries, held under the assumption that $n \geq 3$ (and $p > n/(n-1)$). We turn to the situation when $n = 2$ (and $p > 2$).

(a) Consider first the translation-invariant case. Let $S \subset \mathbf{R}^2$ be a smooth curve whose curvature does not vanish anywhere, and let $d\sigma$ be the induced measure (modified by multiplication by a smooth cut-off function of compact support). Define

$$(\mathcal{A}f)(x) = \sup_{t>0} \left| \int_S f(x - ty)\, d\sigma(y) \right|.$$

Then $\|\mathcal{A}f\|_{L^p(\mathbf{R}^2)} \leq A_p\|f\|_{L^p(\mathbf{R}^2)}$, for $2 < p \leq \infty$.

(b) The general setting of Theorem 2 requires an additional hypothesis, as the example of varying line segments in §3.5.4 shows. This further assumption can be understood through the following considerations. The curvature hypothesis (66) imposes a condition on the family of hypersurfaces (curves) $S_{x,t}$ that is "static" in nature: the condition effectively bears on the functions $x \mapsto S_{x,t}$ for each $t > 0$, and does not take into account the change in $S_{x,t}$ as t varies. In the case of $\mathbf{R}^2$, what is required is a "dynamic" condition; this is the assumption of *cinematic* curvature, which we now formulate.

We know (by §4.8) that the averaging operators A_t can be expressed as Fourier integral operators; that is, when $0 < t \leq 1$, we have (essentially) that

$$(A_t f)(x) = \int_{\mathbf{R}^2} e^{i\Phi_t(x,\xi)}\, a_t(x,\xi)\, \widehat{f}(\xi)\, d\xi,$$

where the phases Φ_t and symbols a_t depend smoothly on t, and $a_t \in S^{-1/2}$ (since here $n = 2$).

To begin with, the curvature assumption (66) guarantees the usual non-degeneracy condition, namely that $\det\left(\dfrac{\partial^2 \Phi_t(x,\xi)}{\partial x_i \partial \xi_j}\right) \neq 0$, for each $t > 0$. Now the additional hypothesis we impose is strikingly parallel to an analogous condition occurring for the oscillatory integral treated in Chapter 9, §1.2 (given by (7) there). For each (x_0, t_0), one considers the mapping

$$\xi \mapsto \nabla_{x,t} \Phi_{t_0}(x_0, \xi) \in \mathbf{R}^3,$$

which is a conic submanifold of codimension one. Let u be a unit vector normal to this submanifold at (x_0, t_0, ξ_0), $\xi_0 \neq 0$; that is,

$$\nabla_\xi \langle \nabla_{x,t} \Phi_{t_0}(x_0,\xi), u\rangle \Big|_{\xi=\xi_0} = 0.$$

We then require that the 2×2 matrix

$$\left(\frac{\partial^2}{\partial \xi_i \partial \xi_j} \langle \nabla_{x,t} \Phi_{t_0}(x_0,\xi), u\rangle\right)\Big|_{\xi=\xi_0}$$

have maximal rank ($= 1$). This condition has an essentially equivalent geometric interpretation which is as follows: If two curves $S_{x,t}$ and $S_{x,t'}$ intersect tangentially at a point $\bar{x}$, then the difference of their curvatures at $\bar{x}$ is roughly $|x - x'|$.

Under the above hypothesis, one can assert that, with A_t as in (69) and with $(\mathcal{A}f)(x) = \sup_{0<t\leq 1} |(A_t f)(x)|$, one has

$$\|\mathcal{A}(f)\|_{L^p(\mathbf{R}^2)} \leq A_p \|f\|_{L^p(\mathbf{R}^2)}, \qquad \text{for } 2 < p \leq \infty.$$

The result in (a) is a theorem of Bourgain [1986a]. The adaptation of the idea of its proof to the variable-coefficient case and the development of the notion of cinematic curvature are due to Sogge [1991].

4.12 An alternative approach to the results in §4.11 has as its starting point the relation of the spherical maximal operator with the wave equation (alluded to in §4.10). This analysis has led to further insight regarding these questions and their relation to ideas we have already discussed, such as the estimates for Fourier integral operators (Chapter 9, §4) and Bochner-Riesz summability (Chapter 9, §6.7). The main results for $n = 2$ are as follows.

(a) If $\mathcal{A}^\alpha$ is as in §4.10, then we have the *a priori* inequality

$$\|\mathcal{A}^\alpha(f)\|_{L^p(\mathbf{R}^2)} \leq A\|f\|_{L^p(\mathbf{R}^2)}, \qquad \text{when } \alpha > -1/8 \text{ and } p = 4.$$

By interpolation, one gets a corresponding result for $\alpha > \alpha(p)$, with $\alpha(p) < 0$, when $2 < p < \infty$; the case $\alpha = 0$, $p = \infty$ is of course trivial.

(b) Let $u(x,t)$ be the solution of the wave equation

$$\frac{\partial^2 u}{\partial t^2} = \Delta_x u, \qquad u(x,0) \equiv 0, \qquad \frac{\partial u}{\partial t}(x,0) = f(x),$$

for $(x,t) \in \mathbf{R}^2 \times \mathbf{R}^1$, $f \in \mathcal{S}$. The one has the *a priori* inequality

$$\int_{[0,T]} \int_{\mathbf{R}^2} |u(x,t)|^p\, dx\, dt \leq A_p^p \|f\|^p_{L^p_\gamma(\mathbf{R}^2)}$$

if $p = 4$ and $\gamma > -\dfrac{3}{4} - \dfrac{1}{8}$, when $T < \infty$. This result is referred to as "local smoothing", since for fixed t, the best estimate when $p = 4$ is

$$\|u(x,t)\|_{L^p(\mathbf{R}^2,dx)} \leq A\|f\|_{L^p_\gamma(\mathbf{R}^2)}, \qquad \text{with } \gamma = -3/4;$$

thus there is a gain of essentially $1/8$ when integrating in t.

(c) The fact that distributions $u(x,t)$ satisfying the wave equation have their Fourier transforms concentrated on the cone $\tau = \pm|\xi|$ (here τ is the dual variable to t), combined with the analogy between results for the sphere and the cone (mentioned in Chapter 10, §3.8), leads us to define the following operator T^δ, which is analogous to the Bochner-Riesz summability operator S^δ of Chapter 9, §2. We consider the multiplier $m(\xi,\tau)$ given by

$$m(\xi,\tau) = \left(1 - \frac{|\xi|^2}{\tau^2}\right)^\delta \qquad \text{for } |\xi| \leq |\tau|,$$

with $m(\xi,\tau) = 0$ when $|\xi| \geq |\tau|$. Then if

$$\widehat{T^\delta f}(\xi,\tau) = m(\xi,\tau)\widehat{f}(\xi,\tau),$$

one can assert that

$$\|T^\delta(f)\|_{L^p(\mathbf{R}^3)} \leq A\|f\|_{L^p(\mathbf{R}^3)}, \qquad \text{for } 4/3 \leq p \leq 4 \text{ and } \delta > 1/8.$$

Of course, if the analogy of T^δ on $\mathbf{R}^3$ with S^δ on $\mathbf{R}^2$ were complete, this inequality would hold for all $\delta > 0$ (in the range $4/3 \leq p \leq 4$); but the truth of this more general assertion is unknown at present.

To stress further the relation of the above results with the approach in Chapter 9, §6.7, we describe two key points in the proof of (a), (b), and (c).

(i) The first is a maximal theorem (which plays a role analogous to (i) in Chapter 9, §6.7) It states that, for every $\varepsilon > 0$,

$$\int_{\mathbf{R}^2} \sup_\theta \left| |C_\theta^\delta|^{-1} \int_{C_\theta^\delta} f(x-y,t)\, dy\, dt \right|^2 dx \leq A_\varepsilon \delta^\varepsilon \|f\|^2_{L^2(\mathbf{R}^3)}$$

where C_θ^δ is a cylinder of unit length passing through the origin whose base has radius δ, and whose principal axis lies in the direction $(\cos\theta, \sin\theta, 1)$ along the surface of the cone $t = \pm|x|$.[‡]

[‡] See also the closely related two-dimensional analogue in Chapter 10, §3.11 and §3.12.

(ii) One also uses the following geometric observation. Let U_m^ν be the set of $(\xi,\tau) \in \mathbf{R}^2 \times \mathbf{R}^1$ so that

$$2^{j-1} \le |\xi| \le 2^{j+1}, \qquad ||\xi| - \tau| \le 1, \qquad |2^{-j/2}\tau - m| < c,$$

and ξ is restricted to lie in the thin sector Γ_j^ν that arose in the "second dyadic decomposition" for Fourier integral operators (Chapter 9, §4.4). Then for each fixed m and m', the number of times a point (ξ,τ) can belong to the algebraic sum $U_m^\nu + U_{m'}^{\nu'}$ is controlled by the inequality

$$\sum_{\nu,\nu'} \chi_{U_m^\nu + U_{m'}^{\nu'}} \le c.$$

Assertions (a) and (b) are in Mockenhaupt, Seeger, and Sogge [1992a] and [1992b]; the second paper deals with extensions to Fourier integral operators, subsuming the results in §4.11; see also Sogge [1993]. The inequality in (c) was obtained earlier in Mockenhaupt [1989]. The notion of "local smoothing" arose in the context of Schrödinger equations (see, e.g., Sjölin [1987], Vega [1988], Constantin and Saut [1988]); it is, however, of a different nature from that occurring for the wave equation.

A brief sketch of how (i) and (ii) can be used in the proofs of (a) and (b) follows immediately below.

4.13 We now outline the proof of (a) and (b) in §4.12. The essential point can be formulated as follows. Let F be the operator, mapping functions on $\mathbf{R}^2$ to functions on $\mathbf{R}^2 \times \mathbf{R}^1$, given by

$$(Ff)(x,t) = \eta(t) \int_{\mathbf{R}^2} e^{2\pi i|\xi|t} e^{2\pi i x\cdot\xi} a(\xi)\, \widehat{f}(\xi)\, d\xi,$$

where $\eta \in \mathcal{S}(\mathbf{R}^1)$ is a cut-off function whose Fourier transform is supported in $(-1,1)$, and $a \in S^0$ is a symbol supported in $2^{j-1} \le |\xi| \le 2^{j+1}$ (for some $j \ge 0$); we have $|\partial_\xi^\alpha a(\xi)| \le A_\alpha |\xi|^{-\alpha}$. Then one has the estimate

$$\|F(f)\|_{L^4(\mathbf{R}^3)} \le A_\mu 2^{\mu j} \|f\|_{L^4(\mathbf{R}^2)}, \qquad \text{if } \mu > 1/8. \tag{$*$}$$

To prove $(*)$, we decompose the spectrum of $F(f)$, discarding inessential contributions; note that in reducing to the problem $(*)$, we have already restricted to $|\xi| \approx 2^j$, which is suggested by the Littlewood-Paley decomposition. We now use also the second dyadic decomposition used in the study of Fourier integral operators (Chapter 9, §4.4). Thus, inserting the factors $\chi^\nu(\xi)$ (supported in the thin sectors Γ^ν) in the definition of F, we obtain operators F^ν with $F = \sum_\nu F^\nu$.

We require a further decomposition of $\sum_\nu F^\nu$ that arises by taking Fourier transforms in the t variable (whose dual variable is τ). We choose a cut-off function $\alpha \in C_0^\infty(\mathbf{R}^1)$ so that $\sum_{m\in\mathbf{Z}} \alpha(\tau - m) \equiv 1$, and define

$$(F_m^\nu f)(x,t) = \int_{\mathbf{R}^3} e^{2\pi i x\cdot\xi} e^{2\pi i t\cdot\tau} \widehat{\eta}(\tau - |\xi|)\, \alpha(2^{-j/2}\tau - m)\, \chi^\nu(\xi)\, a(\xi)\, \widehat{f}(\xi)\, d\xi\, d\tau.$$

First, one has that if $F_m = \sum_\nu F_m^\nu$, then

$$\Big\|\sum_m F_m\Big\|_{L^p(\mathbf{R}^3)} \le c 2^{(\frac14 - \frac{1}{2p})j} \Big\|\Big(\sum_m |F_m|^2\Big)^{1/2}\Big\|_{L^p(\mathbf{R}^3)}$$

for $2 \le p \le \infty$. Indeed, when $p = 2$ this follows by Plancherel's theorem, and the case $p = \infty$ holds because $F_m \not\equiv 0$ only when $|m| \le c2^{j/2}$. Thus when $p = 4$ this gives

$$\Big\|\sum_m F_m\Big\|_{L^4(\mathbf{R}^3)} \le c 2^{j/8} \Big\|\Big(\sum_m |F_m|^2\Big)^{1/2}\Big\|_{L^4(\mathbf{R}^3)}. \tag{$**$}$$

We now refer to §4.12(ii), and note that the spectrum of F_m^ν is contained in the set U_m^ν. Thus by the control given there on the number of points lying in the sum $U_m^\nu + U_{m'}^{\nu'}$, one sees that

$$\Big\|\Big(\sum_m |F_m|^2\Big)^{\frac12}\Big\|_{L^4(\mathbf{R}^3)} = \Big\|\Big(\sum_m \Big|\sum_\nu F_m^\nu\Big|^2\Big)^{\frac12}\Big\|_{L^4(\mathbf{R}^3)} \le c \Big\|\Big(\sum_{m,\nu} |F_m^\nu|^2\Big)^{\frac12}\Big\|_{L^4(\mathbf{R}^3)}.$$

Next, using the maximal inequality §4.12(i), one can show, by much the same method as used in Chapter 9 §6.7(i) and (ii), that

$$\Big\|\Big(\sum_{m,\nu} |F_m^\nu|^2\Big)^{1/2}\Big\|_{L^4(\mathbf{R}^3)} \le c 2^{\epsilon j} \Big\|\Big(\sum_{m,\nu} |f_m^\nu|^2\Big)^{1/2}\Big\|_{L^4(\mathbf{R}^2)}, \tag{$***$}$$

where $F_m^\nu(f) = F_m^\nu(f_m^\nu)$ and $\widehat{f_m^\nu}(\xi) = \tilde{\alpha}(2^{-j/2}|\xi| - m)\, \tilde{\chi}^\nu(\xi)\, \gamma(2^{-j}|\xi|)\, \widehat{f}(\xi)$, with $\gamma \in C_0^\infty(\mathbf{R}^1)$ vanishing near the origin. Finally

$$\Big\|\Big(\sum_{m,\nu} |f_m^\nu|^2\Big)^{1/2}\Big\|_{L^p(\mathbf{R}^2)} \le A\|f\|_{L^p(\mathbf{R}^2)}, \qquad 2 \le p \le \infty,$$

by arguments similar to the proof of Chapter 9 §6.7(iii). Combining this with $(**)$ and $(***)$ yields $(*)$.

For the complete proofs, the reader should consult the papers cited in §4.12 above.

E. Other topics

4.14 The estimates for the spherical maximal function actually imply that the standard maximal operator on $\mathbf{R}^n$,

$$(Mf)(x) = \sup_{r>0} \frac{1}{c_n r^n} \int_{|y|\le r} |f(x-y)|\, dy,$$

has L^p bounds, $p > 1$, that are independent of the dimension n. Indeed, one can show:

(a) For $p > 1$, there is a constant A_p, independent of n, so that

$$\|M(f)\|_{L^p(\mathbf{R}^n)} \leq A_p \|f\|_{L^p(\mathbf{R}^n)}, \qquad 1 < p \leq \infty.$$

(b) As far as the weak-type (1,1) inequality is concerned, bounds independent of the dimension are not known, but it is true that

$$|\{x \in \mathbf{R}^n : (Mf)(x) > \alpha\}| \leq \frac{A_1^{(n)}}{\alpha} \|f\|_{L^1(\mathbf{R}^n)}, \qquad \alpha > 0,$$

with $A_1^{(n)} = O(n)$ as $n \to \infty$.

The connection between (a) and the boundedness of the spherical maximal operator is indicated by the fact that, when the dimension n is large, a preponderance of the volume of the unit ball is concentrated near the boundary. In fact, (a) can be shown by using the L^p bounds for the spherical maximal operator on $\mathbf{R}^k$, where $k < n$ and $p > k/(k-1)$, together with a variant of the method of rotations used in §4.1 of Chapter 2. The proof of (b) is based on different ideas. See Stein [1982], Stein and Strömberg [1983].

The assertion in (b) should be compared with the bound 3^n that arises from the covering lemma arguments in §3 of Chapter 1; this argument only gives the bound $O(3^{n/p})$ for M on $L^p(\mathbf{R}^n)$, $p > 1$.

4.15 The above results raise the question as to whether other standard operators in harmonic analysis on $\mathbf{R}^n$ have bounds that are independent of the dimension n. This is indeed the case for the Riesz transforms, Poisson integrals, and Littlewood-Paley g functions.

(a) Let $R = (R_1, \ldots, R_n)$ be the Riesz transforms on $\mathbf{R}^n$ (see Chapter 1, §6.2.1 for a definition). Then there exist bounds A_p, A_p', independent of n, so that $A_p'\|f\|_{L^p(\mathbf{R}^n)} \leq \|R(f)\|_{L^p(\mathbf{R}^n)} \leq A_p\|f\|_{L^p(\mathbf{R}^n)}$, for $1 < p < \infty$.

(b) Suppose $u(x,t) = (f * P_t)(x)$ is the Poisson integral of f; let

$$g(f)(x) = \left(\int_0^\infty |\nabla u(x,t)|^2 \, t\, dt\right)^{1/2}$$

and

$$g_1(f)(x) = \left(\int_0^\infty \left|\frac{\partial u}{\partial t}(x,t)\right|^2 t\, dt\right)^{1/2}.$$

We then have $A_p'\|f\|_{L^p(\mathbf{R}^n)} \leq \|g_1(f)\|_{L^p(\mathbf{R}^n)} \leq \|g(f)\|_{L^p(\mathbf{R}^n)} \leq A_p\|f\|_{L^p(\mathbf{R}^n)}$, for $1 < p < \infty$, with A_p and A_p' independent of n.

(c) The mapping $f \mapsto \sup_{t>0} |u(x,t)|$ has a weak-type (1,1) bound that is independent of n.

Note that (b) implies (a) since $g_1(Rf)(x) \leq g(f)(x)$. The above results are implicit in the treatment of the square functions g and g_1 and maximal operators given in Chapter 4 of *Singular Integrals*, and Stein [1970a]. Other approaches that implicitly contain some of these results are in P. Meyer [1976], Gundy and Varopolous [1979]. See also Stein [1985], Duoandikoetxea and Rubio de Francia [1985].

4.16 In connection with §4.14, we may ask what happens when we replace the unit ball in $\mathbf{R}^n$ with an arbitrary open, bounded, convex, symmetric set B. Let

$$(M_B f)(x) = \sup_{r>0} |rB|^{-1} \int_{rB} |f(x-y)|\, dy.$$

Then

(a) There is a bound A_p, independent of B and n, so that

$$\|M_B(f)\|_{L^p(\mathbf{R}^n)} \leq A_p\|f\|_{L^p(\mathbf{R}^n)}, \qquad \text{if } 3/2 < p \leq \infty.$$

(b) While similar results are not known for $1 < p \leq 3/2$ (or for $p = 1$), one can assert that

$$|\{x \in \mathbf{R}^n : (Mf)(x) > \alpha\}| \leq \frac{A_n}{\alpha}\|f\|_{L^1(\mathbf{R}^n)}, \qquad \alpha > 0,$$

with $A_n = O(n \log n)$ as $n \to \infty$, and with A_n otherwise independent of B.

The occurrence of the exponent 3/2 may be explained as follows. One proves (a) by showing that when $|B| = 1$, there is an invertible linear transformation $L = L_B$ on $\mathbf{R}^n$ so that if $m(\xi) = \widehat{\chi_B}(L(\xi))$, then

$$|m(\xi)| \leq \frac{c}{|\xi|}, \qquad |m(\xi) - 1| \leq c|\xi|, \qquad \text{and} \quad |\langle \xi, \nabla_\xi m(\xi)\rangle| \leq c,$$

where c is a universal constant. The decay $O(|\xi|^{-1})$ is the same as that for the Fourier transform of surface measure on the unit sphere of $\mathbf{R}^3$, for whose maximal operator L^p estimates hold only when $p > 3/2$. See Bourgain [1986c], [1986d], Carbery [1986b]; further results for some special classes of convex sets are in Bourgain [1987], D. Müller [1990a]. The proof of the weak-type estimate in (b) depends on different ideas; see Stein and Strömberg [1983].

4.17 It turns out that, despite the example given in §4.3, there are positive results for the maximal operator (22) and its singular integral variant in §4.5 for the "flat" case, when $\gamma(t)$ is infinitely flat at the origin.† The most far-reaching results have been obtained when $k = 1$ and $n = 2$ (i.e., for curves in the plane); here it is natural to impose a convexity condition to exclude the example in §4.3. We assume that $\gamma(t) = (t, \gamma_2(t))$ where γ_2 is smooth, $\gamma_2(0) = \gamma_2'(0) = 0$, and $\gamma_2(t)$ is convex for $t \geq 0$. We set

$$(\mathcal{M}_\gamma f)(x) = \sup_{0<h<1} h^{-1} \int_0^h |f(x - \gamma(t))|\, dt,$$

and

$$(S_\gamma f)(x) = \text{p.v.} \int_{-1}^1 f(x - \gamma(t)) \frac{dt}{t}.$$

(a) If $\gamma_2(t)$ is odd, then $S_\gamma : L^2(\mathbf{R}^2) \to L^2(\mathbf{R}^2)$ is bounded exactly when there is a c > 0 so that $h(ct) \geq 2h(t)$, where $h(t) = t\gamma_2'(t) - \gamma_2(t)$. If this holds, $\mathcal{M}_\gamma$ is also bounded on $L^2(\mathbf{R}^2)$.

† Note, of course, there is always the "trivial" case when γ is linear, which easily reduces to the more standard theory.

(b) If γ_2 is even, then S_γ is bounded on $L^2(\mathbf{R}^2)$ if and only if there is a $c > 0$ so that $\gamma_2'(ct) \geq \gamma_2'(t)$ for $t > 0$. If this condition holds, then both S_γ and $\mathcal{M}_\gamma$ are bounded on $L^p(\mathbf{R}^2)$, $1 < p < \infty$.

(c) Suppose γ_2 is odd and there is an $\varepsilon > 0$ with $th'(t) > \varepsilon h(t)$, then both S_γ and $\mathcal{M}_\gamma$ are bounded on $L^p(\mathbf{R}^2)$, $1 < p < \infty$.

Nagel, Vance, Wainger, and Weinberg [1983] and [1985]; Carlsson, Christ, Cordoba, Duoandikoetxea, Rubio de Francia, Vance, Wainger, and Weinberg [1986]; Carbery, Christ, Vance, Wainger, and D. Watson [1989].

4.18 While weak-type (1,1) inequalities are not known to hold for the maximal operator $\mathcal{M}$ in §2 and its singular integral analogue in §4.5, there is a limiting result for p near 1 which involves the class $L \log L$. More precisely, let $\Phi(u) = u \log(u+2)$ for $u \geq 0$; then Φ is convex and increasing, and we denote by L^Φ the corresponding Orlicz space (as in Chapter 10, §3.6). With $\mathcal{M}$ as in Theorem 1, we have the following *a priori* inequality: There is an A so that

$$|\{x : (\mathcal{M}f)(x) > \alpha\}| \leq \int_{\mathbf{R}^n} \Phi\left(\frac{A|f|}{\alpha}\right) dx, \qquad \alpha > 0.$$

One proves first that $\mathcal{M} : L^\Phi \to L^{1,\infty}$ is bounded[‡] (see Christ and Stein [1987]); one then invokes Chapter 10, §3.6. There are similar results for the singular integrals in §4.5.

Notes

§1. The use of square functions to prove maximal theorems in this context originated in Stein [1976a], [1976b], and was developed in Stein and Wainger [1978]. For a historical account of the role of square functions, see Stein [1982]. A modification of some of the arguments is in Duoandikoetxea and Rubio de Francia [1986].

The use of square functions in this setting was preceded by a suggestive paper of Nagel, Rivière, and Wainger [1976], which proved the result in §1.2 by a decomposition in the Fourier transform space. However, it seems that that argument is restricted essentially to the case treated.

§2. When $k = 1$, the theorem was proved in Stein and Wainger [1978]. The additional element needed for $k \geq 1$ is the technique of lifting to $\mathbf{R}^N$, as in §2.1. For this, see Ricci and Stein [1988]; in a different context, it was already used in Rothschild and Stein [1976]. The related tool of "descent" goes back to de Leeuw [1965] in the setting of $\mathbf{R}^n$ and the Fourier transform, and has been developed in several other settings, in particular in the form of "transference". For more about this method, the reader is referred to Coifman and G. Weiss [1977b].

§3. Theorem 2 is in Sogge and Stein [1990]. Earlier relevant papers are Stein [1976a], Stein and Wainger [1978], Greenleaf [1981], Phong and Stein [1986a]. The case $n = 2$ in the translation-invariant context is due to Bourgain [1986a]; a generalization may be found in Sogge [1991].

‡ Here $L^{1,\infty}$ is the weak-type L^1 space, as in Chapter 6, §7.10.

CHAPTER XII

Introduction to the Heisenberg Group

In the final part of this book we take up the Heisenberg group $\mathbf{H}^n$. Its study is particularly appropriate at this stage because it affords us an opportunity to glance backward at ideas already treated and, at the same time, gives us a chance to look ahead at some further developments of interest.

It is a remarkable fact that the Heisenberg group arises in two fundamental but different settings in analysis. On the one hand, it can be realized as the boundary of the unit ball in several complex variables, and this fact has a decisive influence on our understanding of that subject (in particular, as it relates to the study of the Cauchy integral and the $\bar{\partial}$-complex). On the other hand, there is its genesis in the context of quantum mechanics, which emphasizes its "symplectic" role in connection with the Fourier transform, pseudo-differential operators, and related matters.

We can better appreciate the dual nature of the Heisenberg group if we bear in mind the following facts. Regarding complex analysis: the domain $\mathcal{U}^n$ and the singular integrals on its boundary $\mathbf{H}^n$ provide the simplest and most natural extension to several complex variables of the roles played in classical analysis by the upper half-plane and the Hilbert transform on its boundary $\mathbf{R}^1$. Regarding phase space and symplectic invariance: the pair of operators x and $(2\pi i)^{-1}(d/dx)$, which may be thought of as generating the pseudo-differential operators, have their interplay governed by the commutation relations inherent in the Heisenberg group.[†]

In this chapter we develop these ideas along the following lines.

(i) *The domain $\mathcal{U}^n$.* This domain is holomorphically equivalent to the unit ball in $\mathbf{C}^{n+1}$ and it stands in the same relation to that ball as does the upper half-plane to the unit disc in $\mathbf{C}^1$. The Heisenberg group arises as the group of translations of $\mathcal{U}^n$; this leads to its identification with the boundary $b\mathcal{U}^n$.

† Emblematic of the two-sided character of the Heisenberg group is the coincidence of notation occurring when parametrizing $\mathbf{H}^n$ as $\{(z,t) \in \mathbf{C}^n \times \mathbf{R}\}$: When we write $z = x + iy$, we are in the complex domain, while if we set $z = x + i\xi$, we have passed to phase space.

(ii) *Cauchy-Szegö integral.* Because of the above identification, and by the use of further symmetries of $\mathcal{U}^n$, the Cauchy-Szegö projection can be realized as a convolution operator on $\mathbf{H}^n$ with a simple and explicitly given singular kernel. The singular integral theory of Chapter 1 applies here and the setting of the Heisenberg group provides us with a natural example of the real-variable structures discussed in that chapter.

(iii) *Quantum mechanics.* The heuristic process of "quantization" requires us to fashion appropriate functions $a(Q, P)$ of operators Q_j and P_j, $1 \leq j \leq n$, which satisfy the Heisenberg commutation relations. When we choose the representations $Q_j = x_j$, $P_j = (2\pi i)^{-1}(\partial/\partial x_j)$, then the functions $a(Q, P)$ formed in a symplectically-invariant manner are variants of the pseudo-differential operators with symbols $a(x, \xi)$. This determines the *Weyl correspondence* $a \mapsto \mathrm{Op}(a)$.

(iv) *Other connections.* Two are particularly striking. First is the identification of the tangential Cauchy-Riemann operators on $\mathcal{U}^n$ with (complex) elements of the Lie algebra of $\mathbf{H}^n$. We shall see (in the next chapter) that the commutation relations of these elements and their conjugates provide a simple expression of the strict pseudoconvexity of the domain $\mathcal{U}^n$. Second is the notion of "twisted convolution", which reflects the composition of pseudo-differential operators (in the Weyl formalism), and is also useful in proving L^2 boundedness of singular integrals on $\mathbf{H}^n$.

1. Geometry of the complex ball and the Heisenberg group

1.1 In the complex plane, the unit disc $|w| < 1$ is seen to be holomorphically equivalent to the upper half-plane $\mathrm{Im}(z) > 0$ by the correspondence

$$w = \frac{i - z}{i + z}, \qquad z = i\left(\frac{1 - w}{1 + w}\right). \tag{1}$$

Now the real line $\mathbf{R}^1$ acts on the upper half-plane by translations

$$z \mapsto z + h, \qquad h \in \mathbf{R}^1,$$

and this action gives an identification of the boundary of the half-plane with the group $\mathbf{R}^1$. It is also responsible for the fact that such basic operators as the Cauchy integral (and its close relative, the Hilbert transform) are given by convolution. We now extend these ideas to higher dimensions.

1.2 In the complex space $\mathbf{C}^{n+1}$, we consider the unit ball

$$\{w \in \mathbf{C}^{n+1} : \sum_{j=1}^{n+1} |w_j|^2 < 1\}$$

and a corresponding upper half-space

$$\mathcal{U}^n = \{z \in \mathbf{C}^{n+1} : \mathrm{Im}\, z_{n+1} > \sum_{j=1}^{n} |z_j|^2\}. \tag{2}$$

These two domains are biholomorphically equivalent via mappings analogous to (1):

$$\begin{aligned} w_{n+1} &= \frac{i - z_{n+1}}{i + z_{n+1}}, & w_k &= \frac{2iz_k}{i + z_{n+1}}, & k &= 1, \ldots, n, \\ z_{n+1} &= i\left(\frac{1 - w_{n+1}}{1 + w_{n+1}}\right), & z_k &= \frac{w_k}{1 + w_{n+1}}, & k &= 1, \ldots, n. \end{aligned} \tag{3}$$

Indeed, the unit ball is given by $|w_{n+1}|^2 + \sum_1^n |w_k|^2 < 1$, and according to (3) this is equivalent to

$$|i - z_{n+1}|^2 + 4 \sum_{k=1}^{n} |z_k|^2 < |i + z_{n+1}|^2.$$

However $|i + z_{n+1}|^2 - |i - z_{n+1}|^2 = 4\, \mathrm{Im}(z_{n+1})$, and this shows that the complex ball $|w| < 1$ corresponds to the domain (2).

Observe that the equivalence (3) extends also to the boundaries of the above domains. Thus the boundary of $\mathcal{U}^n$, which is

$$b\mathcal{U}^n = \{z \in \mathbf{C}^{n+1} : \mathrm{Im}\, z_{n+1} = \sum_{j=1}^{n} |z_j|^2\},$$

corresponds via (3) to the boundary $|w| = 1$ of the unit ball, except for the "south pole" $(0, \ldots, 0, -1)$; this point may be viewed as the image of the "point at infinity" of $b\mathcal{U}^n$. Notice also that in this correspondence the north pole $(0, \ldots, 0, 1)$ is the image of the origin in $b\mathcal{U}^n$.

1.3 An important feature of the domain $\mathcal{U}^n$ is its invariance under a large group of "symmetries" (biholomorphic self-mappings). For our immediate purposes, we shall need to make explicit three subgroups of these symmetries: "dilations", "rotations", and "translations".

It will be convenient to use the notation

$$z = (z', z_{n+1}), \quad \text{where } z' = (z_1, \ldots, z_n) \in \mathbf{C}^n \text{ and } z_{n+1} \in \mathbf{C}^1.$$

Now for each positive number δ, we define the *dilation* $\delta \circ z$ by

$$\delta \circ z = \delta \circ (z', z_{n+1}) = (\delta z', \delta^2 z_{n+1}). \tag{4}$$

Notice that the nonisotropy, represented by the fact that z' is dilated by δ and z_{n+1} by δ^2, is consistent with the definition (2) of $\mathcal{U}^n$. It is obvious that the dilations map $\mathcal{U}^n$ to $\mathcal{U}^n$ and $b\mathcal{U}^n$ to $b\mathcal{U}^n$.

For each unitary linear transformation u on $\mathbf{C}^n$, we define $u(z)$ by

$$u(z) = u(z', z_{n+1}) = (u(z'), z_{n+1}). \tag{5}$$

Again it is clear that these *rotations* give holomorphic self-mappings of $\mathcal{U}^n$ and extend to mappings of $b\mathcal{U}^n$ to itself.

1.4 Heisenberg group. We come now to the Heisenberg group, which gives the *translations* of the domain $\mathcal{U}^n$. Abstractly, this group consists of the set‡

$$\mathbf{C}^n \times \mathbf{R} = \{[\zeta, t] : \zeta \in \mathbf{C}^n, t \in \mathbf{R}\}$$

with the multiplication law

$$[\zeta, t] \cdot [\eta, s] = [\zeta + \eta, t + s + 2\,\mathrm{Im}(\zeta \cdot \bar{\eta})]. \tag{6}$$

It is easy to check that the law (6) does indeed make $\mathbf{C}^n \times \mathbf{R}$ into a group whose identity is the origin $[0, 0]$ and where the inverse is given by $[\zeta, t]^{-1} = [-\zeta, -t]$. The space $\mathbf{C}^n \times \mathbf{R}$ with the multiplication structure (6) is the Heisenberg group and will be denoted by $\mathbf{H}^n$.

To each element $[\zeta, t]$ of $\mathbf{H}^n$, we associate the following (holomorphic) affine self-mapping of $\mathcal{U}^n$:

$$[\zeta, t] : (z', z_{n+1}) \mapsto (z' + \zeta, z_{n+1} + t + 2iz' \cdot \bar{\zeta} + i|\zeta|^2). \tag{7}$$

In fact, since $|z'+\zeta|^2 - |z'|^2 = \mathrm{Im}\{i(z' \cdot \bar{\zeta} + |\zeta|^2)\}$, the mapping preserves the defining function r, which is given by

$$r(z) = \mathrm{Im}(z_{n+1}) - |z'|^2. \tag{8}$$

Hence the transformation (7) maps $\mathcal{U}^n = \{z : r(z) > 0\}$ to itself and preserves the boundary $b\mathcal{U}^n = \{z : r(z) = 0\}$.

Observe next that the mapping (7) defines an *action* of the group $\mathbf{H}^n$ on the space $\mathcal{U}^n$: If we compose the mappings (7) corresponding to elements $[\zeta, t]$, $[\eta, s]$ of $\mathbf{H}^n$, the resulting transformation corresponds to the element $[\zeta, t] \cdot [\eta, s]$. This follows easily from the identity

$$2i\zeta \cdot \bar{\eta} + i|\zeta|^2 + i|\eta|^2 = 2\,\mathrm{Im}(\zeta \cdot \bar{\eta}) + i|\zeta + \eta|^2.$$

Thus (7) gives us a realization of $\mathbf{H}^n$ as a group of affine holomorphic bijections of $\mathcal{U}^n$ (as we observe below, this realization is "faithful"— different elements of $\mathbf{H}^n$ give different mappings).

‡ Here we shall use square brackets [ ] for elements of the Heisenberg group to distinguish them from points in $\mathbf{C}^{n+1}$, for which parentheses ( ) are used.

1.4.1 It is important to note that the mappings (7) are simply transitive on the boundary $b\mathcal{U}^n$: For every two points in $b\mathcal{U}^n$, there is exactly one element of $\mathbf{H}^n$ mapping the first to the second. In particular, we have that

$$[\zeta, t] : (0, 0) \mapsto (\zeta, t + i|\zeta|^2);$$

so we can identify the Heisenberg group with $b\mathcal{U}^n$ via its action on the origin:

$$\mathbf{H}^n \ni [\zeta, t] \mapsto (\zeta, t + i|\zeta|^2) \in b\mathcal{U}^n. \tag{9}$$

Since the Heisenberg group $\mathbf{H}^n$ preserves the boundary $b\mathcal{U}^n$, and since we have identified $b\mathcal{U}^n$ with $\mathbf{H}^n$, we get as a result an action of $\mathbf{H}^n$ on itself. Because the general element of $b\mathcal{U}^n$ is of the form $h(0)$ for some $h \in \mathbf{H}^n$, the action of of another $h_1 \in \mathbf{H}^n$ maps it to $(h_1 h)(0)$, and so the action of $\mathbf{H}^n$ on itself is simply by *left* translation:

$$h_1 : h \mapsto h_1 h.$$

1.4.2 The above considerations have an important heuristic consequence. Since $\mathbf{H}^n$ acts on $\mathcal{U}^n$ by holomorphic automorphisms, it can be expected that the basic operators of complex analysis on $\mathcal{U}^n$ should be invariant under the action of $\mathbf{H}^n$. Hence we can expect them to be realized in terms of (left-invariant) convolution operators on $\mathbf{H}^n$.† This is indeed the case, as we shall see below.

1.4.3 There is a natural measure dh on $\mathbf{H}^n$ which is given by the Euclidean Lebesgue measure $d\zeta\, dt$ on $\mathbf{C}^n \times \mathbf{R}$; here we write $h = [\zeta, t]$, $\zeta \in \mathbf{C}^n$, $t \in \mathbf{R}$. In view of (6), the left translations of $\mathbf{H}^n$ on itself are affine when considered as mappings of $\mathbf{C}^n \times \mathbf{R}$, and their linear parts have determinant 1. Thus the measure dh is left-invariant (by the same token, it is also right-invariant); it is the *Haar measure* for $\mathbf{H}^n$.

1.4.4 For later purposes it is useful to consider the "Heisenberg coordinates" on $\mathcal{U}^n$, which are implicit in our considerations above. They are given by

$$[\zeta, t, r] = [z', \mathrm{Re}\, z_{n+1}, \mathrm{Im}(z_{n+1}) - |z'|^2].$$

Thus r represents the height of the point $z = (z', z_{n+1}) \in \mathcal{U}^n$ and $[\zeta, t]$ represents its projection onto $b\mathcal{U}^n$, identified with $\mathbf{H}^n$. Note, however, that the correspondence $z \mapsto [\zeta, t, r]$ is *not* holomorphic.

† The dilation and rotation groups will also play a role in the determination of these operators.

2. The Cauchy-Szegö integral

The Cauchy-Szegö integral may be viewed as the orthogonal projection of $L^2(\mathbf{H}^n)$ onto its subspace of boundary values of holomorphic functions. To make this idea precise, we first discuss the analogue (for $\mathbf{H}^n$) of the classical Hardy space $\mathcal{H}^2$ of holomorphic functions in the upper half-plane.[‡]

2.1 The space $\mathcal{H}^2(\mathcal{U}^n)$. We continue our identification of $b\mathcal{U}^n$ (the boundary of $\mathcal{U}^n$) with $\mathbf{H}^n$, as given by (9). This identification allows us to transport the Haar measure dh on $\mathbf{H}^n$ to a measure $d\beta$ on $b\mathcal{U}^n$; that is, we have the integration formula

$$\int_{b\mathcal{U}^n} F(z)\, d\beta(z) = \int_{\mathbf{C}^n\times\mathbf{R}} F(z', t+i|z'|^2)\, dz'\, dt,$$

for (say) continuous F of compact support. With these measures we can define the space $L^2(\mathbf{H}^n) = L^2(b\mathcal{U}^n)$.

For any function F defined on $\mathcal{U}^n$, we write F_ε for its "vertical translate":

$$F_\varepsilon(z) = F(z+\varepsilon\mathbf{i}), \qquad \text{where } \mathbf{i} = (0,\dots,0,i).$$

If $\varepsilon > 0$, then F_ε is defined in a neighborhood of $\overline{\mathcal{U}}^n$; in particular, F_ε is defined on $b\mathcal{U}^n$.

The space $\mathcal{H}^2(\mathcal{U}^n)$ consists of all functions F holomorphic on $\mathcal{U}^n$, for which

$$\sup_{\varepsilon>0} \int_{b\mathcal{U}^n} |F_\varepsilon(z)|^2\, d\beta(z) < \infty. \tag{10}$$

The norm $\|F\|_{\mathcal{H}^2}$ of F is then the square root of the left side of (10). That the space $\mathcal{H}^2(\mathcal{U}^n)$ is actually a Hilbert space, and that its elements can be identified with their boundary values, is the content of the following proposition.

PROPOSITION. *Suppose F belongs to $\mathcal{H}^2(\mathcal{U}^n)$. Then*

(i) *There exists an $F^b \in L^2(b\mathcal{U}^n)$ so that $F(z+\varepsilon\mathbf{i})|_{b\mathcal{U}^n} \to F^b$ in the $L^2(b\mathcal{U}^n)$ norm, as $\varepsilon \to 0$.*

(ii) *The space of F^b so obtained is a closed subspace of $L^2(b\mathcal{U}^n)$; moreover,*

(iii) $\|F^b\|_{L^2(b\mathcal{U}^n)} = \|F\|_{\mathcal{H}^2}$.

[‡] For background and other generalizations see, e.g., *Fourier Analysis*, Chapter 3.

2.2 The proof of the proposition can be reduced to a version of itself for the classical case of one variable; in effect, we consider $F(z) = f(z', z_{n+1})$ as a function of z_{n+1}, for different fixed values of z'. What is needed to carry this out is the classical theorem concerning the (holomorphic) H^2 space on the upper half plane, which can be formulated as follows. We define $\mathcal{H}^2(\mathbf{R}^2_+)$ to be the set of functions $f(w)$, holomorphic for $w = u+iv$ in the upper half-plane $v > 0$, which satisfy

$$\left(\sup_{v>0}\int_{\mathbf{R}} |f(u+iv)|^2\, du\right)^{1/2} = \|f\|_{\mathcal{H}^2(\mathbf{R}^2_+)} < \infty. \tag{11}$$

As a consequence of (11), one has the maximal inequality

$$\int_{\mathbf{R}} \sup_{v>0} |f(u+iv)|^2\, du \le A^2 \|f\|^2_{\mathcal{H}^2(\mathbf{R}^2_+)}. \tag{12}$$

This implies the almost-everywhere result, namely that there exists an $f^b \in L^2(\mathbf{R})$ so that

$$f(u+iv) \to f^b(u) \quad \text{as } v \to 0, \qquad \text{for a.e. } u,$$

as well as the identity

$$\|f^b\|_{L^2(\mathbf{R})} = \|f\|_{\mathcal{H}^2(\mathbf{R}^2_+)}. \tag{13}$$

Indeed, the Hardy space theory of Chapter 3, §4.2 shows that (11) implies (12). For the almost-everywhere result see Chapter 3, §5.12.[†]

2.2.1 To apply the one-dimensional theory, we shall consider, for each $z' \in \mathbf{C}^n$ and $\delta > 0$, the function f defined on the upper half-plane by $f(z_{n+1}) = F(z', z_{n+1} + i[\delta + |z'|^2])$. The first step is to establish that $f \in \mathcal{H}^2(\mathbf{R}^2_+)$ for *every* z' and $\delta > 0$. The second is to show that $f \in \mathcal{H}^2(\mathbf{R}^2_+)$ for almost every z', when $\delta = 0$. The latter conclusion will allow us to prove the proposition.

LEMMA. *Suppose $F \in \mathcal{H}^2(\mathcal{U}^n)$ and $\delta > 0$. Then for each $z' \in \mathbf{C}^n$, the function*

$$f(z_{n+1}) = F(z', z_{n+1} + i[\delta + |z'|^2])$$

with $z_{n+1} = u+iv$, belongs to $\mathcal{H}^2(\mathbf{R}^2_+)$.

[†] These results may also be found in *Fourier Analysis*, Chapter 2, §3, and Chapter 3, §5.

We first simplify the statement by using the Heisenberg group. Indeed, if $h = [\zeta, t] \in \mathbf{H}^n$ with $h(z)$ given by (7), it is easy to check that $h(z+\varepsilon\mathbf{i}) = h(z) + \varepsilon\mathbf{i}$; thus by the invariance of the measure $d\beta$ on $b\mathcal{U}^n$, we see that whenever $F \in \mathcal{H}^2(\mathcal{U}^n)$, then also $F_h \in \mathcal{H}^2(\mathcal{U}^n)$ with $F_h(z) = F(h(z))$. Choosing $h = [z', 0]$ allows us to reduce the proof of the lemma to the special case $z' = 0$.

Next, by using the dilations (4), it is easy to further simplify the statement by reducing matters to the case where $\delta = 1$. Thus we can take $z' = 0$, $\delta = 1$, and set

$$f(u+iv) = F(0, u+iv+i), \qquad v > 0.$$

Now by the mean-value property of holomorphic functions,

$$|f(u+iv)|^2 \le c \int_{|z|<1/2} |F(z', z_{n+1} + u + iv + i)|^2 \, dz. \tag{14}$$

Here $dz = dz'\, dz_{n+1}$ denotes the Lebesgue measure on $\mathbf{C}^{n+1}$, and the constant c is the reciprocal of the volume of the ball of radius $1/2$. Notice that, in the integral above, the imaginary part of $z_{n+1} + u + iv + i$ is at least $1/2$, and therefore exceeds $|z'|^2$, guaranteeing that the integration in (14) is taken over a subset of $\mathcal{U}^n$.

We now integrate (14) with respect to u, write $z_{n+1} = x_{n+1} + iy_{n+1}$, and assimilate the integration in u with that in x_{n+1}. The result is

$$\begin{aligned}&\int_{\mathbf{R}} |f(u+iv)|^2 \, du \\ &\le c' \int\limits_{\substack{|z'|<1/2 \\ |y_{n+1}|<1/2}} |F(z', x_{n+1} + i[y_{n+1} + v + 1])|^2 \, dz' \, dx_{n+1} \, dy_{n+1}.\end{aligned} \tag{15}$$

At this stage we make the change of variables $y_{n+1} \to \varepsilon$, where $y_{n+1} + v + 1 = \varepsilon + |z'|^2$, so $\varepsilon = y_{n+1} + v + 1 - |z'|^2$. Since $|y_{n+1}| < 1/2$ and $|z'|^2 < 1/4$, the range of ε is then contained in the interval $(1/4 + v, 3/2 + v)$. Moreover, if we use the definition of the measure $d\beta$ on $b\mathcal{U}^n$ described in §2.1, we see that the integral on the right side of (15) is majorized by

$$c' \cdot \int_{1/4+v}^{3/2+v} \left(\int_{b\mathcal{U}^n} |F(z+\varepsilon\mathbf{i})|^2 \, d\beta(z) \right) d\varepsilon,$$

which, because of (10), is dominated by

$$c' \cdot \frac{5}{4} \cdot \|f\|_{\mathcal{H}^2(\mathcal{U}^n)}.$$

This proves that $f \in \mathcal{H}^2(\mathbf{R}^2_+)$, and the lemma is established.

2.2.2 We now apply the maximal inequality (12) to

$$f(z_{n+1}) = F(z', z_{n+1} + i[\delta + |z'|^2]),$$

for each z'; together with (13), it shows that

$$\begin{aligned}\int_{\mathbf{R}} \sup_{y_{n+1}>0} &|F(z', x_{n+1} + i[y_{n+1} + \delta + |z'|^2])|^2 \, dx_{n+1} \\ &\le A^2 \int_{\mathbf{R}} |F(z', x_{n+1} + i[\delta + |z'|^2])|^2 \, dx_{n+1}.\end{aligned}$$

We integrate this with respect to z', and what is obtained can be reinterpreted to state

$$\int_{b\mathcal{U}^n} \sup_{\varepsilon>0} |F(z + \mathbf{i}[\delta + \varepsilon])|^2 \, d\beta(z) \le A^2 \int_{b\mathcal{U}^n} |F(z + \mathbf{i}\delta)|^2 \, d\beta(z).$$

If we let $\delta \to 0$ in the above, we have

$$\int_{b\mathcal{U}^n} \sup_{\varepsilon>0} |F(z + \varepsilon\mathbf{i})|^2 \, d\beta(z) \le A^2 \|F\|^2_{\mathcal{H}^2(\mathcal{U}^n)}. \tag{16}$$

From (16), we see that for a.e. $z' \in \mathbf{C}^n$, the function

$$z_{n+1} \mapsto F(z', z_{n+1} + i|z'|^2)$$

is in $\mathcal{H}^2(\mathbf{R}^2_+)$ and by §2.2 we get that

$$\lim_{\varepsilon\to 0} F(z + \varepsilon\mathbf{i}) = F^b(z)$$

exists for almost every $z \in b\mathcal{U}^n$.

2.2.3 By Fatou's lemma, it follows that

$$\int_{b\mathcal{U}^n} |F^b(z)|^2 \, d\beta(z) \le \sup_{\varepsilon>0} \int_{b\mathcal{U}^n} |F(z + \varepsilon\mathbf{i})|^2 \, d\beta(z) = \|F\|^2_{\mathcal{H}^2(\mathcal{U}^n)}. \tag{17}$$

On the other hand, because of (13),

$$\int_{\mathbf{R}} |F^b(z', x_{n+1} + i|z'|^2)|^2 \, dx_{n+1} \ge \sup_{\varepsilon>0} \int_{\mathbf{R}} |F(z', x_{n+1} + i[\varepsilon + |z'|^2])|^2 \, dx_{n+1}$$

for almost every z'. Thus an integration in z' proves the reverse inequality to (17), and hence the identification of $L^2(\mathbf{H}^n)$ with $L^2(b\mathcal{U}^n)$ gives us conclusion (iii) of the proposition.

Finally, the inequality (14) actually shows that whenever K is a compact subset of $\mathcal{U}^n$, then there is a bound c_K so that

$$\sup_{z\in K} |F(z)| \le c_K \|F\|_{\mathcal{H}^2(\mathcal{U}^n)}, \tag{18}$$

from which it is clear that convergence of a sequence in $\mathcal{H}^2(\mathcal{U}^n)$ norm implies its uniform convergence over compact subsets of $\mathcal{U}^n$. From this it follows that the space $\mathcal{H}^2(\mathcal{U}^n)$ is complete in its norm, and this establishes conclusion (ii) of the proposition. Since we have already established the existence of the limit asserted in (i), the proof of the proposition is complete.

2.3 The Cauchy-Szegö kernel. We will now determine the Cauchy-Szegö kernel $S(z,w)$ for the domain $\mathcal{U}^n$. It is the function, defined on $\mathcal{U}^n \times \mathcal{U}^n$, characterized by the following properties:

(a) For each $w \in \mathcal{U}^n$, the function $z \mapsto S(z,w)$ is holomorphic for $z \in \mathcal{U}^n$, and in fact belongs to $\mathcal{H}^2(\mathcal{U}^n)$. This allows us to define, for each $w \in \mathcal{U}^n$, the boundary function $S^b(z,w)$ (which is defined for almost all $z \in b\mathcal{U}^n$).

(b) The kernel S is symmetric, in the sense that $S(z,w) = \overline{S(w,z)}$ for each $(z,w) \in \mathcal{U}^n \times \mathcal{U}^n$. This symmetry of course permits us to extend the definition of S so that for each $z \in \mathcal{U}^n$, $S(z,w)$ is defined for almost every $w \in b\mathcal{U}^n$.

(c) The kernel S satisfies the reproducing property

$$F(z) = \int_{b\mathcal{U}^n} S(z,w)\, F^b(w)\, d\beta(w), \quad z \in \mathcal{U}^n, \tag{19}$$

whenever $F \in \mathcal{H}^2(\mathcal{U}^n)$.

THEOREM 1. *Let* $S(z,w) = c_n[r(z,w)]^{-n-1}$, *where*

$$r(z,w) = \frac{i}{2}(\bar{w}_{n+1} - z_{n+1}) - \sum_{k=1}^{n} z_k \bar{w}_k, \tag{20}$$

and

$$c_n = \frac{n!}{4\pi^{n+1}}. \tag{21}$$

Then S is the (unique) function that enjoys the properties (a), (b), and (c) *above.*

It is worth pointing out the following:

(i) The function $r(z,w)$ is holomorphic in z, antiholomorphic in w, and, when restricted to the diagonal $z = w$, agrees with the defining function (8). This property uniquely characterizes $r(z,w)$.

(ii) Note that, for each fixed $w \in \mathcal{U}^n$, $r(z,w)$ (and hence $S(z,w)$) is holomorphic for z in a neighborhood of $\overline{\mathcal{U}}^n$, with a parallel assertion when the roles of z and w are interchanged. In particular, if $w \in \mathcal{U}^n$ is fixed, the boundary function $S^b(\cdot,w)$ is actually defined on all of $b\mathcal{U}^n$.

2.3.1 We now turn to the proof of the theorem.

Let C denote the *Cauchy-Szegö projection operator*: it is the orthogonal projection from $L^2(b\mathcal{U}^n)$ to the subspace of functions $\{F^b\}$ that are boundary values of functions $F \in \mathcal{H}^2(\mathcal{U}^n)$. In other words, for each $f \in L^2(b\mathcal{U}^n)$, we have that $C(f) = F^b$ for some $F \in \mathcal{H}^2(\mathcal{U}^n)$; moreover, $C(F^b) = F^b$ and C is self-adjoint, i.e., $C^* = C$.

Fix now $z \in \mathcal{U}^n$; then $(Cf)(z) = F(z)$ is well-defined,‡ and the kernel $S(z,w)$ will be determined by the representation

$$F(z) = \int_{b\mathcal{U}^n} S(z,w)\, f(w)\, d\beta(w). \tag{22}$$

Note that this is a more general version than (19).

Heuristically, we can proceed as follows. Let $\{\phi_j\}$ be an orthonormal basis of the Hilbert space $\mathcal{H}^2(\mathcal{U}^n)$. By the identification of $\mathcal{H}^2(\mathcal{U}^n)$ with a subspace of $L^2(b\mathcal{U}^n)$, we may also take $\{\phi_j\}$ as a basis of this subspace. We can then expect that

$$S(z,w) = \sum_j \phi_j(z)\bar{\phi}_j(w). \tag{23}$$

Indeed, whenever $\sum_j |a_j|^2 < \infty$, then $\sum_j a_j\phi_j \in \mathcal{H}^2(\mathcal{U}^n)$ and $\|\sum_j a_j\phi_j\|^2_{\mathcal{H}^2(\mathcal{U}^n)} = \sum_j |a_j|^2$. Moreover, for any compact set $K \subset \mathcal{U}^n$, we have by (18) that

$$\sup_{z\in K}\Big|\sum_j a_j\phi_j(z)\Big| \le c_K\Big(\sum_j |a_j|^2\Big)^{1/2}.$$

Therefore, by the converse of Schwarz's inequality,

$$\Big(\sum_j |\phi_j(z)|^2\Big)^{1/2} \le c_K, \qquad \text{for all } z \in K,$$

and thus the sum (23) converges uniformly whenever (z,w) belongs to a compact subset of $\mathcal{U}^n \times \mathcal{U}^n$.

We take (23) as the definition of $S(z,w)$. Note that $\bar{S}(z,w) = S(w,z)$, and that for each fixed $z \in \mathcal{U}^n$, $\bar{S}(z,w) \in \mathcal{H}^2(\mathcal{U}^n)$ as a function of w; moreover, $\bar{S}(z,w)$ extends to $z \in \mathcal{U}^n$, $w \in b\mathcal{U}^n$, by the identity $S(z,w) = \sum \phi_j(z)\bar{\phi}_j(w)$, with the series converging in the norm of $L^2(b\mathcal{U}^n)$.

These considerations therefore establish the existence of a function S that satisfies (22) and the properties (a), (b), and (c) above. The reproducing property (19) uniquely determines $\bar{S}(z,w)$ as an element of $\mathcal{H}^2(\mathcal{U}^n)$, for each fixed z. Together with conclusion (iii) of the proposition in §2.1, this shows that S is uniquely determined by the properties (a), (b), and (c).

‡ We use the symbol $C(f)$ to denote either F or its boundary values F^b, depending on context.

2.3.2 We now exploit the symmetry properties of $\mathcal{U}^n$ to establish an explicit formula for S.

Using first the dilations given by (4) in §1.3, we assert that

$$S(\delta \circ z, \delta \circ w) = \delta^{-2n-2} S(z,w), \qquad \delta > 0. \tag{24}$$

Indeed, looking back at the identification of $b\mathcal{U}^n$ with $\mathbf{H}^n$, and the corresponding identification of $d\beta(z)$ with Euclidean measure on $\mathbf{H}^n$ (see §2.1), we observe that $d\beta(\delta \circ z) = \delta^{2n+2}\, d\beta(z)$. Since the mapping $F(z) \mapsto F(\delta \circ z)$ also takes $\mathcal{H}^2(\mathcal{U}^n)$ to itself, we see that

$$F(\delta^{-1} \circ z) = \int_{b\mathcal{U}^n} S(z,w)\, F^b(\delta^{-1} \circ w)\, d\beta(w).$$

Making the indicated change of variables $z \mapsto \delta \circ z$, $w \mapsto \delta \circ w$ then shows that

$$F(z) = \int_{b\mathcal{U}^n} S(\delta z, \delta w)\, \delta^{2n+2}\, F^b(w)\, d\beta(w).$$

Thus $S(\delta \circ z, \delta \circ w)\delta^{2n+2}$ satisfies the properties (a), (b), and (c), and therefore must equal $S(z,w)$. This proves (24).

Similarly, using the unitary rotations given by (5), we see that

$$S(u(z), u(w)) = S(z,w), \tag{25}$$

because, as is easily verified, $d\beta(u(z)) = d\beta(z)$.

In the same way we can also exploit the action of the Heisenberg group on $\mathcal{U}^n$, given by (7), to obtain

$$S(h(z), h(w)) = S(z,w), \qquad h \in \mathbf{H}^n. \tag{26}$$

The identities (24), (25), and (26) hold for each $(z,w) \in \mathcal{U}^n \times \mathcal{U}^n$. They are also valid for $(z,w) \in \mathcal{U}^n \times b\mathcal{U}^n$ in that, for each $z \in \mathcal{U}^n$, they are identities for elements of $L^2(b\mathcal{U}^n)$ and therefore hold for almost every $w \in b\mathcal{U}^n$. However, because of the translation identity (26), and the fact that $\mathbf{H}^n$ acts transitively on $b\mathcal{U}^n$, the assertion (26) can be made more precise: For each $z \in \mathcal{U}^n$, $S(z,w)$ is actually a smooth function of $w \in b\mathcal{U}^n$, and (26) holds for all $(z,w) \in \mathcal{U}^n \times b\mathcal{U}^n$. As a result, (24) and (25) also hold for all $(z,w) \in \mathcal{U}^n \times b\mathcal{U}^n$. We postpone (until §2.3.4 below) the technical proof that, for each fixed $z \in \mathcal{U}^n$, $S(z,w)$ is a smooth function of $w \in b\mathcal{U}^n$.

Let $S(z) = S(z,0)$. Then $S(z)$ is holomorphic on $\mathcal{U}^n$, and is invariant under the unitary rotations $z = (z', z_{n+1}) \mapsto (u(z'), z_{n+1})$. As a consequence, $S(z)$ is constant in z' and thus depends on z_{n+1} only, so we may write $S(z) = s(z_{n+1})$. Using (24) shows that $s(\delta^2 z_{n+1}) = \delta^{-2n-2} s(z_{n+1})$ for all $\delta > 0$, which implies that $S(z) = c z_{n+1}^{-n-1}$.

Finally, we determine the form of $S(z,w)$ for $z \in \mathcal{U}^n$, $w \in b\mathcal{U}^n$ by using (26). Indeed, $S(z, h(0)) = S(h^{-1}(z), 0) = S(h^{-1}z)$. If $w = (w', w_{n+1}) \in b\mathcal{U}^n$, then (according to (9)) $w = h(0)$, where

$$h = [\zeta, t] = [w', w_{n+1} - i|w'|^2].$$

Using the action (7), we see that

$$[h^{-1}(z)]_{n+1} = z_{n+1} - \bar{w}_{n+1} - 2i \sum_{k=1}^{n} z_k \bar{w}_k = 2ir(z,w).$$

Therefore $S(z,w) = c_n[r(z,w)]^{-n-1}$ with $c_n = (2i)^{-n-1}c$.

2.3.3 To conclude the proof of the theorem, we need only show that the constant c_n is given by (21). We do this by testing the reproducing formula (19) for the (naturally arising) function

$$F(z) = (z_{n+1} + i)^{-n-1} = (\overline{2ir(\mathbf{i}, z)})^{-n-1}.$$

In fact, applying (19) to this F gives

$$\begin{aligned}(2i)^{-n-1} = F(\mathbf{i}) &= c_n \int_{b\mathcal{U}^n} r(\mathbf{i}, z)^{-n-1}\, F(z)\, d\beta(z)\\ &= c_n(-2i)^{n+1} \int_{b\mathcal{U}^n} |F(z)|^2\, d\beta(z),\end{aligned}$$

so

$$c_n^{-1} = 4^{n+1} \int_{b\mathcal{U}^n} |F(z)|^2\, d\beta(z) = 4^{n+1} \int_{\zeta \in \mathbf{C}^n} \int_{t \in \mathbf{R}} \frac{dt\, d\zeta}{(t^2 + (|\zeta|^2 + 1)^2)^{n+1}}.$$

An evaluation of the last integral (see §7.4 below) gives $c_n^{-1} = 4\pi^{n+1}/n!$, and our proof is complete.

2.3.4 We shall now dispose of the technical problem deferred above, namely to show that, for each $z \in \mathcal{U}^n$, the function $S(z,w)$ can be redefined so as to be smooth in $w \in b\mathcal{U}^n$.

We have that $S(z,w) = S(h(z), h(w))$, for each $h \in \mathbf{H}^n$. Now every $w \in b\mathcal{U}^n$ can be written as $g(0)$, for a unique $g \in \mathbf{H}^n$; thus $S(z, g(0)) = S(h(z), hg(0))$ for almost all $g \in \mathbf{H}^n$. Suppose ϕ is a smooth, compactly supported function on $\mathbf{H}^n$, with

$$\int_{\mathbf{H}^n} \phi(h)\, dh = 1;$$

here dh is the invariant measure described in §2.1. An integration in h then shows that

$$\begin{aligned}S(z, g(0)) &= \int_{\mathbf{H}^n} S(h(z), hg(0))\, \phi(h)\, dh\\ &= \int_{\mathbf{H}^n} S(hg^{-1}(z), h(0))\, \phi(hg^{-1})\, dh,\end{aligned}$$

from which the smoothness of $S(z, g(0))$ in g is a straightforward matter. Further details are left to the reader.

2.4 The projection operator. The essence of the Cauchy-Szegö integral is the resulting projection operator C, which can be represented as a limit arising from (22). In view of the action of the Heisenberg group, the operator C can be explicitly described as a convolution operator on this group. We can see this as follows.

The mapping $f \mapsto C(f)$ assigns to each element $f \in L^2(b\mathcal{U}^n)$ another element of $L^2(b\mathcal{U}^n)$, of the form $C(f) = F^b$, for some $F \in \mathcal{H}^2(\mathcal{U}^n)$. As a consequence of (22) and the proposition in §2.1, we have that

$$(Cf)(z) = \lim_{\varepsilon\to 0} \int_{b\mathcal{U}^n} S(z+\varepsilon\mathbf{i}, w)\, f(w)\, d\beta(w), \qquad z \in b\mathcal{U}^n, \tag{27}$$

where the limit exists in the $L^2(b\mathcal{U}^n)$ norm.

We now use the identification of $b\mathcal{U}^n$ with $\mathbf{H}^n$. We can write each $w \in b\mathcal{U}^n$ as $w = g(0)$ for a unique $g \in \mathbf{H}^n$. In this correspondence we have that $d\beta(w) = dg$, the invariant measure on $\mathbf{H}^n$. Similarly, we write $z \in b\mathcal{U}^n$ in the form $z = h(0)$. Now

$$\begin{aligned} S(z+\varepsilon\mathbf{i}, w) &= S(g^{-1}(z+\varepsilon\mathbf{i}), g^{-1}(w)) \\ &= S(g^{-1}(z)+\varepsilon\mathbf{i}, g^{-1}(w)) = S(g^{-1}h(0)+\varepsilon\mathbf{i}, 0). \end{aligned}$$

Let us put $K_\varepsilon(h) = S(h(0)+\varepsilon\mathbf{i}, 0)$; also set $f(g) = f(g(0)) = f(w)$ and $(Cf)(h) = (Cf)(h(0)) = (Cf)(z)$. Then

$$(Cf)(h) = \lim_{\varepsilon\to 0} \int_{\mathbf{H}^n} K_\varepsilon(g^{-1}\cdot h)\, f(g)\, dg, \qquad f \in L^2(\mathbf{H}^n), \tag{28}$$

where the limit is taken in $L^2(\mathbf{H}^n)$.†

At this stage it will be convenient to change our notation slightly, writing $x, y, \dots$ for points in $\mathbf{H}^n$, and also to recall the general definition of convolution:

$$(f * g)(x) = \int f(y)\, g(y^{-1}\cdot x)\, dy.$$

We see that (28) can be formally rewritten as

$$(Cf)(x) = (f * K)(x),$$

where K is the distribution given by $\lim\limits_{\varepsilon\to 0} K_\varepsilon$. Now if $x = [\zeta, t]$, then

$$K_\varepsilon(x) = c[t + i|\zeta|^2 + i\varepsilon]^{-n-1}$$

† Note that here products such as $g\cdot h$ denote group multiplication, and g^{-1} denotes the group inverse of g.

because of (9) and (20), with $c = (2i)^{n+1}c_n = 2^{n-1}i^{n+1}n!/\pi^{n+1}$. Since

$$K_\varepsilon(x) = -\frac{\partial}{\partial t}\left(\frac{c}{n}[t + i|\zeta|^2 + i\varepsilon]^{-n}\right),$$

and the function $[t + i|\zeta|^2]^{-n}$ is locally integrable on $\mathbf{H}^n$, we see that the distribution K is given by

$$K = -\frac{\partial}{\partial t}\left(\frac{c}{n}[t + i|\zeta|^2]^{-n}\right), \tag{29}$$

and that this distribution equals the function

$$c(t + i|\zeta|^2)^{-n-1}$$

away from the origin.

Another immediate consequence of (28) is that we can write

$$(Cf)(x) = \int_{\mathbf{H}^n} K(x,y)\, f(y)\, dy \tag{30}$$

where $K(x,y) = K(y^{-1}\cdot x)$ for $x \neq y$ (i.e., $y^{-1}\cdot x \neq 0$); as such, (30) holds whenever f is an L^2 function supported in a compact set, for every x outside the support of f.

2.5 L^p theory for the Cauchy-Szegö projection. We are now in a position to apply the general L^p theory of singular integrals (as treated in Chapter 1) to the Cauchy-Szegö projection operator. Our underlying space $\mathbf{H}^n$ is of course $\mathbf{R}^{2n+1}$; however the role of the additive structure in $\mathbf{R}^{2n+1}$ is supplanted by the Heisenberg group multiplication law (6). In addition to this multiplication structure, an important part is played by appropriate *nonisotropic* dilations of $\mathbf{H}^n$, induced by the dilations (4) acting on $\mathcal{U}^n$. These are given by $x \mapsto \delta\circ x$ for $\delta > 0$; if we write $x = [\zeta, t]$, then $\delta\circ x = [\delta\zeta, \delta^2 t]$. These dilations are automorphisms of the group $\mathbf{H}^n$:

$$\delta\circ(x\cdot y) = (\delta\circ x)\cdot(\delta\circ y);$$

the standard isotropic dilations of $\mathbf{R}^{2n+1}$ are *not* automorphisms of $\mathbf{H}^n$.

We next consider the "norm" function ρ given by

$$\rho(x) = \rho([\zeta, t]) = \max(|\zeta|, |t|^{1/2}).$$

Note that $\rho(x^{-1}) = \rho(-x) = \rho(x)$ and $\rho(\delta\circ x) = \delta\rho(x)$. In addition, the function ρ satisfies a quasi-triangle inequality:

$$\rho(x\cdot y) \le c(\rho(x) + \rho(y)). \tag{31}$$

To see this, it suffices to verify it when $\rho(x) \geq \rho(y)$. Hence, by using our dilations, we can reduce matters to the case where $\rho(x) = 1$ and $\rho(y) \leq 1$, where it is clear by compactness.

On $\mathbf{H}^n$, we define the quasi-distance $\rho(x,y) = \rho(y^{-1}\cdot x)$; ρ is clearly symmetric. It is obvious from (31) that ρ satisfies the generalized triangle inequality

$$\rho(x,y) \leq c(\rho(x,z) + \rho(y,z)).$$

Using ρ, we define the balls $B(x,\delta)$ in $\mathbf{H}^n$ by

$$B(x,\delta) = \{y : \rho(y,x) < \delta\}.$$

We remark that the quasi-distance ρ, as well as the corresponding balls $B(x,\delta)$, are left-invariant under the action of $\mathbf{H}^n$; in other words

$$\rho(a\cdot x, a\cdot y) = \rho(x,y) \qquad \text{and} \qquad B(a\cdot x,\delta) = a\cdot B(x,\delta)$$

for each $a \in \mathbf{H}^n$. If we take our underlying measure $d\mu$ to be the invariant (Lebesgue) measure on $\mathbf{H}^n$, then it is easily seen that this family of balls satisfies all of the conditions set forth in Chapter 1, §1.

2.5.1 In carrying out estimates on the Heisenberg group, it is useful to observe the following integration formula. There is a positive constant c, so that whenever F is a nonnegative function on $(0,\infty)$, then

$$\int_{\mathbf{H}^n} F(\rho(x))\,dx = c\int_0^\infty F(r)r^{2n+1}\,dr.$$

To see this, we first determine c by requiring the formula to hold when F is the characteristic function of the unit interval $[0,1]$. By the homogeneity given by the dilations, it then holds when F is the characteristic function of any interval $[0,\ell]$. Taking appropriate linear combinations and limits establishes the result generally.

Since $\rho(x) \approx |t+i|\zeta|^2|^{1/2}$, the formula shows that $(t+i|\zeta|^2)^{-n-1}$ is not locally integrable and that $(t+i|\zeta|^2)^{-n}$ is locally integrable, as was asserted above.

2.5.2 Next, we consider the volume function $V(x,y)$ described in Chapter 1, §6.5. Note that $V(x,y) = |B(y,\delta)|$ with $\delta = \rho(x,y)$, thus

$$V(x,y) = |B(0,\delta)| = c'\delta^{2n+2} = c'\rho(x,y)^{2n+2}.$$

Observe also that the kernel

$$K(x) = c(|\zeta|^2 + it)^{-n-1}$$

of the Cauchy-Szegö projection satisfies $|K(x)| \approx \rho(x)^{-2n-2}$. Writing $K(x,y) = K(y^{-1}\cdot x)$, we then have that $K(x,y) \approx \rho(x,y)^{-2n-2}$; in particular

$$|K(x,y)| \leq \frac{A}{V(x,y)}. \tag{32}$$

We remark that (32) can also be seen directly by exploiting the fact that $K(x)$ is homogeneous of degree $-2n-2$ with respect to our dilations, which allows one to reduce (32) to the special case when $\rho(x,y) = 1$.

A similar scaling argument shows that

$$|K(x,y) - K(x,y_0)| \leq A\cdot\frac{\rho(y,y_0)}{\rho(x,y_0)}\cdot V(x,y_0)^{-1} \tag{33}$$

whenever $\rho(x,y_0) \geq \bar{c}\rho(y,y_0)$, for some appropriately large constant $\bar{c}$. Note also that our kernel is formally self-adjoint; i.e., $K(x,y) = \bar{K}(y,x)$.

Thus the Cauchy-Szegö projection operator C satisfies the conditions set forth in Chapter 1, §6.5. Theorem 3 of that chapter is therefore applicable to C (of course, C is bounded on L^q, $q = 2$, because it is an orthogonal projection); the results described in Chapter 1, §7 also apply to C. We summarize this by stating here one of the main conclusions that follow:

THEOREM 2. *The Cauchy-Szegö projection has an extension to a bounded operator from $L^p(\mathbf{H}^n)$ to itself, for $1 < p < \infty$.*

This theorem is in fact a special case of a series of analogous results for singular integrals on the Heisenberg group and its generalizations. These may be found in §5.2 of this chapter and in §5.3 of the next chapter. The main difficulty that distinguishes these generalizations from the Cauchy-Szegö operator just treated is that the L^2 boundedness is no longer automatic.

2.6 The Lie algebra of $\mathbf{H}^n$ and Cauchy-Riemann operators.

The Lie algebra of the Heisenberg group is a vector space which, together with a Lie bracket operation defined on it, represents the "infinitesimal action" of $\mathbf{H}^n$. This structure can be defined as follows.

To begin with, consider the tangent space of $\mathbf{H}^n$ at the origin. Using the coordinates $[\zeta,t]$ for points in $\mathbf{H}^n$ (here $\zeta = (\zeta_j) = (x_j+iy_j) \in \mathbf{C}^n$), we can write any tangent vector τ (at the origin) as

$$\tau = \sum_{j=1}^n\left(a_j\frac{\partial}{\partial x_j} + b_j\frac{\partial}{\partial y_j}\right) + c\frac{\partial}{\partial t}.$$

Here a_j, b_j, and c are real scalars which can be taken as the coordinates of τ. Next we can identify such a tangent vector τ with a left-invariant vector field $V = V^{(\tau)}$ on $\mathbf{H}^n$. Recall that a vector field V can be thought of as a first-order differential operator that annihilates constants. In our coordinate system, a vector field can be written in the form

$$(Vf) = \sum_{j=1}^{n}\left(A_j\frac{\partial f}{\partial x_j} + B_j\frac{\partial f}{\partial y_j}\right) + C\frac{\partial f}{\partial t}, \tag{34}$$

where now the A_j, B_j, and C are real-valued *functions* on $\mathbf{H}^n$.

A vector field is *left-invariant* whenever, for all smooth functions f,

$$(Vf_h) = (Vf)_h; \tag{35}$$

here f_h is the left translate of f on $\mathbf{H}^n$, given by

$$f_h(g) = f(hg).$$

The identification of a left-invariant vector field $V^{(\tau)}$ with a tangent vector τ comes about by requiring that $V^{(\tau)}$ at the origin equals τ; that is, we take $a_j = A_j(0)$, $b_j = B_j(0)$, and $c = C(0)$. Thus it is clear that to each such vector field there is naturally associated a tangent vector at the origin. In the other direction, given a tangent vector τ at the origin, let $s \mapsto \gamma(s)$ be any curve in $\mathbf{H}^n$ with $\gamma(0) = 0$ and $\dot\gamma(0) = \tau$. Then the corresponding vector field $V^{(\tau)}$ is given by

$$(V^{(\tau)}f)(h) = \frac{d}{ds}f(h\cdot\gamma(s))|_{s=0}. \tag{36}$$

Note that $V^{(\tau)}$ is left-invariant, because right group translations commute with left group translations. Observe also that (35) gives us

$$(V^{(\tau)}f)(h) = (V^{(\tau)}f_h)(0),$$

and hence $V^{(\tau)}$ is completely determined by τ.

Let $\mathfrak{h}^n$ denote the vector space of left-invariant vector fields on $\mathbf{H}^n$. Note that this linear space is closed with respect to the bracket operation

$$[V_1, V_2] = V_1V_2 - V_2V_1.$$

The space $\mathfrak{h}^n$, equipped with this bracket, is referred to as the *Lie algebra* of $\mathbf{H}^n$.

2.6.1 The Lie algebra structure of $\mathfrak{h}^n$ is most readily understood by describing it in terms of an appropriate basis.

Let X_j denote the left-invariant vector field on $\mathbf{H}^n$ that equals $\partial/\partial x_j$ at the origin. To determine it explicitly, let $\gamma(s) = \tau\cdot s$ be the straight line in $\mathbf{H}^n$ passing through the origin whose tangent vector τ has coordinates

$$a_j = 1, \qquad a_k = 0 \quad k \neq j, \qquad b_k = 0 \quad \text{all } k, \qquad c = 0.$$

Then according to (36)

$$(X_jf)(h) = \frac{d}{ds}f(h\cdot\tau s)|_{s=0}.$$

Writing $h = [\zeta, t]$ and using the multiplication law (6) shows that

$$h\cdot\tau s = [\zeta + \tau' s, t + 2y_js];$$

here we have written τ' for the projection of τ on $\mathbf{C}^n$. Thus

$$X_j = \frac{\partial}{\partial x_j} + 2y_j\frac{\partial}{\partial t}. \tag{37}$$

Similarly, if Y_j is the left-invariant vector field on $\mathbf{H}^n$ that equals $\partial/\partial y_j$ at the origin, then

$$Y_j = \frac{\partial}{\partial y_j} - 2x_j\frac{\partial}{\partial t}. \tag{38}$$

Finally, if T is the left-invariant vector field on $\mathbf{H}^n$ that equals $\partial/\partial t$ at the origin, then

$$T = \frac{\partial}{\partial t}. \tag{39}$$

The $2n+1$ vector fields $X_1, \ldots, X_n, Y_1, \ldots, Y_n, T$ form a basis for $\mathfrak{h}^n$. Note that we have the commutation relations

$$[Y_j, X_k] = 4\delta_{jk}T, \tag{40}$$

and that all other commutators vanish.

2.6.2 The Lie algebra structure of $\mathfrak{h}^n$ closely mirrors the group structure of $\mathbf{H}^n$. The passage from $\mathbf{H}^n$ to $\mathfrak{h}^n$ is given by *differentiation* (as in (36)). In the reverse direction, one goes from $\mathfrak{h}^n$ to $\mathbf{H}^n$ by using the integration process referred to as *exponentiation*; this allows one to pass from a vector field to a corresponding (local) one-parameter group of transformations. Further details are in §7.14 below.

2.6.3 We now return to our basic half-space $\mathcal{U}^n \subset \mathbf{C}^{n+1}$, whose boundary $b\mathcal{U}^n$ has been identified with $\mathbf{H}^n$. The underlying complex structure of $\mathbf{C}^{n+1}$ leads us to consider the *tangential Cauchy-Riemann operators*. These are complex vector fields on $b\mathcal{U}^n$ that are characterized by the following two properties:

(i) The vector field on $b\mathcal{U}^n$ arises by restricting (to $b\mathcal{U}^n$) a vector field that, in the coordinates $z_1, \ldots, z_{n+1}$ of $\mathbf{C}^{n+1}$, can be written in the form

$$\sum_{j=1}^{n+1} \alpha_j \frac{\partial}{\partial \bar{z}_j}, \tag{41}$$

where the α_j are complex valued functions. This is precisely the class of first-order differential operators that annihilate holomorphic functions.

(ii) The vector field (41) is tangential at $b\mathcal{U}^n$; this can be restated by the requirement that

$$\sum_{j=1}^{n+1} \alpha_j \frac{\partial}{\partial \bar{z}_j} r(z) = 0, \qquad \text{wherever} \quad r(z) = 0.$$

Here r is the defining function given by (8).

An obvious counting argument shows that, at each point of $b\mathcal{U}^n$, the vector space of restrictions of the Cauchy-Riemann operators has complex dimension n. The connection of the above with the Lie algebra $\mathfrak{h}^n$ is as follows. We define the complex vector fields

$$\bar{Z}_j = \frac{1}{2}(X_j + iY_j) = \frac{\partial}{\partial \bar{\zeta}_j} - i\zeta_j \frac{\partial}{\partial t}, \qquad j = 1, \ldots, n. \tag{42}$$

We also set

$$Z_j = \frac{1}{2}(X_j - iY_j) = \frac{\partial}{\partial \zeta_j} + i\bar{\zeta}_j \frac{\partial}{\partial t}, \qquad j = 1, \ldots, n. \tag{43}$$

Here $\dfrac{\partial}{\partial \zeta_j} = \dfrac{1}{2}\left(\dfrac{\partial}{\partial x_j} - i\dfrac{\partial}{\partial y_j}\right)$, $\dfrac{\partial}{\partial \bar{\zeta}_j} = \dfrac{1}{2}\left(\dfrac{\partial}{\partial x_j} + i\dfrac{\partial}{\partial y_j}\right)$. The commutation relations (40) then become

$$[\bar{Z}_j, Z_k] = 2i\delta_{jk}T, \tag{44}$$

with all other commutators among the Z_j, $\bar{Z}_j$, and T vanishing.

PROPOSITION. *Using the identification of* $\mathbf{H}^n$ *with* $b\mathcal{U}^n$, *the vector fields* $\bar{Z}_j$, $j = 1, \ldots, n$, *are tangential Cauchy-Riemann operators on* $b\mathcal{U}^n$ *and form a basis of such operators.*

Proof. The fact that the $\bar{Z}_j$ are tangential follows from their very definition as vector fields on $\mathbf{H}^n$ (alias $b\mathcal{U}^n$). One may check by direct computation that they are of the form (41), but it is easier to proceed as follows.

At the origin of $\mathbf{H}^n$ (which corresponds to the origin of $b\mathcal{U}^n$) $\bar{Z}_j = \partial/\partial\bar{\zeta}_j$, agreeing with the form (41). Next, the vector field $\bar{Z}_j$ is left-invariant, and we have observed (in §1.4.1) that this left action on $\mathbf{H}^n$ is induced by the holomorphic self-mappings of $\mathbf{C}^{n+1}$ given by the translations (7). Since holomorphic mappings preserve the form of (41), we have shown that each $\bar{Z}_j$ satisfies (i). Finally, at each point of $b\mathcal{U}^n$ there are n linearly independent restrictions of the $\bar{Z}_j$. Allowing variable coefficients, we see that the $\bar{Z}_j$ form a basis for the tangential Cauchy-Riemann operators on $b\mathcal{U}^n$.

These assertions concerning the $\bar{Z}_j$ can also be shown by direct calculation, without recourse to invariance. Since this calculation is useful, we present it. Recall the coordinate change $\zeta_j = z_j$ for $1 \le j \le n$, $t = \mathrm{Re} z_{n+1}$, and $r = \mathrm{Im} z_{n+1} - |z'|^2$ that was described in §1.4.4. Note that as a result

$$\frac{\partial}{\partial \bar{z}_j} = \frac{\partial}{\partial \bar{\zeta}_j} - z_j \frac{\partial}{\partial r}, j = 1, \ldots, n, \qquad \text{and} \qquad \frac{\partial}{\partial \bar{z}_{n+1}} = \frac{1}{2}\left(\frac{\partial}{\partial t} + i\frac{\partial}{\partial r}\right).$$

Next observe that $\dfrac{\partial}{\partial \bar{z}_j} - 2iz_j \dfrac{\partial}{\partial \bar{z}_{n+1}}$ is a Cauchy-Riemann operator that is tangential, because it annihilates the function $r(z)$. Substituting in the above gives that

$$\frac{\partial}{\partial \bar{z}_j} - 2iz_j \frac{\partial}{\partial \bar{z}_{n+1}} = \frac{\partial}{\partial \bar{\zeta}_j} - i\zeta_j \frac{\partial}{\partial t};$$

this yields the formula (42).

2.6.4 We shall pursue in the next chapter the above identification of the tangential Cauchy-Riemann operators, when we continue our study of the relation of $\mathbf{H}^n$ with complex analysis. Now, however, we take up another important aspect of the Heisenberg group, namely the relationship of its Lie algebra with the theory of quantum mechanics.

3. Formalism of quantum mechanics and the Heisenberg group

We come now to the second main source of the study of the Heisenberg group: the mathematical ideas connected with the fundamental notions of quantum mechanics. It is our intention here to explain briefly the formalism involved and to describe some of the underlying heuristics. In this spirit we shall not limit ourselves to assertions that have a precise formulation and proof. However, related rigorous mathematical propositions will be presented in §4 and §5 below, as well as in the appendix to this chapter.

3.1 Commutation relations and quantization. The first point to recall is the recipe, which was developed in the early part of this century, that prescribes the construction of the "quantized" version of a system originally formulated in terms of "classical" mechanics.

To begin with, a classical system is determined by (say) $2n$ coordinates

$$p_1, \ldots, p_n, q_1, \ldots, q_n$$

in phase space, where the q_j are variables representing the positions of the constituents of the system and the p_j represent the corresponding momenta. The relevant physical quantities (the "observables") are then real-valued functions of the $2n$ variables.

The crucial idea involved in the quantum-mechanical description of such a system is that the coordinate variables p_j and q_j should be replaced by operators P_j and Q_j, acting on a Hilbert space, that satisfy the commutation relations[‡]

$$[P_j, Q_k] = \frac{h}{2\pi i}\delta_{jk} I; \tag{45}$$

here I is the identity operator and h is a (physically determined) constant. Moreover, if $a(q_1, \ldots, q_n, p_1, \ldots, p_n)$ is a relevant classical observable, its quantum-mechanical analogue is the "corresponding" operator

$$a(Q_1, \ldots, Q_n, P_1, \ldots, P_n). \tag{46}$$

The mathematical issues that arise are then the explicit presentation of the operators P_j and Q_j, and the question as to what meaning should be attached to (46).

Turning to (45), and the first issue mentioned, we consider the standard "representation" of the operators P_j and Q_j in the situation that arises when $h = 1$ (this value of h can always be achieved by a simple change of scale). We take our underlying Hilbert space to be $L^2(\mathbf{R}^n)$, and set

$$(Q_j f)(x) = x_j f(x), \qquad (P_j f)(x) = \frac{1}{2\pi i} \cdot \frac{\partial}{\partial x_j} f(x), \tag{47}$$

where the operators are defined initially on the space of test functions. Observe that (45) is satisfied with $h = 1$.

3.1.1 A simple illustration of the procedure of quantization, where the problem of defining the operator (46) can be solved directly, arises

‡ Notice the close resemblance of the commutation relations (45) to those for the Lie algebra of the Heisenberg group (40). We shall return to this matter below.

for the one-dimensional system underlying simple harmonic motion. In its classical formulation, the energy of a system is given by

$$\frac{1}{2}\left(\frac{p^2}{m} + kq^2\right).$$

Here m is the mass of the particle, $p^2/2m$ represents the kinetic energy, and $kq^2/2$ represents the potential energy (k is a positive physical constant). The quantum version of the harmonic oscillator is then determined by the second-order differential operator

$$\frac{-1}{8\pi^2 m}\frac{d^2}{dx^2} + \frac{k}{2}x^2,$$

which is reducible to the Hermite operator $x^2 - \frac{d^2}{dx^2}$ (after a change of scale). This operator occurs also, implicitly, in the study of the $\bar{\partial}_b$ complex in the next chapter.

3.2 Weyl correspondence. The question we now turn to is that of the meaning to be given to the operator (46) in general, where the P's and Q's are as in (47). What can one require of the rule assigning an operator $a(Q, P)$ to the function $a(q, p)$? The least is that the mapping be linear, and that it has the expected result when $a(q,p) = a_2(q)$ or $a(q,p) = a_1(p)$. In the first case we would want $a_2(Q)$ to be the multiplication operator $f \mapsto a_2(\cdot)f(\cdot)$, and in the second case we would expect the multiplier operator $\widehat{f} \mapsto a_1(\cdot)\widehat{f}(\cdot)$.

In one sense, this problem has already been solved (by the study of pseudo-differential operators, as in Chapter 6). Indeed, if we rewrite $a(q,p)$ as $a(x,\xi)$, then one possible meaning to attach to (46) is to set

$$a(Q, P) = T_a,$$

where T_a is the operator with symbol a, defined by (1) in Chapter 6. However, this assignment of T_a to a has the effect that $T_a = T_{a_2}T_{a_1}$, if $a(q,p) = a_2(q)a_1(p)$, in accordance with its preference for putting the P operations to the right of the Q operations.

Now the rule $a \mapsto T_a$ is not the only one that satisfies our simple requirements above, due to the noncommutativity of the P's and Q's. Another rule arises from the possibility of treating the P and Q operations (i.e., the ξ and x variables) on a more symmetric footing. The idea of formulating a more symmetric definition of (46) goes back to Weyl and is based on the following simple idea: if $a(q,p)$ is the complex-valued function

$$a(q,p) = e^{2\pi i(u\cdot q + v\cdot p)},$$

where u and v are real vectors, then $a(Q,P)$ should be the (unitary) operator given by exponentiating the skew-adjoint operator $2\pi i(u\cdot Q+v\cdot P)$; i.e.,

$$a(Q,P)=e^{2\pi i(u\cdot Q+v\cdot P)}=W(u,v). \tag{48}$$

Moreover, since "any" function $a(q,p)$ can be written, via the Fourier transform, as an integral of exponentials $e^{2\pi i(u\cdot q+v\cdot p)}$ with respect to u and v, we can expect to write the corresponding operator as a similar integral. If we write $W(u,v)$ for the unitary operator given by (48), we have therefore been led to the definition of the *Weyl correspondence*, which is the assignment $a\mapsto \mathrm{Op}(a)$ given by

$$\mathrm{Op}(a)=\int_{\mathbf{R}^n\times\mathbf{R}^n} W(u,v)\,\widehat{a}(u,v)\,du\,dv, \tag{49}$$

where

$$\widehat{a}(u,v)=\int_{\mathbf{R}^n\times\mathbf{R}^n} e^{-2\pi i(x\cdot u+\xi\cdot v)}\,a(x,\xi)\,dx\,d\xi. \tag{50}$$

3.2.1 We shall now make more explicit the formulae describing the Weyl correspondence. First, we have the fact that

$$[W(u,v)(f)](x)=e^{2\pi i x\cdot u}e^{\pi i u\cdot v}f(x+v), \tag{51}$$

for any $f\in L^2(\mathbf{R}^n)$.

Indeed, the definition $W(u,v)=e^{2\pi i(u\cdot Q+v\cdot P)}$ requires that, for s_1 and s_2 real,

$$W(us_1,vs_1)\cdot W(us_2,vs_2)=W(u(s_1+s_2),v(s_1+s_2)) \qquad \text{and}$$
$$\frac{\partial}{\partial s}W(us,vs)|_{s=0}=2\pi i(u\cdot Q+v\cdot P).$$

It is easy to check that the $W(u,v)$ given by (51) satisfies these conditions and, moreover, that this determines W uniquely.

We next substitute (51) and (50) in the definition (49), having rewritten (50) as

$$\widehat{a}(u,v)=\int e^{-2\pi i(y\cdot u+\xi\cdot v)}\,a(y,\xi)\,dy\,d\xi,$$

and obtain a quadruple integral (involving integration with respect to $du\,dv\,dy\,d\xi$). If we carry out the integration in u first, the result is

$$\int_{\mathbf{R}^n\times\mathbf{R}^n\times\mathbf{R}^n}\delta(x-y+v/2)\,f(x+v)\,e^{-2\pi i\xi\cdot v}\,a(y,\xi)\,dv\,dy\,d\xi;$$

here δ is the Dirac delta function. An integration in y and an obvious change of variables then gives

$$(\mathrm{Op}(a)f)(x)=\int_{\mathbf{R}^n\times\mathbf{R}^n} a([x+y]/2,\xi)\,e^{2\pi i\xi\cdot(x-y)}\,f(y)\,dy\,d\xi. \tag{52}$$

This formula is to be compared with the formula for T_a, namely

$$(T_af)(x)=\int_{\mathbf{R}^n\times\mathbf{R}^n} a(x,\xi)\,e^{2\pi i\xi\cdot(x-y)}\,f(y)\,dy\,d\xi.$$

We shall see below that, in a precise sense, the difference $\mathrm{Op}(a)-T_a$ is small.

As we have stated above, one is led to the correspondence (52) by considerations of symmetry in the p and q variables. This is part of a wider symmetry, "symplectic invariance". These considerations in fact uniquely determine the mapping $a\mapsto\mathrm{Op}(a)$; see §7.5 and §7.6 below.

3.2.2 We remark that the basic operator

$$W(u,v)=e^{2\pi i(u\cdot Q+v\cdot P)}$$

is in reality the restriction of a unitary representation of the Heisenberg group.† By this we mean the following. For each $[\zeta,t]\in\mathbf{H}^n$, with $\zeta=u+iv$, define the operator $R(\zeta,t)=R(u,v,t)$ by

$$(R(\zeta,t)f)(x)=e^{2\pi i[u\cdot x+u\cdot v/2+t/4]}\,f(x+v). \tag{53}$$

Then, as is easily verified, the mapping $[\zeta,t]\mapsto R(\zeta,t)$ is a homomorphism from $\mathbf{H}^n$ to the group of unitary operators on $L^2(\mathbf{R}^n)$; i.e.,

$$R(\zeta,t)\,R(\eta,s)=R(\zeta+\eta,t+s+2\,\mathrm{Im}(\zeta\cdot\bar\eta)).$$

Moreover, clearly $R(\zeta,0)=W(u,v)$ for $\zeta=u+iv$.

Given the representation R of $\mathbf{H}^n$, we can pass to the corresponding representation of the Lie algebra $\mathfrak{h}^n$ by differentiation. The basis elements X_j, Y_j, and T of $\mathfrak{h}^n$ then correspond to

$$\frac{\partial R}{\partial u_j}(0),\quad \frac{\partial R}{\partial v_j}(0),\quad \text{and}\quad \frac{\partial R}{\partial t}(0),$$

respectively. Carrying out the differentiation in (53) shows that X_j, Y_j, and T are thereby represented by $2\pi iQ_j$, $2\pi iP_j$, and $\pi i/2I$. In this way, we see that we can consider (45), together with the choice (47), as giving us a representation of the Lie algebra of the Heisenberg group.

† See the appendix below for a discussion of the representation theory of $\mathbf{H}^n$.

3.3 Twisted convolution. We focus our attention on the phase space $\mathbf{R}^n \times \mathbf{R}^n$, which we identify with $\mathbf{C}^n$ via $\zeta \in \mathbf{C}^n$, $\zeta = u + iv \leftrightarrow (u, v) \in \mathbf{R}^n \times \mathbf{R}^n$.

On it, we consider the symplectic form $\langle \cdot, \cdot \rangle$ given to us by the Heisenberg group multiplication law (6) and defined by

$$\langle z, w \rangle = 2 \operatorname{Im} (z \cdot \bar{w}) = 2 \operatorname{Im} \Big(\sum_{j=1}^{n} z_j \bar{w}_j \Big),$$

whenever $z, w \in \mathbf{C}^n$. With λ a fixed real constant, we can define the *twisted convolution* of two functions F and G by

$$(F *_\lambda G)(z) = \int_{\mathbf{C}^n} F(z - w)\, G(w)\, e^{-2\pi i \lambda \langle z, w \rangle}\, dw; \tag{54}$$

here dw is the Euclidean measure on $\mathbf{C}^n$. Notice that, in view of the antisymmetry of $\langle \cdot, \cdot \rangle$, we have that $\langle z - w, w \rangle = -\langle w, z \rangle$; thus

$$G *_\lambda F = F *_{-\lambda} G,$$

and the twisted product is not (in general) commutative.

The interest of twisted convolution is twofold. First, it occurs when we analyze the convolution of functions on the Heisenberg group in terms of the Fourier transform in the t-variable. To see this, let $f(\zeta, t)$ be (say) a test function on $\mathbf{H}^n$. Define

$$f_\lambda^0(\zeta) = f(\zeta, \cdot)\widehat{\ }(\lambda) = \int_{\mathbf{R}} f(\zeta, t)\, e^{-2\pi i \lambda t}\, dt. \tag{55}$$

Similarly define g_λ^0 when g is another test function on $\mathbf{H}^n$. Suppose $f * g$ is the convolution of f and g on $\mathbf{H}^n$, that is

$$(f * g)(x) = \int_{\mathbf{H}^n} f(y)\, g(y^{-1} \cdot x)\, dy.$$

Then

$$(f * g)_\lambda^0 = f_\lambda^0 *_\lambda g_\lambda^0. \tag{56}$$

Indeed, with $g = [w, s]$ and $x = [z, t]$, then by the multiplication law of $\mathbf{H}^n$,

$$\begin{aligned}(f * g)_\lambda^0 &= \int_{\mathbf{H}^n \times \mathbf{R}} e^{-2\pi i \lambda t} f(w, s)\, g(z - w, t - s + \langle z, w \rangle)\, dw\, ds\, dt \\ &= \int f(w, s) g(z - w, t - s) e^{-2\pi i \lambda (t - s)}\, e^{-2\pi i \lambda s}\, e^{2\pi i \lambda \langle z, w \rangle}\, dw\, ds\, dt \\ &= f_\lambda^0 *_\lambda g_\lambda^0,\end{aligned}$$

proving (56).

Plancherel's formula with (56) has the following consequence. If we want to prove the boundedness on $L^2(\mathbf{H}^n)$ of a convolution operator T, given by $Tf = f * K$, then formally this is equivalent to proving the uniform boundedness on $L^2(\mathbf{C}^n)$ of the family of oscillatory operators given by

$$F \mapsto F *_\lambda K_\lambda^0, \qquad -\infty < \lambda < \infty. \tag{57}$$

3.3.1 The second interest of twisted convolution is that it explicitly implements the product formula for the Weyl correspondence in terms of the symbols.

Indeed, if we recall the formula (49), write analogously

$$\operatorname{Op}(b) = \int W(u', v')\, \widehat{b}(u', v')\, du'\, dv',$$

and set $z = u + iv$, $w = u' + iv'$; then

$$\operatorname{Op}(a) \operatorname{Op}(b) = \int_{\mathbf{C}^n \times \mathbf{C}^n} W(z)\, W(w)\, \widehat{a}(z)\, \widehat{b}(w)\, dz\, dw.$$

However, since $W(z) = R(z, 0)$, we have that

$$W(z)\, W(w) = R(z + w, \langle z, w \rangle) = W(z + w) e^{i\pi \langle z, w \rangle / 2},$$

according to (53) and (51). This shows that

$$\operatorname{Op}(a) \operatorname{Op}(b) = \int_{\mathbf{C}^n \times \mathbf{C}^n} W(z + w)\, \widehat{a}(z)\, \widehat{b}(w) e^{i\pi \langle z, w \rangle / 2}\, dz\, dw,$$

which after an obvious change of variables is

$$\int_{\mathbf{C}^n \times \mathbf{C}^n} W(z)\, \widehat{a}(z - w)\, \widehat{b}(w) e^{i\pi \langle z, w \rangle / 2}\, dz\, dw.$$

The result is

$$\operatorname{Op}(a) \cdot \operatorname{Op}(b) = \operatorname{Op}(c), \qquad \text{with} \quad \widehat{c} = \widehat{a} *_{1/4} \widehat{b}. \tag{58}$$

4. Weyl correspondence and pseudo-differential operators

The heuristics and formal considerations above have led us to two concepts we now examine more closely: the Weyl correspondence and twisted convolution.

4.1 Beginning with the Weyl correspondence, we see that our previous manipulations have brought us to the formula (52). We take that as our starting point, and define $\operatorname{Op}(a)$ in terms of it.

Suppose $a = a(x, \xi)$ belongs to the standard symbol class S^m (as in Chapter 6, §1.3). To give a precise definition of $\operatorname{Op}(a)f$, that is, to make sense of

$$(\operatorname{Op}(a) f)(x) = \int_{\mathbf{R}^n \times \mathbf{R}^n} a\Big(\frac{x + y}{2}, \xi\Big) e^{2\pi i \xi \cdot (x - y)}\, f(y)\, dy\, d\xi, \tag{59}$$

notice that if $c(x, y, \xi) = a([x + y]/2, \xi)$, then c is a compound symbol (in the sense of Chapter 6, §6.1). Therefore the operator (59) is defined (as a mapping from $\mathcal{S}$ to itself) by setting $\operatorname{Op}(a) = T_{[c]}$, in the notation of that chapter. Moreover, a direct application of the proposition in Chapter 6, §6.1 then gives the following.

PROPOSITION 1. (i) *If $a \in S^m$ then there exists $a' \in S^m$ so that*

$$\mathrm{Op}(a) = T_{a'}. \tag{60}$$

(ii) *The symbol a' differs from a by a symbol of order $m-1$. More precisely, one has the asymptotic formula*

$$a' \sim \sum_\alpha \frac{1}{\alpha!} \cdot (\partial_\xi)^\alpha \left(\frac{1}{4\pi i}\frac{\partial}{\partial x}\right)^\alpha a(x,\xi). \tag{61}$$

The formula (61) is an immediate consequence of the asymptotic formula (58) of Chapter 6, with $c(x,y,\xi) = a([x+y]/2,\xi)$.

As a result, all the regularity assertions involving estimates for L^p, Sobolev, and Lipschitz spaces that were proved in §5 of Chapter 6 for the operators T_a hold also for the Weyl analogues $\mathrm{Op}(a)$.

4.2 We note a few additional facts about the Weyl correspondence.

The symmetric feature of the Weyl correspondence is reflected in the identity

$$(\mathrm{Op}(a))^* = \mathrm{Op}(\bar{a}), \tag{62}$$

where $\mathrm{Op}(a)^*$ is defined by

$$\langle \mathrm{Op}(a)^* f, g\rangle = \langle f, \mathrm{Op}(a) g\rangle. \tag{63}$$

In fact, (62) is obvious from the definition (59), $\mathrm{Op}(a) = T_{[c]}$, and the fact that $c(x,y,\xi) = c(y,x,\xi)$, if $c(x,y,\xi) = a([x+y]/2,\xi)$.

In contrast, for the usual pseudo-differential operators, the formula for the adjoint T_a^* is more complicated; see §6.2 of Chapter 6.

4.3 We observe next that if we have $a \in L^2(\mathbf{R}^n \times \mathbf{R}^n)$, then the operator T_a is not only bounded on $L^2(\mathbf{R}^n)$, but belongs to the Hilbert-Schmidt class.

Recall that an operator T is in that class if it is representable by a kernel K,

$$(Tf)(x) = \int_{\mathbf{R}^n} K(x,y)\, f(y)\, dy \tag{64}$$

with $K \in L^2(\mathbf{R}^n \times \mathbf{R}^n)$. In that case, $(Tf)(x)$ is well defined for almost every $x \in \mathbf{R}^n$ whenever $f \in L^2(\mathbf{R}^n)$. We set $\|T\|_{\mathrm{HS}} = \|K\|_{L^2(\mathbf{R}^n\times\mathbf{R}^n)}$.

Note that each operator of the above type is automatically bounded on $L^2(\mathbf{R}^n)$; indeed, by Schwarz's inequality

$$|(Tf)(x)|^2 \le \int_{\mathbf{R}^n} |K(x,y)|^2\, dy \cdot \|f\|_{L^2}^2,$$

and an integration in x proves our assertion. It should also be noted that if T_1 and T_2 are Hilbert-Schmidt operators, then so is T_1T_2 and

$$\|T_1T_2\|_{\mathrm{HS}} \le \|T_1\|_{\mathrm{HS}} \|T_2\|_{\mathrm{HS}}. \tag{65}$$

In fact, if T_j is represented by the kernel K_j, then T is represented by

$$K(x,y) = \int_{\mathbf{R}^n} K_1(x,z)\, K_2(z,y)\, dz.$$

Thus Schwarz's inequality shows that

$$|K(x,y)|^2 \le \int_{\mathbf{R}^n} |K_1(x,z)|^2\, dz \cdot \int_{\mathbf{R}^n} |K_2(z,y)|^2\, dz,$$

and an integration in x and y then verifies the assertion (65).

With these preliminaries out of the way, we can state our result.

PROPOSITION 2. (i) *Suppose $a \in S^m \cap L^2(\mathbf{R}^n \times \mathbf{R}^n)$. Then the operator $\mathrm{Op}(a)$ is in the Hilbert-Schmidt class with*

$$\|\mathrm{Op}(a)\|_{\mathrm{HS}} = \|a\|_{L^2(\mathbf{R}^n\times\mathbf{R}^n)}. \tag{66}$$

(ii) *Moreover, the mapping*

$$a \mapsto \mathrm{Op}(a)$$

extends to a unitary isomorphism between $L^2(\mathbf{R}^n \times \mathbf{R}^n)$ and the Hilbert-Schmidt class.

Proof. Let us assume first that $a(x,\xi)$ is a test function on $\mathbf{R}^n \times \mathbf{R}^n$. Then if the kernel $k(x,z)$ is defined as in Chapter 6, §2.2,

$$k(x,z) = \int_{\mathbf{R}^n} e^{2\pi i z\cdot\xi}\, a(x,\xi)\, d\xi,$$

we have, according to (59), that the kernel K of $\mathrm{Op}(a)$ equals

$$K(x,y) = k([x+y]/2, x-y). \tag{67}$$

Notice that, as a result, the mapping $a \mapsto K$ is an isomorphism from test functions (on $\mathbf{R}^n \times \mathbf{R}^n$) to itself. Moreover $\|a\|_{L^2(\mathbf{R}^n\times\mathbf{R}^n)} = \|k\|_{L^2(\mathbf{R}^n\times\mathbf{R}^n)}$, and

$$\begin{aligned}\int_{\mathbf{R}^n\times\mathbf{R}^n} |K(x,y)|^2\, dx\, dy &= \int_{\mathbf{R}^n\times\mathbf{R}^n} |k([x+y]/2, x-y)|^2\, dx\, dy\\ &= \int_{\mathbf{R}^n\times\mathbf{R}^n} |k(x,z)|^2\, dx\, dz,\end{aligned}$$

because the Jacobian determinant of $(x,y) \mapsto ([x+y]/2, x-y)$, as a mapping of $\mathbf{R}^n \times \mathbf{R}^n$ to itself, has absolute value 1. Thus (66) is proved when a is a test function. The extension of $\mathrm{Op}(a)$ to all $a \in L^2$, together with (66) for these a, then follows by a simple limiting argument. To prove that the mapping $a \mapsto \mathrm{Op}(a)$ is unitary, it suffices to see (in view of (66)) that it is invertible and, in particular, that each operator whose kernel K is a test function on $\mathbf{R}^n \times \mathbf{R}^n$ is of the form $\mathrm{Op}(a)$ for some test function a. This follows from the fact that the mapping $a \mapsto K$ is an isomorphism of test functions, as we have already remarked. The proof of the proposition is therefore complete.

4.4 The last proposition, together with the multiplicative law for symbols for the Weyl correspondence given by (58), has a remarkable implication regarding the twisted convolutions (54).

PROPOSITION 3. *Suppose F and G belong to $L^2(\mathbf{C}^n)$, and $\lambda \neq 0$. Then $F *_\lambda G$ is in $L^2(\mathbf{C}^n)$, and*

$$\|F *_\lambda G\|_{L^2(\mathbf{C}^n)} \leq c_\lambda \|F\|_{L^2(\mathbf{C}^n)} \|G\|_{L^2(\mathbf{C}^n)} \tag{68}$$

where $c_\lambda = 2^{-n}|\lambda|^{-n/2}$.

Note. Of course, the conclusion $F *_\lambda G \in L^2(\mathbf{C}^n)$ fails utterly if $\lambda = 0$. The result for $\lambda \neq 0$ is made possible by the particular oscillatory behavior of the factor $e^{2\pi i\lambda\langle z,w\rangle}$ arising in the definition of twisted convolution. For further discussion of this point, see §7.10 below.

Proof. We establish (68) first when $\lambda = 1/4$, under the restriction that both F and G are test functions. We define the symbols a and b by

$$\widehat{a} = F, \quad \widehat{b} = G.$$

In this case, then, our conclusion is a direct consequence of (66), the multiplicativity of the Hilbert-Schmidt class described by (65), the product formula (58), and Plancherel's theorem. The case of general F and G in $L^2(\mathbf{C}^n)$ then follows by a simple limiting argument.

To pass from the case $\lambda = 1/4$ to general λ, we use a scaling argument. For each $\delta > 0$, let $F^\delta(z) = F(z/\delta)$ and $G^\delta(z) = G(z/\delta)$. Then

$$\begin{aligned}(F^\delta *_{1/4} G^\delta)(z) &= \int F(\delta^{-1}[z-w])\, G(\delta^{-1}[w])\, e^{i\pi\langle z,w\rangle/2}\, dw \\ &= \delta^{2n}(F *_{\delta^2/4} G)(\delta^{-1}[z]),\end{aligned}$$

as is seen by the change of variables $w \mapsto \delta(w)$. Since $\|F^\delta\|_{L^2} = \delta^n\|F\|_{L^2}$, $\|G^\delta\|_{L^2} = \delta^n\|G\|_{L^2}$, and $\|(F *_{\delta^2/4} G)^\delta\|_{L^2} = \delta^n\|F *_{\delta^2/4} G\|_{L^2}$, we see that

$$\|F *_{\delta^2/4} G\|_{L^2} \leq \delta^{-n}\|F\|_{L^2}\|G\|_{L^2}.$$

If we set $\lambda = \delta^2/4$, we have therefore established (68) whenever $\lambda > 0$. The result for $\lambda < 0$ then follows by complex conjugation, completing the proof of the proposition.

5. *Twisted convolution and singular integrals on* $\mathbf{H}^n$

We now exploit further the ideas involving twisted convolution. As an application, we will be able to obtain the L^2 estimates for (singular integral) convolution operators on the Heisenberg group, along the following lines. For an appropriate distribution kernel K on $\mathbf{H}^n$, we will consider the operator on $\mathbf{H}^n$ given by

$$T(f) = f * K, \tag{69}$$

where the convolution is taken with respect to the group structure of $\mathbf{H}^n$. Our goal is to show that, for K that satisfy conditions analogous to those of the $\mathbf{R}^n$ theory (namely those given in Chapter 6, §4.5), the operator (69) is bounded on $L^2(\mathbf{H}^n)$.

Needless to say, matters here cannot be directly reduced to the ordinary Fourier transform (as they can in $\mathbf{R}^n$). It is in this regard that twisted convolution enters, and the key point is to prove an analogous estimate for twisted convolution. This estimate, by its very nature, combines features of both singular integrals and oscillatory integrals, and has an interest in its own right. Then, with it, using the approach outlined in §3.3, we will be able to conclude the proof of our theorem.

5.1 Oscillatory singular integrals. We place ourselves in the setting of $\mathbf{R}^m$; the case $m = 2n$ is the one that occurs in the application below. We suppose we are given an integrable function K on $\mathbf{R}^m$ that satisfies conditions of the type set forth in Chapter 6 §4.5, namely

$$\left|\left(\frac{\partial}{\partial x}\right)^\alpha K(x)\right| \leq A_1|x|^{-m-|\alpha|}, \qquad \text{for } |\alpha| \leq 1, \tag{70}$$

and the cancellation condition

$$\left|\int_{\mathbf{R}^m} K(x)\,\phi^R(x)\,dx\right| \leq A, \qquad R > 0. \tag{71}$$

Here $\phi^R(x) = \phi(x/R)$, and ϕ ranges over our normalized "bump functions"; that is, ϕ is supported in the unit ball and

$$\left|\left(\frac{\partial}{\partial x}\right)^\alpha \phi(x)\right| \leq 1, \qquad \text{for } |\alpha| \leq N,$$

where N is a fixed integer.

In addition, here, we shall be given a real bilinear form $B(x,y)$ on $\mathbf{R}^m \times \mathbf{R}^m$. Given K and B, we define the oscillatory singular integral T_B by

$$(T_B f)(x) = \int_{\mathbf{R}^m} f(x-y)\,K(y)\,e^{iB(x,y)}\,dy. \tag{72}$$

PROPOSITION. *T_B is bounded from $L^2(\mathbf{R}^m)$ to itself. Its bound is independent of the bilinear form B, and depends only on the bounds A_1 and A appearing in* (70) *and* (71).

Note the slight difference between the assumptions on K given here and those made in §4.5 of Chapter 6. In Chapter 6, K was assumed merely to be a distribution. Here, in order to avoid irrelevant technicalities, we make the *a priori* assumption that K is integrable. However what matters is that, in our conclusion, the bounds depend only on A and A_1, and not on the L^1 norm of K nor on the bilinear form B.

5.1.1 The proof of the proposition will be in two steps. First, a simple scaling argument will reduce matters to the case when the bilinear form B has been suitably normalized. Once this is done, the operator (72) can be split into two parts: a "local" part involving integration over small y, and a "global" part involving integration over large y. For the local part the singular integral feature predominates, and for the global part the oscillatory character of the integral is decisive.

To begin with, we may assume that B is not identically zero, for otherwise the proposition is a direct consequence of the proposition in §4.5 of Chapter 6 and Plancherel's theorem. In the case $B \not\equiv 0$, we write $F *_B K$ for the integral (72) and note that, as in the scaling argument in §4.4 above, we can easily arrive at the identity

$$\delta^{-m}(F^\delta *_B K^\delta) = (F *_{\delta^2 B} K)^\delta. \tag{73}$$

Moreover, our assumptions (70) and (71) on K are such that they automatically hold also for $\delta^{-m}K(\cdot/\delta) = \delta^{-m}K^\delta$, for all $\delta > 0$, with the same bounds A_1 and A. Thus (73) allows us to replace the bilinear form B with $\delta^2 B$ for an arbitrary $\delta > 0$, and we may therefore suppose that B has been normalized so that

$$\begin{cases} |B(x,y)| \le |x|\cdot|y|, & \text{all } x, y \in \mathbf{R}^m, \\ B(\bar{x},\bar{y}) = 1 & \text{for some } \bar{x},\bar{y} \text{ with } |\bar{x}| = |\bar{y}| = 1. \end{cases} \tag{74}$$

5.1.2 We next decompose $T_B = T_0 + T_\infty$ by writing $K = K^0 + K^\infty$, where $K^0 = K\eta$, $K^\infty = K(1-\eta)$; here η is a C^∞ function supported in $|x| \le 1$ that equals 1 near the origin. Thus T_0 is the operator (72) with K replaced by K^0, and T_∞ is defined similarly.

We first estimate T_0. Fix any point $x^0 \in \mathbf{R}^m$, and let $D(x^0)$, $D^*(x^0)$ denote the balls about x^0 of radii 1 and 2, respectively. Since the kernel K^0 of T_0 is supported in the unit ball about the origin, $(T_0 f)(x)$, for $x \in D(x^0)$, is completely determined by the restriction of f to $D^*(x^0)$. To exploit this, write

$$B(x,y) = B(x-x^0,y) + B(x^0,y-x) + B(x^0,x).$$

Then if $f_{x^0}(x) = e^{-iB(x^0,x)} f(x)$, we have that

$$(T_0 f)(x) = e^{iB(x^0,x)} \int_{|y|\le 1} f_{x_0}(x-y)\, K^0(y)\, e^{iB(x-x^0,y)}\, dy.$$

This last expression can be written as the sum of two terms $I_1 + I_2$. The first, I_1, is

$$e^{iB(x^0,x)} \int_{|y|\le 1} f_{x_0}(x-y)\, K^0(y)\, dy,$$

and the second is similar, except that $K^0(y)$ is replaced by

$$K^0(y)[e^{iB(x-x^0,y)} - 1].$$

For the first term, the Euclidean singular integral theory is applicable. Indeed, since K satisfies (70) and (71), so does $K^0 = K\eta$, and by the results in §4.5 of Chapter 6,

$$\int_{D(x^0)} |I_1(x)|^2\, dx \le c \int_{D^*(x^0)} |f_{x^0}|^2\, dx \le c \int_{D^*(x^0)} |f|^2\, dx.$$

The term I_2 is estimated crudely. Indeed,

$$|I_2(x)| \le \int_{|y|\le 1} |f_{x_0}(x-y)|\cdot|K(y)|\cdot|y|\, dy.$$

Since $|K(y)|\cdot|y|\cdot\chi_{\{|y|\le 1\}} \in L^1$,

$$\int_{D(x^0)} |I_2(x)|^2\, dx \le c \int_{D^*(x^0)} |f|^2\, dx.$$

Adding these two estimates gives

$$\int_{D(x^0)} |(T_0 f)(x)|^2\, dx \le c \int_{D^*(x^0)} |f(x)|^2\, dx. \tag{75}$$

Notice that the constant c appearing in (75) depends only on A_1 and A, and is independent of x^0. Integrating the above inequality in x^0 yields

$$\int_{\mathbf{R}^m} |(T_0 f)(x)|^2\, dx \le c2^m \int_{\mathbf{R}^m} |f(x)|^2\, dx,$$

concluding our estimate for the local part of T.

5.1.3 We shall estimate the L^2 norm of the operator T_∞ by estimating the norm of $T_\infty^* T_\infty$, which in turn will be controlled by obtaining a majorization for the kernel of the operator $T_\infty^* T_\infty$. After a simple change of variables, we see that

$$(T_\infty f)(x) = e^{iB(x,x)} \int_{\mathbf{R}^m} f(y)\, K^\infty(x-y)\, e^{-iB(x,y)}\, dy.$$

Therefore the kernel $k(x,y)$ of $T_\infty^* T_\infty$ equals

$$\begin{aligned} k(x,y) &= \int_{\mathbf{R}^m} \bar{K}^\infty(z-x)\, e^{iB(z,x)}\, K^\infty(z-y)\, e^{-iB(z,y)}\, dz \\ &= e^{iB(y,x-y)} \int_{\mathbf{R}^m} \bar{K}^\infty(z-x+y)\, K^\infty(z)\, e^{iB(z,x-y)}\, dz \\ &= e^{iB(y,x-y)}\, L(x-y), \end{aligned}$$

where

$$L(u) = \int_{\mathbf{R}^m} \bar{K}^\infty(z-u)\, K^\infty(z)\, e^{iB(z,u)}\, dz. \tag{76}$$

It then suffices to estimate L.

First, there is the crude bound

$$|L(u)| \le c(1+|u|)^{-m} \log(2+|u|), \tag{77}$$

which is obtained by replacing the integrand in (76) by its absolute value. Indeed, because of (70) and the fact that $K^\infty = (1-\eta)\cdot K$,

$$\begin{aligned} |L(u)| &\le c \int (1+|z-u|)^{-m}\,(1+|z|)^{-m}\, dz \\ &= c\Bigl(\int_{|z|\le|u|/2} + \int_{|u|/2\le|z|\le 2|u|} + \int_{|z|\ge 2|u|}\Bigr). \end{aligned}$$

For the first integral we use the fact that

$$(1+|z-u|)^{-m} \le (1+|u|/2)^{-m}.$$

For the second integral, we use $(1+|z|)^{-m} \approx (1+|u|)^{-m}$ and

$$\int_{|z|\le 2|u|} (1+|z-u|)^{-m} \le \int_{|z|\le 3|u|} (1+|z|)^{-m}\, dz.$$

Finally, the third integral is dominated by $c(1+|u|)^{-m}$. This establishes (77).

However, that estimate is not sufficient for our purposes, since it does not guarantee the integrability of L at infinity. A more refined approach is obtained by exploiting the oscillatory properties afforded by $e^{iB(z,u)}$.

To do this, observe that the second condition in (74) can be restated as follows: the linear functional $\ell(u) = \langle \nabla_z B(z,u), \bar{x}\rangle$ takes the value 1 at $\bar{y}$. Now

$$e^{iB(z,u)} = \frac{1}{i\ell(u)} \langle \nabla_z e^{iB(z,u)}, \bar{x}\rangle.$$

If we insert this in (76), carry out the integration by parts, and note that $|\nabla_z K^\infty(z)| \le A_1(1+|z|)^{-m}$, we see as before that

$$|L(u)| \le c(1+|u|)^{-m} \log(2+|u|)\cdot [\ell(u)]^{-1}.$$

Combining this with (77) gives our final estimate for L,

$$|L(u)| \le c(1+|\ell(u)|)^{-1}\,(1+|u|)^{-m}\, \log(1+|u|). \tag{78}$$

The function on the right is integrable on $\mathbf{R}^m$. We show this by choosing an orthogonal coordinate system

$$\{(u_1,\dots,u_m)\}$$

so that $\{(0,u_2,\dots,u_m)\}$ is the hyperplane annihilated by the linear functional ℓ. However, if the unit vector $\bar{y}$ is given by $(\bar{y}_1,\dots,\bar{y}_n)$, then $\ell(\bar{y}) = \bar{y}_1\cdot\ell(1,0,\dots,0) = 1$. Since $|y_1|\le 1$ and $\ell(\bar{y}) = 1$, we have $|\ell(1,0,\dots,0)|\ge 1$, which means that $|\ell(u)| \ge |u_1|$ for every $u\in\mathbf{R}^m$. Now integrate the right side of (78) over $\mathbf{R}^m$ by integrating over $u_2, u_3, \dots, u_m$ first. The result is easily seen to be bounded by $c'(1+|u_1|)^{-1}(1+|u_1|)^{-1}\log(1+|u_1|)$, and this function is clearly integrable over $\mathbf{R}^1$.

To recapitulate, we have that the kernel $k(x,y)$ of $T_\infty^* T_\infty$ is dominated by $L(x-y)$, where L is integrable over $\mathbf{R}^m$. We can now invoke the lemma in §2.4.1 of Chapter 7. The result is that

$$\|T_\infty^* T_\infty\| \le \int_{\mathbf{R}^m} |L(u)|\, du.$$

This concludes the estimates for $T_\infty^* T_\infty$, hence also for T_∞, and finally for $T_B = T_0 + T_\infty$. Since all of our estimates depended only on the bounds A_1 and A, and not otherwise on our kernel K or the bilinear form B, the proposition is therefore demonstrated.

5.1.4 We now specialize to the case $m = 2n$; we identify $\mathbf{R}^m$ with $\mathbf{C}^n$, and to reflect this we use complex notation, writing $z, w, \dots$ rather than $x, y, \dots$. For the bilinear form B we take the symplectic form $\langle\cdot,\cdot\rangle$, with $\langle z,w\rangle = 2\,\mathrm{Im}(z\cdot\bar{w})$. As a special case of the proposition above, we have the following assertion regarding twisted convolution (as defined by (54)).

COROLLARY. *Suppose the integrable function K satisfies the conditions* (70) *and* (71), *with $m = 2n$. Then the operator*

$$f \mapsto f *_\lambda K \tag{79}$$

is bounded from $L^2(\mathbf{C}^n)$ to itself. The bound can be taken to depend only on the bounds A_1 and A arising in (70) *and* (71), *and not to depend on λ.*

We remark that, in the present case, the proof can in fact be simplified due to the special nature of twisted convolution. In particular, the argument for the global part of the operator given in §5.1.3 is unnecessary, since the kernel K^∞ is in $L^2(\mathbf{C}^n)$, and by the proposition in §4.4, the mapping $f \mapsto f *_\lambda K^\infty$ is bounded on $L^2(\mathbf{C}^n)$.

5.2 Singular integrals on the Heisenberg group. We shall now develop that part of the theory of singular integrals on the Heisenberg group which, among other things, extends the results obtained for the Cauchy-Szegö projection in §2.4 and §2.5 above.

From our earlier considerations, we can expect the Heisenberg group dilations

$$x = [z,t] \mapsto \delta \circ x = \delta \circ [z,t] = [\delta z, \delta^2 t]$$

and the norm function $\rho(x) = \max(|z|, |t|^{1/2})$ to play significant roles. In fact, the class of kernels we consider has an invariance property with respect to these dilations, and the inequalities they satisfy will be expressed in terms of the norm function ρ.

More precisely, we assume that K is a distribution on $\mathbf{H}^n$ that agrees with a function $K(x)$ whenever $x = [z,t] \neq 0$, and satisfies the regularity conditions

$$|K(x)| \leq A_1 \rho(x)^{-2n-2}, \tag{80}$$

$$|\nabla_z K(x)| \leq A_1 \rho(x)^{-2n-3}, \qquad \text{and} \tag{81}$$

$$\left|\frac{\partial}{\partial t} K(x)\right| \leq A_1 \rho(x)^{-2n-4}. \tag{82}$$

We also assume that K satisfies the cancellation condition

$$|K(\Phi^R)| \leq A \qquad \text{for all } R > 0, \tag{83}$$

where $\Phi^R([z,t]) = \Phi([z/R, t/R^2])$, whenever Φ is a normalized bump function on $\mathbf{H}^n$, i.e., when Φ is supported in the unit ball $\{x : \rho(x) \leq 1\}$, and $|(\partial/\partial x)^\alpha \Phi(x)| \leq 1$ for all α with $|\alpha| \leq N$, for some fixed N.

Observe that (80)–(82) and (83) are the direct analogues of conditions (38) and (39) of Chapter 6 that occurred in the $\mathbf{R}^n$ case, once we take into account that here we are using the dilations for which z has degree 1 and t has degree 2.

THEOREM 3. *Assume that the distribution K satisfies* (80)–(83) *above. Then the convolution operator T on the Heisenberg group, defined for test functions f by*

$$T(f) = f * K, \tag{84}$$

has an extension to a bounded operator from $L^2(\mathbf{H}^n)$ to itself.

5.2.1 Some examples. Before coming to the proof of the theorem, we list some examples.

(i) Suppose K_1 is a homogeneous function of degree $-2n-1$ on $\mathbf{H}^n$; i.e., $K_1([\delta z, \delta^2 t]) = \delta^{-2n-1} K_1([z,t])$, $\delta > 0$. Assume also that K_1 is smooth away from the origin. Then the distribution $K = \partial K_1/\partial z_j$ (or $K = \partial K_1/\partial \bar{z}_j$) satisfies the above conditions. Indeed, (80)–(82) are obvious by homogeneity. Again by homogeneity

$$\begin{aligned} K(\Phi^R) &= -\int K_1([z,t]) \frac{\partial}{\partial z_j}(\Phi^R)([z,t])\, dz\, dt \\ &= -\int K_1([z,t]) \frac{\partial \Phi}{\partial z_j}([z,t])\, dz\, dt, \end{aligned}$$

from which (83) follows.

(ii) Similarly, if K_2 is homogeneous of degree $-2n$ and is smooth away from the origin, then $K = \partial K_2/\partial t$ satisfies the above conditions. Note that the kernel of the Cauchy-Szegö projection is of this form. Other instances of (i) and (ii) also occur naturally; see §3.1 in Chapter 13 below.

(iii) If K is a homogeneous function of the critical degree $-2n-2$ and is smooth away from the origin, one can define a principal-value distribution that agrees with the function $K(x)$ away from the origin and satisfies (80)–(83) whenever K satisfies the appropriate mean-value property. Moreover, the examples (i) and (ii) are of this type. For further details, see §7.9.

5.2.2 We begin the proof of the theorem by making the *a priori* assumption that K has compact support not containing the origin. Then K is of course an integrable function. We will obtain bounds for the corresponding operator that depend only on the constants A_1 and A appearing in (80)–(83). Once this is done, it will be an easy matter to lift the restriction on K.

For such K, it suffices to estimate the L^2 norm of the operator (84) by testing it on f that are (say) continuous and have compact support. In this case, the formalism set forth in §3.3 for analyzing the L^2 norm of $f * K$ is completely rigorous. Let us recall it. We define f^0_λ by $f^0_\lambda(z) = \int_{\mathbf{R}} f(z,t)\, e^{-2\pi i \lambda t}\, dt$, and similarly K^0_λ by

$$K^0_\lambda(z) = \int_{\mathbf{R}} K(z,t)\, e^{-2\pi i \lambda t}\, dt. \tag{85}$$

Then by Plancherel's theorem

$$\|f * K\|^2_{L^2(\mathbf{H}^n)} = \int_{\mathbf{R}} \|(f * K)^0_\lambda\|^2_{L^2(\mathbf{C}^n)}\, d\lambda.$$

However, by (56), $(f * K)^0_\lambda(z) = (f^0_\lambda *_\lambda K^0_\lambda)(z)$.

So it would suffice to prove the boundedness of

$$F \mapsto F *_\lambda K^0_\lambda \tag{86}$$

on $L^2(\mathbf{C}^n)$, with a bound c independent of λ. As a result, we would have that

$$\|(f * K)^0_\lambda\|_{L^2(\mathbf{C}^n)} \leq c\|f^0_\lambda\|_{L^2(\mathbf{C}^n)},$$

which implies that

$$\|f * K\|^2_{L^2(\mathbf{H}^n)} \leq c^2 \int_{\mathbf{R}} \|f^0_\lambda\|^2_{L^2(\mathbf{C}^n)}\, d\lambda = c^2\|f\|^2_{L^2(\mathbf{H}^n)}.$$

To be able to apply the corollary in §5.1.4 regarding (86), we need the following lemma.

LEMMA. *Suppose K is a function on $\mathbf{H}^n$ whose support is compact and disjoint from the origin, and which satisfies (80)–(83). For each real λ, let K^0_λ be defined by (85). Then the function K^0_λ is integrable on $\mathbf{C}^n$ and satisfies the hypotheses (70) and (71) of the corollary in §5.1.4 uniformly in λ, $0 < \lambda < \infty$.*

Proof. Since

$$|K^0_\lambda(z)| \leq \int_{\mathbf{R}} |K(z,t)|\, dt \leq A_1\left(\int_{|t|\geq|z|^2} t^{-n-1}\, dt + |z|^{-2n-2}\int_{|t|<|z|^2} dt\right)$$

by (80), we have that $|K^0_\lambda(z)| \leq A_1'|z|^{-2n}$. Similarly,

$$|\nabla_z K^0_\lambda(z)| \leq \int_{\mathbf{R}} |\nabla_z K(z,t)|\, dt \leq A_1'|z|^{-2n-1}.$$

Next,

$$\begin{aligned}\int_{\mathbf{C}^n} K^0_\lambda(z)\,\phi(z/R)\, dz &= \int_{\mathbf{C}^n\times\mathbf{R}} K(z,t)\,\phi(z/R)\, e^{-2\pi i\lambda t}\, dz\, dt\\ &= R^{2n}\int_{\mathbf{C}^n\times\mathbf{R}} K(Rz,t)\,\phi(z)\, e^{-2\pi i\lambda t}\, dz\, dt\\ &= R^{2n+2}\int_{\mathbf{C}^n\times\mathbf{R}} K(Rz,R^2t)\,\phi(z)\, e^{-2\pi i\lambda^2 Rt}\, dz\, dt.\end{aligned}$$

However, for each $R > 0$, $R^{2n+2}K(Rz,R^2t)$ satisfies the same estimates (80)–(83) as $K(z,t)$ does; we see from the above that it suffices to show

$$\left|\int_{\mathbf{C}^n\times\mathbf{R}} K(z,t)\,\phi(z)\, e^{-2\pi i\lambda t}\, dz\, dt\right| \leq A', \qquad \text{for all } \lambda, \tag{87}$$

whenever ϕ is a normalized bump function on $\mathbf{C}^n$, with the bound A' depending only on the bounds A_1 and A appearing in (80)–(83).

We prove (87) by dividing consideration into two cases: first when $|\lambda| \leq 1$, second when $|\lambda| > 1$. Let $\eta(t)$ be a C^∞ function on $\mathbf{R}^1$ supported in $|t| \leq 1$, with $\eta(t) = 1$ for $|t| \leq 1/2$. Taking $|\lambda| \leq 1$, we write the integral appearing in (87) as

$$\int_{\mathbf{C}^n\times\mathbf{R}} K(z,t)\,\phi(z)\,\eta(t)\, e^{-2\pi i\lambda t}\, dz\, dt + \int_{\mathbf{C}^n\times\mathbf{R}} K(z,t)\,\phi(z)\,[1-\eta(t)]\, e^{-2\pi i\lambda t}\, dz\, dt.$$

The first integral is dealt with by writing $\Phi(z,t) = \phi(z)\eta(t)e^{-2\pi i\lambda t}$ and considering it (up to a bounded multiple) to be a bump function supported in

$$\{x \in \mathbf{H}^n : \rho(x) \leq 1\} = \{[z,t] \in \mathbf{H}^n : |z| \leq 1, |t| \leq 1\}.$$

Thus, that term can be estimated by (83), with $R = 1$. The second integral is obviously majorized by

$$\int_{\mathbf{C}^n\times\mathbf{R}} K(z,t)\,\phi(z)\,|1-\eta(t)|\, dz\, dt \leq A_1\int_{\substack{|z|\leq 1\\ |t|\geq 1/2}} [\rho(z,t)]^{-2n-2}\, dz\, dt \leq A'.$$

This disposes of (87) when $|\lambda| \leq 1$.

When $|\lambda| > 1$, we can write

$$\begin{aligned}1 &= \eta(\lambda t) + [1-\eta(\lambda t)]\\ &= \eta(\lambda t)\psi(|\lambda|^{1/2}z) + \eta(\lambda t)[1-\psi(|\lambda|^{1/2}z)] + [1-\eta(\lambda t)].\end{aligned}$$

Here ψ is a C^∞ function on $\mathbf{C}^n$ supported in $|z| \leq 1$, with $\psi(z) = 1$ for $|z| \leq 1/2$.

If we insert the factor $\eta(\lambda t)\psi(|\lambda|^{1/2}z)$ in (87), we see that

$$\phi(z)\,\eta(\lambda t)\,\psi(|\lambda|^{1/2}z)\, e^{-2\pi i\lambda t}$$

can be written in the form $\Phi(|\lambda|^{1/2}z, \lambda t)$ where, since $|\lambda| > 1$, Φ is (up to a bounded multiple) a bump function as above. Hence, the bound for the corresponding term in (87) follows from (83) with $R = |\lambda|^{-1/2}$.

The term arising in (87) when we insert the factor

$$\eta(\lambda t)[1-\psi(|\lambda|^{1/2}z)]$$

can be estimated directly, since it is supported on

$$\{[z,t] : |t| \leq |\lambda|^{-1}, |z| > |\lambda|^{-1/2}/2\}.$$

Thus, that term is majorized by

$$A_1 \int\limits_{\substack{|t|\le|\lambda|^{-1}\\ 2|z|\ge|\lambda|^{-1/2}}} [\rho(z,t)]^{-2n-2}\,dz\,dt \le A \int\limits_{2|z|\ge|\lambda|^{-1/2}} |z|^{-2n-2}\,dz \cdot \int\limits_{|t|\le|\lambda|^{-1}} dt,$$

which is bounded.

Finally, we deal with the term

$$\int_{\mathbf{C}^n\times\mathbf{R}} K(z,t)\,\phi(z)\,[1-\eta(\lambda t)]\,e^{-2\pi i\lambda t}\,dz\,dt. \tag{88}$$

Integrating (88) by parts in the t-variable allows us to replace

$$K(z,t)\,[1-\eta(\lambda t)]\,e^{2\pi i\lambda t} \qquad \text{by} \qquad \frac{-1}{2\pi i\lambda}\frac{\partial}{\partial t}(K(z,t)[1-\eta(\lambda t)]);$$

then (80) and (82) show that the integral is majorized by

$$\frac{cA_1}{|\lambda|}\int_{2|t|\ge|\lambda|^{-1}}\left(\int_{\mathbf{C}^n}\rho(z,t)^{-2n-4}\,dz\right)dt \le A',$$

since, as is easily seen,

$$\int_{\mathbf{C}^n}[\rho(z,t)]^{-2n-4}\,dz = c/t^2.$$

The proof of the lemma is therefore complete.

5.2.3 We can now finish the proof of the theorem. Because of the lemma just proved and §5.1.4, we have that the operators (86) are bounded on $L^2(\mathbf{C}^n)$, with a bound depending only on the constants appearing in (80)–(83). This proves our desired conclusion under the additional assumption that K has compact support that is disjoint from the origin. We now show how this restriction can be removed.

For each $0<\varepsilon<N<\infty$, let

$$K_{\varepsilon,N}(z,t) = K(z,t)\,\Psi(z/N,t/N^2)\,[1-\Psi(z/\varepsilon,t/\varepsilon^2)]$$

where Ψ is a fixed smooth function on $\mathbf{H}^n$, supported in the unit ball $\rho(z,t)\le 1$, with $\Phi(z,t)=1$ for $\rho(z,t)\le 1/2$. Note that $K_{\varepsilon,N}$ satisfies the assumptions (80)–(83) uniformly in ε and N and, moreover, that $K_{\varepsilon,N}$ satisfies the support restriction used above.

Now the condition (83), which the $K_{\varepsilon,N}$ satisfy uniformly, then states that the family $\{K_{\varepsilon,N}\}$ is a bounded set of distributions. Thus, there exist (ε_j,N_j) with $\varepsilon_j\to 0$ and $N_j\to\infty$, so that K_{ε_j,N_j} converges (in the sense of distributions) to a distribution K' as $\varepsilon_j\to 0$ and $N_j\to\infty$. As is easily verified, K' also satisfies conditions (80)–(83), and equals the function $K(x)$ away from the origin. Moreover, convolution with K' is a bounded operator on $L^2(\mathbf{H}^n)$, since the operators $f\mapsto f*K_{\varepsilon,N}$ are uniformly bounded.

Finally, since both K and K' agree with the same function away from the origin, their difference $K-K'$ is a distribution that is supported at the origin and hence is a linear combination of the Dirac delta function and finitely many of its derivatives. Therefore the analogue of (83) holds, replacing K with $K-K'$. By using appropriate Φ, one notes that, in fact, no derivatives of the delta function appear in $K-K'$; i.e., $K-K'$ is a constant multiple of the delta function. Thus convolution with $K-K'$ is a constant multiple of the identity operator and $f\mapsto f*K$ is bounded on $L^2(\mathbf{H}^n)$, completing the proof of the theorem.

5.2.4 Corollary. *The operator $Tf=f*K$ treated in Theorem 3 above has an extension to a bounded operator from $L^p(\mathbf{H}^n)$ to itself, for every p with $1<p<\infty$.*

The proof follows much the same lines as the argument given in §2.5 for the Cauchy-Szegö projection, except that in that case the L^2 boundedness was a consequence of the fact that the operator was an orthogonal projection, while in the present case it is a consequence of Theorem 3.

What needs to be verified is that the kernel $K(x,y)=K(y^{-1}\cdot x)$ satisfies the inequalities (32) and (33). Since $V(x,y)=c\rho(x,y)^{2n+2}$, and $\rho(x,y)=\rho(y^{-1}\cdot x)$, these amount to the two inequalities

$$|K(x)|\le A[\rho(x)]^{-2n-2}, \tag{89}$$

and

$$|K(x)-K(u\cdot x)|\le A\rho(u)\cdot[\rho(x)]^{-2n-3}, \qquad \text{if } \rho(x)\ge\bar{c}\rho(u), \tag{90}$$

where $\bar{c}$ is an appropriate (large) constant.

The first of these is the hypothesis (80). For the second we may re-scale, using the fact that $\rho(\delta\circ x)=\delta\rho(x)$ and noting that if $K(x)$ satisfies (80)–(82), then so does $\delta^{2n+2}K(\delta\circ x)$, with the same bounds. Thus the second inequality is reduced to the case $\rho(x)=1$, $\rho(u)\le 1/\bar{c}$ and holds because K is assumed have bounded first derivatives in any compact set that excludes the origin.

Finally, observe that while the kernel $K(x,y)$ is not, in general, formally self-adjoint, T^* is given by convolution with the kernel

$$K^*(y,x)=K^*(y^{-1}\cdot x), \qquad K^*(x)=\bar{K}(x^{-1}).$$

Now K^* satisfies the same estimates as K, because

$$K^*(x)=K^*([\zeta,t])=\bar{K}([-\zeta,-t]).$$

The proof of the corollary is therefore concluded.

5.2.5 There is a different proof of Theorem 3, which in addition allows it to be formulated for a large class of groups. It will be given in the next chapter. The present proof, however, has the virtue of arising naturally from the structure of the Heisenberg group, and it gives us a glimpse of the further significance of oscillatory integrals in making L^2 estimates. See also §7.10 below.

6. Appendix: Representations of the Heisenberg group

We sketch here the basic facts concerning the representation theory of the Heisenberg group. Further details will be found in the cited literature.

6.1 For each real λ, we define a mapping R^λ from the Heisenberg group $\mathbf{H}^n$ to the group of unitary operators on $L^2(\mathbf{R}^n)$ as follows. For $[\zeta, t] \in \mathbf{H}^n$, $\zeta = u + iv \in \mathbf{C}^n$, $t \in \mathbf{R}$, and $\phi \in L^2(\mathbf{R}^n)$, we have

$$[R^\lambda(\zeta,t)\phi](x) = e^{2\pi i\lambda[u\cdot x + u\cdot v/2 + t/4]}\phi(x+v). \tag{91}$$

Observe that (53) corresponds to the case $\lambda = 1$. As in that special case, we see that each operator $R^\lambda(\zeta,t)$ is unitary on $L^2(\mathbf{R}^n)$ and that, for each λ, the mapping

$$[\zeta,t] \mapsto R^\lambda(\zeta,t)$$

is a homomorphism from $\mathbf{H}^n$ to the group of unitary operators on $L^2(\mathbf{R}^n)$. Moreover it is continuous, in the sense that

$$\|R^\lambda(\zeta,t)\phi - \phi\|_{L^2} \to 0 \quad \text{as} \quad [\zeta,t] \to 0, \qquad \text{for } \phi \in L^2.$$

Thus R^λ defines a *unitary representation* of $\mathbf{H}^n$.

We observe next that, whenever $\lambda \neq 0$, this representation is *irreducible*; in other words, that there are no nontrivial subspaces of $L^2(\mathbf{R}^n)$ that are invariant under the induced action of $\mathbf{H}^n$. Since R^λ is unitary, this is equivalent to the following assertion: If A is a bounded operator on $L^2(\mathbf{R}^n)$ that commutes with every $R^\lambda(\zeta,t)$, as $[\zeta,t]$ varies over $\mathbf{H}^n$, then A is a constant multiple of the identity.

Indeed, if A is such an operator then, specializing (91) to the case $u = 0$, $t = 0$, we see that A commutes with translations on $\mathbf{R}^n$. Therefore A is given by a Fourier multiplier; i.e., there is a bounded function a on $\mathbf{R}^n$ so that

$$\widehat{A\phi}(\xi) = a(\xi)\widehat{\phi}(\xi),$$

for $\phi \in L^2(\mathbf{R}^n)$.‡ Next, using (91) with $v = 0$, $t = 0$, we see that $a(\xi+u) = a(\xi)$, for all $u \in \mathbf{R}^n$, so a is a constant function. Thus, A is a constant multiple of the identity and the representation R^λ is irreducible.

‡ A proof may be found in *Fourier Analysis*, Chapter 1, §3.

6.2 The next issue we address is the classification of the irreducible unitary representations of $\mathbf{H}^n$. To describe this, recall the equivalence relation: two representations R_1, R_2 are said to be *equivalent* if there is a unitary operator U so that

$$U \circ R_1(\zeta,t) = R_2(\zeta,t) \circ U$$

for all $[\zeta,t] \in \mathbf{H}^n$.

(a) Two representations R^{λ_1}, R^{λ_2} of the type described above are equivalent only when $\lambda_1 = \lambda_2$. To see this, restrict (91) to $[\zeta,t]$ with $\zeta = 0$. The resulting operators are scalar multiples of the identity, namely

$$R^{\lambda_1}(0,t) = e^{\pi i\lambda_1 t/2}I, \qquad \text{and} \quad R^{\lambda_2}(0,t) = e^{\pi i\lambda_2 t/2}I.$$

If the representations are equivalent, we have $\lambda_1 = \lambda_2$.

(b) Besides the (infinite-dimensional) representations R^λ above, there are the "trivial" one-dimensional representations, indexed by pairs $(\alpha,\beta) \in \mathbf{R}^n \times \mathbf{R}^n$, given by multiplication operators

$$r_{\alpha,\beta}(\zeta,t) = e^{i(\alpha\cdot u + \beta\cdot v)}, \qquad \zeta = u + iv.$$

These representations are (vacuously) irreducible, and are clearly inequivalent as (α,β) vary.

(c) A well-known theorem of Stone and von Neumann, which we do not prove here, states that, up to unitary equivalence, the families $\{R^\lambda\}$, $\{r_{\alpha,\beta}\}$ comprise the totality of irreducible unitary representations of the Heisenberg group.†

6.3 Group Fourier transform. The representations of the Heisenberg group, in particular those appearing in §6.1, allow us to study $L^2(\mathbf{H}^n)$ via the *group Fourier transform*. Starting with a function $f \in L^1(\mathbf{H}^n)$ (say), we define an operator-valued function $\lambda \mapsto \widehat{f}_{\mathrm{H}}(\lambda)$ on $\mathbf{R}^1$, given by

$$\widehat{f}_{\mathrm{H}}(\lambda) = \int_{\mathbf{H}^n} R^\lambda(\zeta,t)\, f(\zeta,t)\, d\zeta\, dt. \tag{92}$$

We see that $\widehat{f}_{\mathrm{H}}(\lambda)$ is (for each $\lambda \in \mathbf{R}$) a bounded operator on $L^2(\mathbf{R}^n)$, either by the theory of Banach-space valued integration, or more directly by noting that the operator $\widehat{f}_{\mathrm{H}}$ is determined by

$$\langle \widehat{f}_{\mathrm{H}}(\lambda)\phi, \psi\rangle = \int_{\mathbf{H}^n} \langle R^\lambda(\zeta,t)\phi,\psi\rangle f(\zeta,t)\, d\zeta\, dt,$$

for any pair $\phi, \psi \in L^2(\mathbf{R}^n)$.

Some of the properties we are accustomed to (for the standard Fourier transform) continue to hold in this setting. First, if $f, g \in L^1(\mathbf{H}^n)$ then

$$(f * g)_{\mathrm{H}}\widehat{\ }(\lambda) = \widehat{f}_{\mathrm{H}}(\lambda)\cdot \widehat{g}_{\mathrm{H}}(\lambda), \tag{93}$$

† For a proof see, e.g., Folland [1989].

where the multiplication on the right is composition of operators.

To check this, note that with $x, y \in \mathbf{H}^n$, we have

$$(f * g)(x) = \int_{\mathbf{H}^n} f(y)\, g(y^{-1} \cdot x)\, dy,$$

and so

$$(f * g)_{\mathrm{H}}\widehat{\ }(\lambda) = \int_{\mathbf{H}^n \times \mathbf{H}^n} R^\lambda(y)\, f(y)\, R^\lambda(y^{-1} \cdot x)\, g(y^{-1} \cdot x)\, dy\, dx,$$

which, after a change of variables, gives (93). These formal manipulations are easily justified.

Observe next that if f is a test function and $\lambda \neq 0$, then the operator $\widehat{f}_{\mathrm{H}}(\lambda)$ is actually a Hilbert-Schmidt operator, whose kernel is given by

$$K_\lambda(x, y) = \tilde{f}(-\lambda(y + x)/2, y - x, -\lambda/4), \tag{94}$$

where $\tilde{f}(\xi, v, \tau)$ is the usual Fourier transform in the first and third variables:

$$\tilde{f}(\xi, v, \tau) = \int_{\mathbf{R}^n \times \mathbf{R}^n} e^{-2\pi i(\xi \cdot u + \tau t)}\, f(u, v, t)\, du\, dt. \tag{95}$$

In fact, looking back at the definitions (91) and (92), we see that

$$[\widehat{f}_{\mathrm{H}}(\lambda)\phi](x) = \int_{\mathbf{H}^n} e^{2\pi i \lambda[u \cdot x + u \cdot v/2 + t/4]}\, \phi(x + v)\, f(u, v, t)\, du\, dv\, dt.$$

Making the change of variables $x + v = y$, $v = y - x$ then gives

$$[\widehat{f}_{\mathrm{H}}(\lambda)\phi](x) = \int_{\mathbf{R}^n} K_\lambda(x, y)\, \phi(y)\, dy,$$

with

$$K_\lambda(x, y) = \int_{\mathbf{R}^n \times \mathbf{R}^n} e^{2\pi i \lambda(u \cdot (y+x)/2 + t/4)}\, f(u, y - x, t)\, du\, dt.$$

This establishes (94). With it, we get the basic identity expressing the group version of Plancherel's theorem for $\mathbf{H}^n$, namely that if f is a test function, then

$$\int_{\mathbf{H}^n} |f(\zeta, t)|^2\, d\zeta\, dt = \frac{1}{4} \int_{\mathbf{R}} \|\widehat{f}_{\mathrm{H}}(\lambda)\|_{\mathrm{HS}}^2\, |\lambda|^n\, d\lambda. \tag{96}$$

Indeed, by (94),

$$\begin{aligned} \|\widehat{f}_{\mathrm{H}}(\lambda)\|_{\mathrm{HS}}^2 &= \int_{\mathbf{R}^n \times \mathbf{R}^n} |K_\lambda(x, y)|^2\, dx\, dy \\ &= \int_{\mathbf{R}^n \times \mathbf{R}^n} |\tilde{f}(-\lambda(x + y)/2, y - x, -\lambda/4)|^2\, dx\, dy \\ &= |\lambda|^{-n} \int_{\mathbf{R}^n \times \mathbf{R}^n} |\tilde{f}(x, y, -\lambda/4)|^2\, dx\, dy. \end{aligned}$$

Therefore an integration in λ, together with the usual Plancherel formula for $\mathbf{R}^n \times \mathbf{R}^n$ applied to (95), yields (96). An additional limiting argument then proves the following.

PROPOSITION. *The group Fourier transform*

$$f \mapsto \widehat{f}_{\mathrm{H}}$$

extends to a unitary mapping from $L^2(\mathbf{H}^n)$ to the Hilbert space of Hilbert-Schmidt valued functions; the norm on the latter space is given by the square root of the right side of (96).

An important formal consequence of the above considerations, which can actually be implemented in appropriate circumstances, is the fact that a convolution operator $f \mapsto f * K$ can be realized on the "Fourier transform side" as a multiplier operator

$$\widehat{f}_{\mathrm{H}}(\lambda) \mapsto \widehat{f}_{\mathrm{H}}(\lambda) \cdot \widehat{K}_{\mathrm{H}}(\lambda).$$

Moreover, the boundedness of the convolution operator on $L^2(\mathbf{H}^n)$ is equivalent to the uniform boundedness of the operator norms of the $\widehat{K}_{\mathrm{H}}(\lambda)$ over $\lambda \neq 0$.

6.4 Hermite expansions. The deeper study of Fourier analysis on the Heisenberg group is intimately connected with the Hermite expansion in $\mathbf{R}^n$. We begin be considering the situation for $n = 1$.

The Hermite functions $H_k(x)$ are defined by the formula

$$H_k(x) = (-1)^k e^{x^2/2} \left(\frac{d}{dx}\right)^k e^{-x^2}, \qquad k = 0, 1, 2, \ldots, \tag{97}$$

from which it is clear that $H_k(x)$ is a polynomial of degree k in x times the function $e^{-x^2/2}$. A useful alternative definition arises from the generating relation

$$\sum_{k=0}^{\infty} H_k(x) \frac{\mu^k}{k!} = e^{-(x^2/2 - 2\mu x + \mu^2)}. \tag{98}$$

Indeed, notice that $e^{-(x^2/2 - 2\mu x + \mu^2)} = e^{x^2/2}\, e^{-(x-\mu)^2}$. The formula for the Taylor expansion of $\mu \mapsto e^{-(x-\mu)^2}$ about $\mu = 0$ then gives (98), since

$$\left(\frac{d}{d\mu}\right)^k e^{-(x-\mu)^2} = (-1)^k \left(\frac{d}{dx}\right)^k e^{-(x-\mu)^2}.$$

We note that the H_k are eigenfunctions of the Hermite operator

$$L = \frac{-d^2}{dx^2} + x^2.$$

In fact

$$LH_k = (2k + 1)H_k, \qquad k = 0, 1, 2, \ldots. \tag{99}$$

To verify (99), note that

$$\begin{aligned} L[e^{-(x^2/2 - 2\mu x + \mu^2)}] &= \left[x^2 - \frac{d^2}{dx^2}\right] e^{-(x^2/2 - 2\mu x + \mu^2)} \\ &= [1 + 4x\mu - 4\mu^2] e^{-(x^2/2 - 2\mu x + \mu^2)}. \end{aligned}$$

However, $(d/d\mu) e^{x^2/2 - (x-\mu)^2} = (-2\mu + 2x) e^{x^2/2 - (x-\mu)^2}$; therefore

$$L(e^{x^2/2}\, e^{-(x-\mu)^2}) = \left(2\mu \frac{d}{d\mu} + 1\right) e^{x^2/2}\, e^{-(x-\mu)^2}.$$

Applying this term-by-term to the generating relation (98) gives us (99).

Our next observation is that the H_k are mutually orthogonal; i.e.,

$$\int_{\mathbf{R}} H_k(x)\, H_j(x)\, dx = 0, \qquad \text{if } j \neq k. \tag{100}$$

Indeed, using (99), we see that $(2k+1)\langle H_k, H_j\rangle = \langle LH_k, H_j\rangle$, if we write

$$\langle f, g\rangle = \int_{\mathbf{R}} f(x)\, g(x)\, dx.$$

Since the H_k are in $\mathcal{S}$, we may integrate by parts and obtain

$$\langle LH_k, H_j\rangle = \langle H_k, LH_j\rangle = (2k+1)\langle H_k, H_j\rangle = (2j+1)\langle H_k, H_j\rangle;$$

this proves (100).

To normalize the H_k, one uses the identity

$$\int_{\mathbf{R}} [H_k(x)]^2\, dx = \pi^{1/2} 2^k k!. \tag{101}$$

To see this, square the identity (98), integrate in x, and apply (100). The result is

$$\sum_{k=0}^{\infty} \int_{\mathbf{R}} [H_k(x)]^2\, dx \frac{\mu^{2k}}{(k!)^2} = \int_{\mathbf{R}} e^{x^2 - 2(x-\mu)^2}\, dx = e^{2\mu^2}\pi^{1/2}.$$

However, $e^{2\mu^2} = \sum_{k=0}^{\infty} 2^k \mu^{2k}/k!$, and comparing terms gives us (101).

Finally, we point out that the $\{H_k\}$ are complete in $L^2(\mathbf{R}^1)$. This is equivalent to the following: If $f \in L^2(\mathbf{R}^1)$ and

$$\int_{\mathbf{R}^1} f(x)\, H_k(x)\, dx = 0, \qquad \text{for all } k,$$

then $f = 0$. So let f be such a function. Because of (98), our assumption on f is the same as the assertion

$$I(\mu) = \int_{\mathbf{R}^1} f(x)\, e^{-x^2/2}\, e^{2\mu x}\, dx = 0. \tag{102}$$

Now $I(\mu)$ is clearly an entire function of μ so, in particular, (102) holds for all imaginary μ. The Fourier inversion formula applied to $f(x)e^{-x^2/2}$ then shows that f is identically zero, proving our assertion. Note, as a corollary, that the H_k give a complete eigenfunction expansion of the operator L.

6.4.1 The passage to n dimensions is straightforward. Define

$$L = -\Delta + |x|^2.$$

For each n-tuple of nonnegative integers $k = (k_1, \ldots, k_n)$, let

$$H_k(x) = H_{k_1}(x_1) \cdot H_{k_2}(x_2) \cdots H_{k_n}(x_n), \qquad x = (x_1, \ldots, x_n) \in \mathbf{R}^n.$$

An easy application of the one-dimensional case shows that the H_k are orthogonal and complete in $L^2(\mathbf{R}^n)$; they can be explicitly normalized by (101). Finally, they are eigenfunctions of L,

$$LH_k = (2|k| + n)H_k, \tag{103}$$

with $|k| = k_1 + \cdots + k_n$, and they give the complete eigenfunction expansion of the operator L. In particular, because of (103), the operator $L - \beta I$ is invertible exactly when $\beta \neq n, n+2, n+4, \ldots$.

6.5 An example. A typical example of the application of Hermite expansions to Fourier analysis on $\mathbf{H}^n$ arises when one examines the Cauchy-Szegö operator in terms of the group Fourier transform. We describe here the formal derivation of the result, leaving the rigorous details to the cited literature.

Recall the Cauchy-Szegö projection $f \mapsto C(f) = f * K$; it can be viewed as the orthogonal projection to the null space of the tangential Cauchy-Riemann operators $\bar{Z}_j$, $j = 1, \ldots, n$ (see also Chapter 13, §7.16). Now

$$\bar{Z}_j(f)_{\mathbf{H}}\widehat{\ }(\lambda) = \hat{f}_{\mathbf{H}}(\lambda) \cdot \widehat{\bar{Z}_j}(\lambda),$$

where $\widehat{\bar{Z}_j}(\lambda) = 1/2(\widehat{X}_j(\lambda) + i\widehat{Y}_j(\lambda))$, and (see (91))

$$\widehat{X}_j(\lambda) = 2\pi i\lambda x_j, \qquad \widehat{Y}_j(\lambda) = \frac{\partial}{\partial x_j}, \qquad j = 1, \ldots, n. \tag{104}$$

For each λ, if ϕ is annihilated by $\widehat{X}_j(\lambda) + i\widehat{Y}_j(\lambda)$, $j = 1, \ldots, n$, then

$$-\sum_{j=1}^{n} (\widehat{X}_j(\lambda) - i\widehat{Y}_j(\lambda))(\widehat{X}_j(\lambda) + i\widehat{Y}_j(\lambda))\phi = 0.$$

However the sum above is

$$\left(\sum_{j=1}^{n} \widehat{X}_j(\lambda)^2 + \widehat{Y}_j(\lambda)^2 + i[\widehat{X}_j(\lambda), \widehat{Y}_j(\lambda)]\right)\phi,$$

which, in view of (104), gives the relation

$$(4\pi^2\lambda^2|x|^2 - \Delta - 2\pi\lambda n)\phi = 0. \tag{105}$$

To understand the implications of (105) let, for each $\lambda \neq 0$, $\{H_k^\lambda\}$ be the orthogonal basis of $L^2(\mathbf{R}^n)$ defined by

$$H_k^\lambda(x) = H_k((2\pi|\lambda|)^{1/2} x).$$

Observe next that

$$(4\pi^2\lambda^2|x|^2 - \Delta - 2\pi\lambda n)H_k^\lambda(x) = 2\pi|\lambda|\{[LH_k - \operatorname{sign}(\lambda) n H_k]((2\pi|\lambda|)^{1/2} x)\}.$$

Thus, by (103), a solution of (105) in L^2 is possible only if $\lambda > 0$, and then

$$\phi(x) = cH_0((2\pi|\lambda|)^{1/2} x) = ce^{-\pi\lambda|x|^2}.$$

We can summarize this as follows.

PROPOSITION. *Let K be the convolution kernel of the Cauchy-Szegö projection operator on $\mathbf{H}^n$, given by (29). Then*

(a) *When $\lambda < 0$, then $\widehat{K}_{\mathbf{H}}(\lambda) = 0$.*

(b) *When $\lambda > 0$, then $\widehat{K}_{\mathbf{H}}(\lambda)$ is the orthogonal projection to the one-dimensional subspace of $L^2(\mathbf{R}^n)$ consisting of all constant multiples of $e^{-\pi\lambda|x|^2}$.*

Related results may be found in Ogden and Vagi [1972], Geller [1977].

The monograph of Thangavelu [1993] deals with a variety of topics concerning Hermite expansions.

7. Further results

A. Fractional linear transformations and the Cauchy-Szegö kernel

7.1 We describe the full group of biholomorphic mappings of the domain $\mathcal{U}^n$; to do this, we deal also with the (equivalent) unit ball $\mathbf{B} \subset \mathbf{C}^{n+1}$.

(a) To begin with, let H be the Hermitian quadratic form

$$H(\zeta) = |\zeta_{n+2}|^2 - \sum_{j=1}^{n+1} |\zeta_j|^2, \qquad \text{for } \zeta = (\zeta_1, \ldots, \zeta_{n+2}) \in \mathbf{C}^{n+2}.$$

Consider the group $G_0 = \mathrm{SU}(n+1,1)$ of complex linear transformations of $\mathbf{C}^{n+2}$, having determinant 1, that preserve $H(\zeta)$. Notice that the elements of G_0 map the set $\{\zeta \in \mathbf{C}^{n+2} : H(\zeta) > 0\}$ to itself. Passing to projective space via $w_j = \zeta_j/\zeta_{n+2}$, $j = 1, \ldots, n+1$, we see that each element of G_0 induces a *fractional linear* transformation of the unit ball

$$\mathbf{B} = \{w = (w_1, \ldots, w_{n+1}) \in \mathbf{C}^{n+1} : \sum_{j=1}^{n+1} |w_j|^2 < 1\}$$

to itself.

If the matrix realization of an element $g \in G_0 = \mathrm{SU}(n+1,1)$ is given by $g = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$, where the submatrices A, B, C, and D have sizes $(n+1) \times (n+1)$, $(n+1) \times 1$, $1 \times (n+1)$, and 1×1, respectively, then the induced fractional linear transformation is

$$g(w) = \frac{Aw + B}{Cw + D}, \tag{$*$}$$

where the vector w is thought of as an $(n+1) \times 1$ matrix. The requirement that g preserves H is equivalent to the relations

$$A^*A - C^*C = I, \qquad |D|^2 - B^*B = 1, \qquad A^*B = C^*D;$$

here I is the $(n+1) \times (n+1)$ identity matrix. G_0 is then the group of matrices $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$ of determinant 1 whose submatrices enjoy the above relations.

(b) It is easily seen that G_0 acts transitively on the unit ball. Next let $K_0 \subset G_0$ be the subgroup that fixes the origin; K_0 is isomorphic to (and acts identically to) the unitary group on $\mathbf{C}^{n+1}$. Thus the unit ball $\mathbf{B}$ can be identified with the homogeneous space G_0/K_0. It can also be seen that every holomorphic automorphism f of $\mathbf{B}$ is represented by an element of G_0: first choose a $g \in G_0$ so that $f_1 = g^{-1} \circ f$ fixes the origin and $f_1'(0) = I$; a lemma of H. Cartan then implies that $f_1 = I$.

(c) We now use the mapping $\gamma : \mathbf{B} \to \mathcal{U}^n$, $\gamma(w) = z$, given by (3). While of course $\gamma \notin G_0$, it is of the form ($*$) with A the diagonal matrix with entries $(1, \ldots, 1, -i)$, $B^* = (0, \ldots, 0, -1)$, $C = (0, \ldots, 0, 1)$, and $D = 1$.

Let $G = \gamma G_0 \gamma^{-1}$ be the corresponding group of holomorphic automorphisms of $\mathcal{U}^n$. Note that $K = \gamma K_0 \gamma^{-1}$ is the subgroup that fixes the point $\mathbf{i} = (0, \ldots, 0, i)$. Another important subgroup of G consists of the affine-linear mappings in G; in the standard notation for semisimple Lie groups, it is a product of three further subgroups A, M, and N. Here A is the subgroup of "dilations" (see §1.3), M is the subgroup of "rotations" (or equivalently those elements of K that commute with every element of A, see §1.3), and N is the Heisenberg group (see §1.4). One also has the Iwasawa decomposition $G = KAN$.

Further details are in Siegel [1950], Hua [1963], Piatetski-Shapiro [1961]; see also Stein [1972], Helgason [1984]. Cartan's lemma can be found in, e.g., Krantz [1982b].

7.2 The use of the fractional linear transformations above allow one to give an alternate derivation of the Cauchy-Szegö kernel.

(a) Consider first the case of the unit ball $\mathbf{B} \subset \mathbf{C}^{n+1}$. Writing $d\sigma$ for the normalized surface measure on the unit sphere, one has

$$F(w) = \int_{|u|=1} S_{\mathbf{B}}(w,u)\, F(u)\, d\sigma(u), \qquad F \in \mathcal{H}^2(\mathbf{B}).$$

Here $\mathcal{H}^2(\mathbf{B})$ is the analogue of the holomorphic Hardy space $\mathcal{H}^2(\mathcal{U}^n)$ considered in §2.1. It consists of those F, holomorphic on $\mathbf{B}$, for which

$$\|F\|_{\mathcal{H}^2(\mathbf{B})} = \left(\sup_{r<1} \int_{|u|=1} |F(ru)|^2\, d\sigma(u)\right)^{1/2} < \infty.$$

The Cauchy-Szegö kernel $S_{\mathbf{B}}$ is given by

$$S_{\mathbf{B}}(w,u) = (1 - \langle w, u\rangle)^{-n-1}, \qquad \langle w, u\rangle = \sum_{j=1}^{n+1} w_j \bar{u}_j.$$

The above reproducing formula may be proved by first checking it for $w = 0$, in which case it is a direct consequence of the mean-value property of harmonic functions. To derive it for $w \neq 0$, one uses the uniqueness of $S_{\mathbf{B}}$, together with the transitivity of the automorphism group G_0 described in §7.1 above. Indeed, if $g \in G_0$, we have

$$S_{\mathbf{B}}(w,u) = S_{\mathbf{B}}(g(w), g(u))\, (Cw + D)^{-n-1}\, (\overline{Cu + D})^{-n-1}.$$

(b) One computes next, using the mapping $\gamma : \mathbf{B} \to \mathcal{U}^n$ (given by (3)), that

$$d\sigma(u) = |J(z)|^2\, d\beta(z), \qquad z = \gamma(u), \qquad J(z) = c_n^{1/2} 2^{n+1} (1 - iz_{n+1})^{-n-1}.$$

As a result, if $\Gamma : F \mapsto \tilde{F}$, where $\tilde{F}(z) = F(\gamma^{-1}(z)) \cdot J(z)$ for $z \in \mathcal{U}^n$, then Γ gives a unitary correspondence between $\mathcal{H}^2(\mathbf{B})$ and $\mathcal{H}^2(\mathcal{U}^n)$. It follows that

$$S_{\mathbf{B}}(w,u) = S(\gamma(w), \gamma(u))\, J^{-1}(\gamma(w))\, \overline{J^{-1}(\gamma(u))},$$

so that S can be determined from $S_{\mathbf{B}}$.

The argument in (a) is in Hua [1963]; see also Stein [1972], Rudin [1980]. For (b), see Chapter 16 of Zygmund [1959] for the case $n = 0$, Korányi and Vági [1971] for general n.

7.3 We mention two further derivations of the Cauchy-Szegö kernel.

(a) The first arises by taking the Fourier transform in the $t = \operatorname{Re}(z_{n+1})$ variable. This gives an isomorphism between the space $\mathcal{H}^2(\mathcal{U}^n)$ and the Hilbert space of measurable functions $f(z,\lambda)$ on $\mathbf{C}^n \times \mathbf{R}_+$ that, for a.e. λ, are entire in $z \in \mathbf{C}^n$, and for which the norm

$$\left(\int_0^\infty \int_{\mathbf{C}^n} |f(z,\lambda)|^2 e^{-4\pi\lambda|z|^2}\, dz\, d\lambda\right)^{1/2}$$

is finite.

This approach represents the original derivation of the Cauchy-Szegö kernel on $\mathcal{U}^n$; see Gindikin [1964].

(b) A second method is based on the same idea as was used in §2.2 to reduce matters to the case of one complex variable. In fact, it is not difficult to see that one can obtain (19) as a consequence of the following one-variable reproducing identity

$$f(z) = \frac{n(2i)^n}{2\pi i} \int_{\operatorname{Im}(w)>0} \frac{f(w)\,(\operatorname{Im} w)^{n-1}}{(\overline{w} - z)^{n+1}}\, dw,$$

for integers $n > 0$ and appropriate f that are holomorphic in the upper half-plane; here dw is Lebesgue measure on $\mathbf{C}^1$. Nagel and Stein [1979].

7.4 We complete the calculation in §2.3.3 and show that

$$\int_{\zeta\in\mathbf{C}^n} \int_{t\in\mathbf{R}} \frac{dt\, d\zeta}{(t^2 + (|\zeta|^2+1)^2)^{n+1}} = \frac{4^{-n}\pi^{n+1}}{n!}$$

A simple scaling argument (which involves carrying out the t-integration first) shows that our integral equals the product $\alpha_n \cdot \beta_n \cdot \gamma_n$, where

$$\alpha_n = \int_{-\infty}^{\infty} \frac{dt}{(1+t^2)^{n+1}},$$

$$\beta_n = \int_0^\infty \frac{r^{2n-1}}{(1+r^2)^{2n+1}}\, dr,$$

and γ_n is the volume of the unit sphere in $\mathbf{C}^n$.

A classical identity of Euler for the Beta function states:

$$\int_0^\infty (1+u^2)^{-a} u^{2b-1}\, du = \frac{\Gamma(b)\Gamma(a-b)}{2\Gamma(a)},$$

if $a > b$. Thus

$$\alpha_n = \frac{\Gamma(1/2)\,\Gamma(n+1/2)}{\Gamma(n+1)} \qquad \text{and} \qquad \beta_n = \frac{\Gamma(n)\,\Gamma(n+1)}{2\cdot\Gamma(2n+1)}.$$

Now (as is well known) the volume of the unit sphere in $\mathbf{C}^n$ is $\gamma_n = 2\pi^n/\Gamma(n)$. Carrying out the indicated multiplication, using the "duplication formula"

$$\Gamma(1/2)\Gamma(2n+1) = 4^n\Gamma(n+1/2)\Gamma(n+1),$$

gives

$$\frac{4^{-n}\pi^{n+1}}{\Gamma(n+1)}$$

as the value of our integral.

B. Symplectic invariance

7.5 We now discuss the symplectic invariance of the notions attached to the Heisenberg group, and in particular the pseudo-differential correspondence (52).

We consider the space $\mathbf{R}^n \times \mathbf{R}^n = \{(x,\xi) : x \in \mathbf{R}^n, \xi \in \mathbf{R}^n\}$ and the nondegenerate antisymmetric bilinear form $\langle , \rangle$ on $\mathbf{R}^n \times \mathbf{R}^n$ given by

$$\langle z, z' \rangle = 2\operatorname{Im}(z\cdot\overline{z'}) = 2\sum_{j=1}^n x_j'\xi_j - x_j\xi_j',$$

where we have written $z = x + i\xi$, $z' = x' + i\xi'$. We say that a real linear transformation $\sigma : \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}^n \times \mathbf{R}^n$ is *symplectic* if σ preserves $\langle , \rangle$, that is, when $\langle \sigma(z), \sigma(z') \rangle = \langle z, z' \rangle$ for all $z, z' \in \mathbf{C}^n \approx \mathbf{R}^n \times \mathbf{R}^n$. Each linear $\sigma : \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}^n \times \mathbf{R}^n$ can be written in block form $\sigma = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$, where A, B, C, and D are real $n \times n$ matrices; σ is symplectic exactly when $CD^t = DC^t$, $AB^t = BA^t$, and $AD^t - BC^t = I$. We write $\operatorname{Sp}(n,\mathbf{R})$ for the group of symplectic transformations.

The intimate connection between $\operatorname{Sp}(n,\mathbf{R})$ and the Heisenberg group is due to the fact that the form $\langle , \rangle$ arises in the group law (6) and hence the mapping $[z,t] \mapsto [\sigma(z),t]$ is an *automorphism* of the Heisenberg group whenever $\sigma \in \operatorname{Sp}(n,\mathbf{R})$. As a result, if $[z,t] \to R^\lambda(z,t)$ is a representation of $\mathbf{H}^n$ (as in §6) then $[z,t] \to R^\lambda(\sigma(z),t)$ is another representation; since they are the same on the center $\{[0,t] : t \in \mathbf{R}\}$, these two representations are equivalent (see §6.2). We restrict our attention to $\lambda = 1$ and $t = 0$, and thus (by §3.2.2) to the operators (see (51))

$$[W(u,v)f](x) = e^{2\pi i x\cdot u}\, e^{\pi i u\cdot v}\, f(x+v)$$

acting on $f \in L^2(\mathbf{R}^n)$. It will be convenient to restate the above for $\tilde{\sigma} = (\sigma^t)^{-1}$ instead of σ ($\tilde{\sigma}$ also belongs to $\operatorname{Sp}(n,\mathbf{R})$). The assertion is that for each $\sigma \in \operatorname{Sp}(n,\mathbf{R})$, there is a unitary operator U_σ on $L^2(\mathbf{R}^n)$ so that

$$W(\tilde{\sigma}(u,v)) = U_\sigma \circ W(u,v) \circ U_\sigma^{-1}. \tag{$*$}$$

(b) By the irreducibility of the representations R^λ, each operator U_σ is actually uniquely characterized by $(*)$ up to a constant multiple. For the following $\sigma \in \operatorname{Sp}(n,\mathbf{R})$, it is easy to determine U_σ and check that they indeed satisfy $(*)$.

(i) If $\sigma = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$, then U_σ is the Fourier transform $U_\sigma f = \widehat{f}$.

(ii) If $\sigma = \begin{pmatrix} I & 0 \\ C & I \end{pmatrix}$, then C is symmetric and $(U_\sigma f)(x) = e^{\pi i Cx\cdot x} f(x)$.

(iii) If $\sigma = \begin{pmatrix} A & 0 \\ 0 & (A^t)^{-1} \end{pmatrix}$, then $(U_\sigma f)(x) = f(A^{-1}x)|\det A|^{-1/2}$.

(c) Furthermore, the group $\mathrm{Sp}(n, \mathbf{R})$ is (finitely) generated by the σ in (i)–(iii), and thus the above formulas allow us (in principle) to compute U_σ for all $\sigma \in \mathrm{Sp}(n, \mathbf{R})$, up to a constant multiple. Notice that this undetermined constant does not affect the conjugacy mapping

$$W(u,v) \mapsto U_\sigma \circ W(u,v) \circ U_\sigma^{-1}.$$

A further analysis allows us to choose the constants so that

$$U_{\sigma_1} U_{\sigma_2} = \pm U_{\sigma_1\cdot\sigma_2}.$$

This permits us to lift $\sigma \mapsto U_\sigma$ to a single-valued representation of the double covering of $\mathrm{Sp}(n, \mathbf{R})$, giving the "metaplectic representation".

(d) One has the following invariance property of the Weyl correspondence. For each symbol a, let $\mathrm{Op}(a)$ be the pseudo-differential operator given by the formula (52). Then for each $\sigma \in \mathrm{Sp}(n, \mathbf{R})$

$$\mathrm{Op}(a(\sigma(x,\xi))) = U_\sigma \circ \mathrm{Op}(a(x,\xi)) \circ U_\sigma^{-1}; \qquad (**)$$

This is a consequence of (*) and (49). Note that from the above formula it follows that each U_σ preserves the Schwartz class $\mathcal{S}(\mathbf{R}^n)$, as well as the tempered distributions $\mathcal{S}'(\mathbf{R}^n)$. Moreover, when $a \in \mathcal{S}'(\mathbf{R}^n \times \mathbf{R}^n)$, the operator $\mathrm{Op}(a)$ given by (52) extends naturally to a mapping $\mathrm{Op}(a) : \mathcal{S}(\mathbf{R}^n) \to \mathcal{S}'(\mathbf{R}^n)$.

The idea of the metaplectic representation is implicit in the work of van Hove [1951], and then appears in Segal [1959] and [1963], Shale [1962], Weil [1963]; see also Mackey [1965] and the account in Folland [1989]. In Weil [1965], as well as in his earlier paper, connections with number theory were developed; these relate in particular to Siegel's theory of quadratic forms where the group $\mathrm{Sp}(n, \mathbf{R})$ had in fact already played a significant role. The metaplectic representation is also of interest in other parts of the representation theory of semisimple groups; see, e.g., Kashiwara and Vergne [1978], Howe [1989].

7.6 The above invariance properties characterize the symmetric pseudo-differential correspondence $a \mapsto \mathrm{Op}(a)$. More precisely, suppose $a \mapsto \tilde{\mathrm{Op}}(a)$ is a linear mapping from $\mathcal{S}'(\mathbf{R}^n \times \mathbf{R}^n)$ to linear operators $\mathcal{S}(\mathbf{R}^n) \to \mathcal{S}'(\mathbf{R}^n)$, that is continuous in the topology of $\mathcal{S}'(\mathbf{R}^n \times \mathbf{R}^n)$. Assume also that:

(i) If $a(x,\xi) = a(x) \in L^\infty(\mathbf{R}^n)$, then $(\tilde{\mathrm{Op}}(a)f)(x) = a(x)f(x)$.

(ii) If $\sigma \in \mathrm{Sp}(n, \mathbf{R})$, then $\tilde{\mathrm{Op}}(a(\sigma(x,\xi))) = U_\sigma \circ \mathrm{Op}(a(x,\xi)) \circ U_\sigma^{-1}$; here U_σ is as in §7.5.

Then $\tilde{\mathrm{Op}}(a) = \mathrm{Op}(a)$, as given by (52).

Indeed, if $a(x) = e^{2\pi i u\cdot x}$ then (by (i)) $\tilde{\mathrm{Op}}(a) = \mathrm{Op}(a)$. As a consequence, (ii) and (**) in §7.5 yield that this identity also holds when $a(x,\xi) = e^{2\pi i(u\cdot x + v\cdot\xi)}$, for any $(u,v) \in \mathbf{R}^n \times \mathbf{R}^n$. The assertion then follows by linearity and a limiting argument.

7.7 The composition formula for pseudo-differential operators given in §3 of Chapter 6 has the following form in the setting of the Weyl correspondence $a \mapsto \mathrm{Op}(a)$.

Let $a \in S^{m_1}$, $b \in S^{m_2}$. Then there exists $c \in S^{m_1+m_2}$ so that $\mathrm{Op}(c) = \mathrm{Op}(a) \circ \mathrm{Op}(b)$. For c we have the asymptotic formula

$$c(x,\xi) \sim \sum_{k=0}^{\infty} \left(\frac{i}{4\pi}\right)^k (\partial_y \cdot \partial_\xi - \partial_x \cdot \partial_\eta)^k a(x,\xi)\, b(y,\eta)\big|_{\substack{y=x \\ \eta=\xi}}.$$

For a detailed proof see, e.g., Folland [1989], Chapter 2.

7.8 The symplectic aspect of the correspondence $a \mapsto \mathrm{Op}(a)$ suggests that it might be relevant to consider symbol classes that display symmetry in x and ξ. In this connection, we define the symmetric class S^m_{sym} to consist of those $a(x,\xi) \in C^\infty(\mathbf{R}^n \times \mathbf{R}^n)$ that satisfy

$$|\partial_x^\beta \partial_\xi^\alpha a(x,\xi)| \le A_{\alpha,\beta}(1 + |x| + |\xi|)^{m-|\alpha|-|\beta|};$$

this should be compared with the nonsymmetric condition (2) of Chapter 6.

It is not difficult to verify that for the correspondence $a \mapsto \mathrm{Op}(a)$ the symbolic calculus formula in §7.7 is still valid; similarly, for the correspondence $a \mapsto T_a$, the formula (13) of Chapter 6 continues to apply.

This calculus is of particular interest in relation to the Hermite operators considered in §6.4. Indeed, if $L = -\Delta + |x|^2$, then L has symbol $|x|^2 + 4\pi^2|\xi|^2 \in S^2_{\mathrm{sym}}$. Also, for any β, $\beta \ne n, n+2, n+4, \ldots$, the operator $(L - \beta I)^{-1}$ is of the form $\mathrm{Op}(a)$ with

$$a(x,\xi) - (1 + |x|^2 + 4\pi^2|\xi|^2)^{-1} \in S^{-4}_{\mathrm{sym}};$$

similarly, $(L - \beta I)^{-1}$ can also be written as $T_{a'}$ with

$$a'(x,\xi) - (1 + |x|^2 + 4\pi^2|\xi|^2)^{-1} \in S^{-4}_{\mathrm{sym}}.$$

A. Grossman, Loupias, and Stein [1969]. Further results may be found in Voros [1978], Hörmander [1979], Howe [1980b].

C. Singular integrals and oscillatory integrals

7.9 We make several observations about singular integrals with homogeneous kernels on the Heisenberg group.

(a) Suppose K_0 is a function defined on $\mathbf{H}^n\backslash\{0\}$ that is smooth and homogeneous of degree $-2n-2$ with respect to the dilations

$$x = [z,t] \mapsto [\delta z, \delta^2 t]$$

on $\mathbf{H}^n$. Let us assume the cancellation condition $\int_{a<\rho(x)<b} K_0(x)\,dx = 0$ for all $a, b > 0$. Then the distribution $K = \text{p.v.}K_0$, defined by

$$K(\Phi) = \lim_{\varepsilon\to 0} \int_{\rho(x)>\varepsilon} K_0(x)\,\Phi(x)\,dx, \qquad \text{for } \Phi \in \mathcal{S},$$

satisfies the hypotheses of Theorem 3. Hence $Tf = f * K$ extends to a bounded operator $L^p(\mathbf{H}^n) \to L^p(\mathbf{H}^n)$ for $1 < p < \infty$.

(b) Let K_1 be smooth away from the origin and homogeneous of degree $-2n-1$. Then the functions

$$K_0(z,t) = \frac{\partial}{\partial z} K_1(z,t), \qquad K_0(z,t) = \frac{\partial}{\partial \bar{z}} K_1(z,t)$$

satisfy the conditions in (a) above. The same holds for $K_0(z,t) = \dfrac{\partial K_2(z,t)}{\partial t}$ if K_2 is assumed to be smooth away from the origin and homogeneous of degree $-2n$.

(c) One should remark however that the distribution p.v.K_0 need not agree with the distribution $\dfrac{\partial K_1(z,t)}{\partial z}$; similarly for $\dfrac{\partial K_1(z,t)}{\partial \bar{z}}$ and $\dfrac{\partial K_2(z,t)}{\partial t}$.

A good example arises in the theory of the Cauchy-Szegö integral (given in §2.4). There

$$K_2(z,t) = \frac{-c}{n}(t + i|z|^2)^{-n}$$

and the distribution kernel of the Cauchy-Szegö projection is $K = \dfrac{\partial K_2}{\partial t}$. However, if

$$K_0(z,t) = \frac{\partial}{\partial t} K_2(z,t),$$

then one has the "jump" relation

$$K = \text{p.v.}K_0 + \frac{1}{2}\delta_0,$$

where δ_0 is the Dirac delta function at the origin.

In (a), the assertion that p.v.K_0 satisfies the key cancellation condition (83) required in Theorem 3 follows from the fact that p.v.$K_0(\Phi^R)$ is independent of R (which holds because of the homogeneity of K_0).

Conversely, to prove (b), one notices first that p.v.$K_0(\Phi^R)$ is bounded independently of R. For some related facts about homogeneous distributions see §8.19–§8.20 in Chapter 1 and §7.5 in Chapter 6.

For the identity in (c) regarding the Cauchy-Szegö kernel, see Korányi and Vági [1971].

7.10 We consider the general version of "twisted convolution" and the corresponding oscillatory integrals which arose in §5.1. We assume that B is a real bilinear form on $\mathbf{R}^m \times \mathbf{R}^m$, and define $F *_B G$ by

$$(F *_B G)(x) = \int_{\mathbf{R}^m} F(x-y)\,G(y)\,e^{iB(x,y)}\,dy.$$

(a) If B is nondegenerate and antisymmetric, then we have

$$\|F *_B G\|_{L^2(\mathbf{R}^m)} \leq c\|F\|_{L^2(\mathbf{R}^m)}\|G\|_{L^2(\mathbf{R}^m)}, \qquad c = c_B.$$

(b) If B is nondegenerate but not antisymmetric, (a) may fail.

(c) However, when B is nondegenerate and K is a function with

$$|\partial_x^\alpha K(x)| \leq A_\alpha |x|^{-\mu-|\alpha|}, \qquad \text{all } \alpha,$$

then the operator $f \mapsto T_B(f) = f *_B K$ is bounded from $L^2(\mathbf{R}^n)$ to itself, whenever $0 \leq \mu < m$. The case corresponding to $\mu = m$ is treated in §5.1 and does not require the nondegeneracy of B.

The assertion (a) follows from the special case dealt with in the proposition in §4.4, and an appropriate change of variables. To see (b), take $B(x,y) = x \cdot y$. For (c), see Phong and Stein [1986a]; the case where $\mu = m$ and B is antisymmetric and nondegenerate is in Mauceri, Picardello, and Ricci [1981].

7.11 The following operator on $\mathbf{H}^n$ is a key example of a singular Radon transform as discussed in Chapter 11, §4.7. Let K be a singular integral distribution of the kind considered in §5.1 with $m = 2n$. Let δ_0 be the Dirac delta function at the origin of $\mathbf{R}^1$. Define the distribution $M = K \otimes \delta_0$ on $\mathbf{H}^n = \{[z,t]\}$, with K acting on the z-variable and δ_0 acting on the t-variable. Set

$$T(f) = f * M, \qquad f \in \mathcal{S}.$$

For each $x \in \mathbf{H}^n$, $(Tf)(x)$ is obtained by integrating over the left translate by x of the hyperplane $\{[z,0]\}$; thus T is of the form described in §4.7 of the previous chapter.

One can assert that T extends to a bounded operator from $L^p(\mathbf{H}^n)$ to itself, when $1 < p < \infty$.

A proof for $p = 2$ can be given as follows. By §3.3, if $F = f * M$, then for each real λ we have $F_\lambda^0 = f_\lambda^0 *_\lambda M_\lambda^0$. However $M_\lambda^0 = K$ for every λ, and thus the L^2 inequality is a consequence of the corollary in §5.1.4. The case $p \neq 2$ requires a further interpolation argument. The full proofs are in Geller and Stein [1984].

Two related remarks about this operator are in order. First, if (in addition) K is homogeneous of degree $-2n$, then the distribution M is homogeneous of the critical degree $-2n-2$ with respect to the dilations $[z,t] \mapsto [\delta z, \delta^2 t]$ of the Heisenberg group; however it is also homogeneous of the critical degree (now $-2n-1$) with respect to the isotropic dilations $[z,t] \mapsto [\delta z, \delta t]$ of the underlying $\mathbf{R}^{2n+1} = \mathbf{C}^n \times \mathbf{R}^1$. Second, the fact that M simultaneously displays these two homogeneities explains in part its role in the $\bar{\partial}$-Neumann problem; in this connection see Phong and Stein [1986b].

D. Hermite expansion

7.12 We note some additional facts about Hermite expansions.

(a) In connection with the discussion of the Cauchy-Szegö projection in §6.5, one has the following "annihilation" and "creation" identities:

$$(x_j + \frac{\partial}{\partial x_j})H_k(x) = 2k_j H_{k-e_j}(x), \qquad (x_j - \frac{\partial}{\partial x_j})H_k(x) = H_{k+e_j}(x),$$

where $k = (k_1, \ldots, k_n)$ and $e_j = (0, \ldots, 0, 1, 0, \ldots, 0)$.

(b) Suppose H_k^* are the re-scaled versions of the Hermite functions, given by

$$H_k^*(x) = c_{k,\lambda} H_k((2\pi|\lambda|)^{1/2}x),$$

where $c_{k,\lambda} = (2|\lambda|)^{n/4}\pi^{|k|}(k!)^{-1/2}$, and $\lambda \neq 0$. Note that the H_k^* are normalized eigenfunctions of the operator $(2\pi\lambda|x|)^2 - \Delta_x$, corresponding to the eigenvalues $2\pi|\lambda|(2|k|+n)$. If we fix $\lambda = 1$, then we can assert that H_k^* is essentially its own Fourier transform. More precisely, when $\lambda = 1$,

$$\widehat{H_k^*} = i^{-|k|}H_k^*.$$

(c) The Mehler kernel is defined by $M(x,y,r) = \sum_k r^{|k|} H_k^*(x)\, H_k^*(y)$, for $|r| < 1$. Thus

$$M(x,y,r) = \left(\frac{2|\lambda|}{1-r^2}\right)^{n/2} \exp\left\{\pi|\lambda|\frac{4rxy - (1+r^2)(x^2+y^2)}{1-r^2}\right\}.$$

The proofs of (a) and (b) follow directly from the generating identity (98) and the fact that

$$H_k(x) = H_{k_1}(x_1)\cdot\cdots\cdot H_{k_n}(x_n).$$

The Fourier transform identity (b) is also a consequence of (c), in the limiting case $r = -i$, $\lambda = 1$. The formula for the Mehler kernel is derived in, e.g., Wiener [1933].

A simple and direct approach to (c) is suggested by the treatment of the Ornstein-Uhlenbeck process in Feller [1966]. If we set $\lambda = 1/2\pi$ and

$$U(x,t) = \int_{\mathbf{R}^n} M(x,y,e^{-t})\, f(y)\, dy,$$

then U satisfies the differential equation

$$\frac{\partial U}{\partial t} = \frac{1}{2}\{(\Delta_x - |x|^2)U(x,t) - U(x,t)\}, \qquad U(x,0) \equiv f(x).$$

Now let $u(x,t) = e^{|x|^2/2}U(x,t)$. Then u satisfies

$$\frac{\partial u}{\partial t} = \frac{1}{2}\{\Delta_x u - 2x\cdot\nabla u\}, \qquad u(x,0) \equiv e^{|x|^2/2}f(x).$$

Next set $w(x,t) = u(xe^t, t)$. Then w satisfies

$$e^{2t}\frac{\partial w}{\partial t} = \frac{1}{2}\Delta_x w, \qquad w(x,0) \equiv e^{|x|^2/2}f(x).$$

Finally, if we let $h(x,\tau) = w(x,t)$, with $\tau = (1 - e^{-2t})/2$, then h satisfies the heat equation

$$\frac{\partial h}{\partial \tau} = \frac{1}{2}\Delta_x h, \qquad h(x,0) \equiv e^{|x|^2/2}f(x),$$

and hence $h(x,\tau) = (2\pi\tau)^{-n/2}\int_{\mathbf{R}^n} e^{-|x^2-y^2|/2\tau} f(y)\, e^{|y|^2/2}\, dy$. Retracing these substitutions gives the desired result.

7.12[a] The relations between the Hermite expansions, the Mehler kernel, and the Fourier transform have further interesting ramifications.

(a) The classical Hausdorff-Young inequality [‡]

$$\|\widehat{f}\|_q \le \|f\|_p, \qquad \text{where } p^{-1} + q^{-1} = 1 \text{ and } 1 \le p \le 2,$$

can be improved so that the norm 1 is replaced by $(p^{1/p}/q^{1/q})^{n/2}$; this bound is attained when f is a Gaussian (say $f = e^{-\pi|x|^2}$). Beckner [1975].

(b) Connected with this are inequalities of the type

$$\|\mathcal{H}_r(f)\|_{L^q(d\mu)} \le \|f\|_{L^p(d\mu)}, \qquad d\mu = (2\pi)^{-n/2}e^{-|x|^2/2}\, dx. \qquad (*)$$

Here $\mathcal{H}_r$ is the operator with kernel

$$(1-r^2)^{-n/2}\exp\left(\frac{-r^2(x^2+y^2) + 2rx\cdot y}{2(1-r^2)}\right),$$

integration is taken with respect to $d\mu$, and p, q, and r are suitably chosen. It can be verified that (a) is identical with (*) when $r = -i(p-1)^{1/2}$; note that r is imaginary. When r is positive, $r \le (\frac{p-1}{q-1})^{1/2}$, $1 < p \le q < \infty$, (*) is an earlier "hypercontractive" inequality of Nelson [1973]; see also Gross [1975].

(c) The inequality (*) can be reduced to its "two-point" version, namely the special case

$$\left(\frac{|1-rz|^q + |1+rz|^q}{2}\right)^{1/q} \le \left(\frac{|1-z|^p + |1+z|^p}{2}\right)^{1/p}, \qquad \text{all } z \in \mathbf{C}.$$

This reduction was initiated in Gross [1975] and extended in Beckner [1975].

7.13 A function f on the Heisenberg group is said to be *radial* if $f(z,t) = f_0(|z|,t)$ whenever $(z,t) \in \mathbf{C}^n \times \mathbf{R}$. More generally, f is said to be *polyradial* if

$$f(z,t) = f_0(|z_1|, \ldots, |z_n|, t), \qquad (z,t) \in \mathbf{C}^n \times \mathbf{R}.$$

(a) The radial $L^1(\mathbf{H}^n)$ functions form a commutative algebra under convolution. The same is true for the polyradial functions.

(b) The group Fourier transform (see §6.3) allows one to re-interpret the above as follows. Suppose $f \in L^1(\mathbf{H}^n)$. Then f is polyradial if and only if for every $\lambda \neq 0$, the operator $\widehat{f_H}(\lambda)$ is diagonal on the Hermite basis $\{H_k^*\}$ described in §7.12(b) above, that is

$$\widehat{f_H}(\lambda)H_k^* = \mu(k,\lambda)H_k^*$$

for appropriate scalars $\mu(k,\lambda)$. Similarly, f is radial if and only if, in addition, we have $\mu(k,\lambda) = \mu(k',\lambda)$ whenever $|k| = |k'|$.

[‡] See, e.g., *Fourier Analysis*, p. 178.

(c) Whenever $f \in L^1(\mathbf{H}^n)$ is radial, the diagonal entries $\mu(k,\lambda)$ can be determined by the formula

$$\mu(k,\lambda) = \int_{\mathbf{H}^n} f(z,t)\, \Lambda_{|k|}(\lambda|z|^2)\, e^{2\pi i\lambda t}\, dz\, dt,$$

where $\Lambda_{|k|}$ can be given explicitly in terms of Laguerre functions. There are similar formulae in the polyradial case and, more generally, whenever f transforms as a character of the n-torus $\mathbf{T}^n$ as z transforms by

$$z \mapsto (e^{i\theta_1} z_1, \ldots, e^{i\theta_n} z_n), \qquad \theta = (\theta_j) \in \mathbf{T}^n.$$

Assertion (a) can be proved directly by a change of variables in the integral defining the convolution of two functions (see, for instance, Folland and Stein [1982]). A systematic theory of the Fourier transform on the Heisenberg group, including (b) and (c), is given in Geller [1977], [1984]; some earlier sources are Miller [1968], Vilenkin [1968], Peetre [1972].

E. Other topics

7.14 (a) If X is any real smooth vector field (given on, say, an open subset of $\mathbf{R}^n$) and s is a sufficiently small number, then one can construct $\exp(sX) = \phi_s$ to be the (local) one-parameter family of diffeomorphisms that satisfy:

(i) $\phi_{s_1} \circ \phi_{s_2} = \phi_{s_1+s_2}$ (where defined),

(ii) ϕ_0 is the identity, and

(iii) $\frac{\partial}{\partial s} f(\phi_s(x))\big|_{s=0} = (Xf)(x)$.

(b) We observe that if $u + iv \in \mathbf{C}^n$, $u, v \in \mathbf{R}^n$, then in the setting where $\mathbf{R}^N = \mathbf{H}^n$, $N = 2n+1$, we have that

$$\exp\Big(\sum_{j=1}^n (u_j X_j + v_j Y_j) + \tau T\Big)(x) = x \cdot [u+iv, \tau], \qquad \text{for } x \in \mathbf{H}^n.$$

Note that in the above the group multiplication is on the right, because X_j, Y_j, T are left-invariant vector fields. This formula means that $[z,t]$ are *canonical coordinates* for the Heisenberg group.

(c) If A and B are operators, then in a formal sense

$$\exp A \cdot \exp B = \exp(A + B + \tfrac{1}{2}[A,B] + \cdots),$$

where $\cdots$ represents sums of commutators of A and B whose length is at least 3; a more precise version is given by the "Campbell-Hausdorff formula". In the case where A and B are (sufficiently small multiples of) elements of the Lie algebra of a Lie group, the formula is actually convergent. So letting

$$A = \sum_{j=1}^n (u_j X_j + v_j Y_j) + \tau T, \qquad B = \sum_{j=1}^n (u'_j X_j + v'_j Y_j) + \tau' T$$

and noting that all commutators of length ≥ 3 vanish, we see by the use of (b) that the Lie algebra structure of $\mathfrak{h}^n$ actually determines the group structure of $\mathbf{H}^n$.

(d) One can also apply the Campbell-Hausdorff formula to obtain a (formal) derivation of the identity (51):

$$W(u,v)f(x) = e^{2\pi i x\cdot u} e^{\pi i u\cdot v} f(x+v).$$

Indeed, let $A = 2\pi i u \cdot Q$, $B = 2\pi i v \cdot P$. Thus

$$(Af)(x) = (2\pi i u \cdot x) f(x), \qquad (Bf)(x) = v \cdot \nabla_x(f).$$

Therefore, $(e^A f)(x) = e^{2\pi i u\cdot x} f(x)$, $(e^B f)(x) = f(x+v)$. Now

$$W(u,v) = e^{2\pi i(u\cdot Q + v\cdot P)} = e^{A+B};$$

but $e^A e^B = e^{A+B+[A,B]/2}$, since $[A,B]f = (-2\pi i u\cdot v) f$, and all other commutators vanish. As a result

$$e^{A+B} = e^{-[A,B]/2}\, e^A\, e^B, \qquad \text{while} \quad e^{-[A,B]/2} f = e^{i\pi u\cdot v} f,$$

and the assertion is established.

For the basic facts concerning the relation between Lie algebras and Lie groups and the use of the Campbell-Hausdorff formula, the reader may consult Hausner and Schwartz [1968], Varadarajan [1974].

7.15 We describe the relation between convolution operators on the Heisenberg group and pseudo-differential operators. The main thrust is that the naturally arising operators on the Heisenberg group have symbols that belong essentially to the exotic class $S_{1/2,1/2}$ (as treated in Chapter 7, §1).

(a) To begin with, we consider the formal set-up and let $Tf = f * K$ with (say) $K \in \mathcal{S}$. Then $Tf = T_a(f)$, where the symbol a is given by

$$a(x,\xi) = \widehat{K}(L_x(\xi)).$$

Here $\widehat{K}$ is the Euclidean Fourier transform of K, and for each x, L_x is a linear transformation, which actually depends linearly on x. Indeed, we can write the group multiplication $y^{-1} \cdot x$ as $M_x(x-y)$ where, for each x, M_x is a linear transformation having determinant 1, and then $L_x = (M_x^t)^{-1}$. The same formula, when properly interpreted, is valid when K is a distribution.

(b) In the case of the Cauchy-Szegö kernel K (given by (29)), one has that

$$\widehat{K} = \begin{cases} 2^n e^{-\pi|\zeta|^2/2\tau}, & \text{if } \tau > 0; \\ 0, & \text{if } \tau < 0. \end{cases}$$

Here $\xi = [\zeta, t]$ where if $x = [z,t]$ then ζ is dual to z, and τ is dual to t; the duality is implemented by the inner product $x \cdot \xi = \mathrm{Re}(z \cdot \overline{\zeta}) + t\tau$. Note that $\widehat{K}$ is homogeneous of degree 0, with respect to the dilations $[\zeta, t] \mapsto [\delta\zeta, \delta^2 t]$, $t > 0$.

(c) In the more general setting of kernels of the kind in §5.2, we assume that K satisfies (80)–(83), but we strengthen the regularity assumptions on the function $K(x)$ so as to hold for derivatives of all orders. That is, we suppose also that for all α, β, $|\partial_t^\beta \partial_{z,\bar z}^\alpha K(x)| \leq A_{\alpha,\beta}\rho(x)^{-2n-2-|\alpha|-2|\beta|}$; here $x = [z,t]$. When K is in this class, we have that $|\partial_\tau^\beta \partial_{\zeta,\bar\zeta}^\alpha \widehat{K}(\xi)| \leq A'_{\alpha,\beta}\rho(\xi)^{-|\alpha|-2|\beta|}$ for all α, β; again $\xi = [\zeta, \tau]$.

(d) As a result, we have the following. Suppose K belongs to the class of kernels described in (c), and T^0 is a localized version of the convolution operator $Tf = f * K$, namely

$$T^0(f) = \psi_1 T(\psi_2 f),$$

where $\psi_1, \psi_2 \in C_0^\infty$ are cut-off functions.

Then $T^0 = T_a$, with $a \in S^0_{1/2,1/2}$. Moreover, for large ξ we have

$$a(x,\xi) = \psi_1(x)\,\psi_2(x)\,\widehat{K}(L_x(\xi)) \qquad \text{modulo } S^{-1/2}_{1/2,1/2}.$$

For the above, see Greiner and Stein [1977]. A proof of the assertion in (b) can be given by starting with the identity

$$\int_{\mathbf{C}^n} e^{-\pi|\zeta|^2/2\tau}\, e^{2\pi i \mathrm{Re}(z\cdot\bar\zeta)}\, d\zeta = 2^n\,\tau^n\, e^{-2\pi\tau|z|^2}, \qquad \text{when } \tau > 0.$$

For results related to (c), see §8.20 in Chapter 1, and §7.4 in Chapter 6.

Notes

§1. The group of fractional linear transformations of the unit ball in $\mathbf{C}^{n+1}$ appears implicitly in 1884 in the work of Poincaré and Picard. This group (which of course contains the Heisenberg group) is detailed explicitly in Siegel [1950]; the identification of $\mathbf{H}^n$ with the boundary of the ball in its unbounded realization is in Piatetski-Shapiro [1961].

§2. The Cauchy-Szegö kernel for the ball in $\mathbf{C}^{n+1}$ was found by Hua [1963], and for $\mathcal{U}^n$ by Gindikin [1964]. Theorem 2, giving the L^p boundedness, is due to Korányi and Vági [1971]. The relation of the Lie algebra $\mathfrak{h}^n$ with Cauchy-Riemann operators (as in §2.6) was stressed in Folland and Stein [1974].

§3. The commutation relations determining the algebra $\mathfrak{h}^n$, and hence the Heisenberg group, have their origin in the development of quantum mechanics around 1925. A systematic treatment of various topics dealing with the relation of $\mathbf{H}^n$ and phase-space considerations is given by Folland [1989].

§4 and §5. The definition of the Weyl correspondence appears in Weyl [1931]. The general L^2 twisted convolution theorem in §4.4 is due to Segal [1963]. The theorems involving singular integral kernels given in §5.1 may be found in Mauceri, Picardello, and Ricci [1981], Phong and Stein [1986a]. The singular integral result (Theorem 3) goes back essentially to Knapp and Stein [1971], but its proof was that given in §5.3 of the next chapter. The present argument is adapted from Ricci [1982].

CHAPTER XIII

More about the Heisenberg Group

We continue here with our study of the application of the Heisenberg group to analysis in several complex variables. Our particular focus will be on the $\bar\partial$ complex, in particular its boundary analogue, and the resulting problems of solvability and regularity, in the setting of the generalized upper half-space $\mathcal{U}^n$. One is led to these questions by considering the nonhomogeneous Cauchy-Riemann equations, and the related $\bar\partial$-Neumann problem. The ideas that are developed here turn out to be relevant in several other contexts as well.

The considerations that underlie the material presented in this chapter can be summarized as follows.

The model role of $\mathbf{H}^n$. The special domain $\mathcal{U}^n$, with its boundary $\mathbf{H}^n$, provides a useful model in the study of more general domains $\mathcal{D} \subset \mathbf{C}^{n+1}$. In fact, the virtues of this model are twofold. First, the group of symmetries operating on $\mathcal{U}^n$ (and $\mathbf{H}^n$) lead to explicit and suggestive formulas that apply in this case. Second, and of equal importance, is the consequence of a suitable "approximation" technique: a strictly pseudoconvex domain can be well approximated (near any boundary point) by $\mathcal{U}^n$ and, as a result, the exact formulas that hold in that special case can be carried over as approximate identities for this more general situation.

The $\bar\partial$-Neumann problem and boundary analogues. These are at the source of our study here and arise when we try to solve the inhomogeneous Cauchy-Riemann equations $\bar\partial u = f$ in $\mathcal{D}$. The formal set-up is simple: we consider the auxiliary (Laplace) equation $\Box U = f$, with U satisfying the $\bar\partial$-Neumann boundary conditions; then $u = \bar\partial^* U$ gives the desired solution, with u orthogonal to holomorphic functions. Thus the problem for u has been transferred to that for U; the latter of which is "nonelliptic" due to the nature of the boundary conditions. To analyze the situation further, one considers a boundary version of the above formalism: this gives a parallel $\bar\partial_b$ problem, and an operator $\Box_b$ (the "Kohn Laplacian") which is the analogue of $\Box$. There are no boundary conditions for $\Box_b$, but now the nonellipticity is inherent in the fact that the quadratic terms in $\Box_b$ are only semidefinite.

The operator $\mathfrak{L}_\alpha$ *on* $\mathbf{H}^n$. In the case corresponding to $\mathcal{U}^n$, the mysteries of $\Box_b$ are revealed by its identification with the operator $\mathfrak{L}_\alpha$. The latter operator can be characterized (without reference to $\bar{\partial}$) purely by symmetry considerations on $\mathbf{H}^n$: it is the analogue of the Laplacian Δ on $\mathbf{R}^n$, in the sense that it is suitably invariant under the action of translations, dilations, and unitary rotations arising from the group structure of $\mathbf{H}^n$. The operator $\mathfrak{L}_\alpha$ depends on a characteristic parameter α, and when restricted to q-forms, we have $\Box_b = \mathfrak{L}_\alpha$, with $\alpha = n - 2q$.

The main point about $\mathfrak{L}_\alpha$ is that, except when α belongs to a discrete set of "forbidden" values, it has an explicitly determined fundamental solution. As a result, one can obtain regularity properties of solutions of $\mathfrak{L}_\alpha$ by the use of the singular integrals on $\mathbf{H}^n$ studied in the previous chapter. One gets sharp estimates that recover a gain of 2 in the "good" directions; these have provided the paradigm for all further "maximal subelliptic" estimates of this kind.

Lewy example. The assertion of Lewy, concerning the $\bar{\partial}_b$ equation, states that the equation $\dfrac{\partial u}{\partial z_1} + i\bar{z}_1\dfrac{\partial u}{\partial t} = f$ has no solution u for "generic" functions f.† The original analysis made no reference to the Heisenberg group. However, using it (and the above ideas) allows one to clarify the situation and, in particular, to determine those f for which the problem is locally solvable. These conclusions are related to $\mathfrak{L}_\alpha$ for $n = 1$ and the forbidden value $\alpha = 1$, and result from consideration of the fundamental solutions for $\alpha \neq 1$: one obtains a "relative" fundamental solution which takes into account the null-space of $\mathfrak{L}_1$, in terms of the Cauchy-Szegö projection.

Homogeneous groups. With our last topic, we reach not the end, but another beginning. It is our purpose here to introduce the reader to a subject that holds promise for interesting further developments: the study of homogeneous groups. These groups are a natural generalization of the Heisenberg group, and have already proved their value by their applicability to a variety of problems.

A word about the limitations of our exposition here. It is not possible, within the confines of this book, to give a systematic account of the $\bar{\partial}$-Neumann problem, nor of the approximation technique referred to above. Instead we focus on that part that deals with analysis on $\mathbf{H}^n$ *per se.* A general introductory survey of the $\bar{\partial}$-Neumann problem is given in §1, while the exact formulas for this problem on $\mathcal{U}^n$ are derived in the Appendix §6, and the approximating technique is outlined in §7.13.

† This equation can be rewritten as $(X_1 - iY_1)u = 2f$, in the notation of (37) and (38) in Chapter 12.

1. The Cauchy-Riemann complex and its boundary analogue

The study of these problems represents an important chapter in complex analysis but, unfortunately, it will not be possible to give here a self-contained exposition of this subject. Nevertheless we will attempt in this section to sketch, without proof, some of the background material so as to provide motivation for the results detailed in the succeeding sections.

1.1 Cauchy-Riemann equations. Our first concern will be with the Cauchy-Riemann equations on a domain $\mathcal{D}$ in $\mathbf{C}^{n+1}$,

$$\bar{\partial} u = f. \tag{1}$$

Equation (1) represents the system $\partial u/\partial \bar{z}_j = f_j$ where, as usual, we write

$$\frac{\partial}{\partial \bar{z}_j} = \frac{1}{2}\left(\frac{\partial}{\partial x_j} + i\frac{\partial}{\partial y_j}\right)$$

for $z_j = x_j + iy_j$, $j = 1, \dots, n+1$. The given functions f_j are required to satisfy the necessary compatibility conditions

$$\frac{\partial f_j}{\partial \bar{z}_k} = \frac{\partial f_k}{\partial \bar{z}_j}, \qquad j = 1, \dots, n+1. \tag{1$'$}$$

Holomorphic functions u on $\mathcal{D}$ are then characterized as solutions of the homogeneous system (1), $\bar{\partial} u = 0$, but the inhomogeneous problem often arises in applications. We give a brief indication of this.

Suppose we are given a holomorphic function h defined (locally) in a neighborhood of some boundary point z_0 of $\mathcal{D}$, and wish to construct a (global) holomorphic function F on $\mathcal{D}$, so that F displays the same behavior as h near z_0.

To produce such an F, take ψ to be a suitable smooth cut-off function that is 1 near z_0, and let $f = \bar{\partial}(\psi h)$. Assuming that f is smooth in the closure of $\mathcal{D}$ (as is the case if h is smooth in $\bar{\mathcal{D}}$, away from z_0), and that for any such f, the problem (1) has a solution u that is also smooth in the closure of $\mathcal{D}$, we set $F = \psi h - u$. Clearly, f is holomorphic in $\mathcal{D}$ and differs from h (near z_0) by a smooth function.

Thus the problem for the inhomogeneous Cauchy-Riemann equations (1) that arises here may be restated as follows. Given an f having a certain degree of smoothness on $\mathcal{D}$ (that satisfies the compatibility conditions (1$'$)), we wish to find a u satisfying (1) that enjoys a corresponding degree of smoothness.

1.2 Passage to Laplace's equation. In attacking the problem (1), it is desirable to require additional conditions, which will guarantee that the solution is unique. While various formulations of conditions of this type are possible, the simplest is to require that the solution u of (1) be orthogonal to holomorphic functions (the null solutions). This means that we are implicitly assuming that u is in $L^2(\mathcal{D})$, and

$$\int_{\mathcal{D}} u\,\bar{h} = 0 \tag{2}$$

for all $h \in L^2(\mathcal{D})$ that are holomorphic in $\mathcal{D}$.

Stated in this way, the problem can be reduced, by the "method of orthogonal projections", to the solution of a boundary-value problem for Laplace's equation. Let us first describe the simple solution that arises in the case of one complex variable. We assume that $\mathcal{D}$ is bounded and that $b\mathcal{D}$ is smooth.

We envisage the auxiliary Dirichlet boundary-value problem

$$\Delta U = f \qquad \text{in } \mathcal{D}, \tag{3}$$

where U is required to satisfy

$$U|_{b\mathcal{D}} = 0. \tag{4}$$

Here $\Delta = \partial^2/\partial x^2 + \partial^2/\partial y^2$, with $z = x + iy$. The existence, uniqueness, and regularity properties of the solution U can be taken as known by the (classical) elliptic theory. Now set

$$u = 4\partial U, \tag{5}$$

where

$$\partial = \frac{1}{2}\left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\right).$$

Then since $4\bar{\partial}\partial = \Delta$, we see that u satisfies (1). Moreover, the orthogonality condition (2) holds because (assuming h to be smooth in $\bar{\mathcal{D}}$), we have

$$\int_{\mathcal{D}} u\bar{h}\,dz = 4\int_{\mathcal{D}} \partial U\,\bar{h}\,dz = -4\int_{\mathcal{D}} U\,\overline{\bar{\partial} h}\,dz = 0$$

by integration by parts; note that the resulting boundary terms vanish because of (4). Finally, the regularity properties of u can be deduced from the corresponding properties of U.

1.3 The $\bar{\partial}$-complex. In higher dimensions we must, to begin with, take into account that the system (1) is over-determined and, connected with this, that f must satisfy the compatibility conditions $\partial f_j/\partial\bar{z}_k = \partial f_k/\partial\bar{z}_j$. These matters are most neatly formulated using wedge products and differential forms.

We start with the basic 1-forms $d\bar{z}_1, \ldots, d\bar{z}_{n+1}$. For any ordered index set $I = \{j_1, \ldots, j_q\}$ where $1 \le j_1 < j_2 < \cdots < j_q \le n+1$ are q distinct integers between 1 and $n+1$, we write

$$d\bar{z}^I = d\bar{z}_{j_1} \wedge d\bar{z}_{j_2} \wedge \cdots \wedge d\bar{z}_{j_q};$$

we also let $|I| = q$. A q-form f is then a sum of the type

$$f = \sum_{|I|=q} f_I d\bar{z}^I, \tag{6}$$

where the f_I are complex-valued functions on $\mathcal{D}$.[‡]

Whenever f is smooth (i.e., the coefficients f_I are smooth), we can apply the $\bar{\partial}$ operator to f, obtaining

$$\bar{\partial} f = \sum_{j=1}^{n+1}\sum_{|I|=q} \frac{\partial f_I}{\partial \bar{z}_j} d\bar{z}_j \wedge d\bar{z}^I. \tag{7}$$

By regrouping the terms in (7) (using the alternating law for wedge products), we see that (7) gives a $q+1$-form, according to our definition (6). It is easy to verify the universal rule $\bar{\partial}^2 = 0$. Moreover, this definition is consistent with (1) if, in (1), we consider u to be a 0-form and take $f = \sum_{j=1}^{n+1} f_j d\bar{z}_j$. Note that the compatibility condition may then be stated as $\bar{\partial} f = 0$.

On q-forms we can define the obvious inner product

$$\langle f, g\rangle_q = \sum_{|I|=q} \int_{\mathcal{D}} f_I(z)\,\bar{g}_I(z)\,dz,$$

when $f = \sum_{|I|=q} f_I d\bar{z}^I$, $g = \sum_{|I|=q} g_I d\bar{z}^I$. This allows us to consider the formal adjoint $\bar{\partial}^*$ of $\bar{\partial}$, which maps q-forms to $(q-1)$-forms, and is defined by

$$\langle \bar{\partial}^* f, g\rangle_{q-1} = \langle f, \bar{\partial} g\rangle_q. \tag{8}$$

For a given q-form f, the relation (8) is required to hold for all smooth $(q-1)$-forms g. This in turn implies a restriction of the forms f for

[‡] In the usual terminology of differential forms, these are referred to as $(0,q)$ forms, the 0 indicating that no holomorphic differential dz_j appears in (6).

which $\bar{\partial}^*$ is definable, namely that the boundary terms arising from the indicated integration by parts must vanish, which is to say that a particular linear relation among the coefficients of f must vanish on the boundary. We abbreviate this by writing

$$f \in \text{domain}(\bar{\partial}_q^*).$$

We can now carry out the analogy with the boundary-value problem (3)–(4) to higher dimensions. The appropriate form to be given to the Laplacian is

$$\Box = \bar{\partial}^*\bar{\partial} + \bar{\partial}\bar{\partial}^*.$$

For a given one-form f on $\mathcal{D}$, we wish to find the one-form U determined by

$$\Box U = f, \tag{9}$$

where U is required to satisfy the boundary conditions implicit in the definition of $\Box U$, namely

$$U \in \text{domain}(\bar{\partial}_1^*), \qquad \bar{\partial}U \in \text{domain}(\bar{\partial}_2^*). \tag{10}$$

This is the $\bar{\partial}$-Neumann problem. It can be shown, as a consequence of the formalism above, that whenever $\bar{\partial}f = 0$, and assuming we can solve the $\bar{\partial}$-Neumann problem (9)–(10), then $u = \bar{\partial}^*U$ satisfies (1) and (2).

The main point we want to emphasize here is that this problem actually decouples into two boundary-value problems for the Laplacian. One is the usual Dirichlet problem, and is determined by the first boundary condition in (10). The other arises because of the second condition in (10), and is a "nonelliptic" boundary value problem for the Laplacian (this can be seen graphically in the case where $\mathcal{D}$ is the half-space $\mathcal{U}^n$; see §6.2 below). Inherent in the latter are some of the chief features that differentiate analysis in several complex variables from that in the classical case of one variable. We shall now touch on two of these aspects.

1.4 The boundary complex. A significant difference that occurs in higher dimensions is the presence of tangential Cauchy-Riemann operators. The importance of this is made evident by the remarkable principle (of Bochner, Lewy, and others) that a function on the boundary of a suitable domain $\mathcal{D}$ is the restriction of some holomorphic function on $\mathcal{D}$ exactly when it satisfies the tangential Cauchy-Riemann equations.

These considerations give rise to the $\bar{\partial}_b$-complex, which is the boundary analogue of the $\bar{\partial}$-complex discussed above. The idea behind the definition of $\bar{\partial}_b$ is as follows.

Suppose first that f is a (smooth) function defined on the boundary of $\mathcal{D}$ (i.e., f is a 0-form on $b\mathcal{D}$). To define $\bar{\partial}_b f$, we first extend f to F, a smooth function on all of $\bar{\mathcal{D}}$, and form $\bar{\partial}F$. We then restrict $\bar{\partial}F$ to $b\mathcal{D}$, and focus on the part of $\bar{\partial}F|_{b\mathcal{D}}$ that is independent of the particular extension F of f. Clearly, this is given by the combination of components that involve only tangential differentiation on the boundary. Put another way, if F_1 and F_2 are two extensions of f, then $\bar{\partial}F_1|_{b\mathcal{D}}$ and $\bar{\partial}F_2|_{b\mathcal{D}}$ differ by a multiple of $\bar{\partial}r$, where r is a defining function for the boundary of $\mathcal{D}$. Thus boundary 1-forms can be defined as restrictions of ordinary 1-forms to the boundary, modulo multiples of $\bar{\partial}r$.

Boundary q-forms can be defined similarly when $q > 1$, as well as the operator $\bar{\partial}_b$, which maps q-forms to $(q+1)$-forms. In the presence of a suitable inner product, we can also define the formal adjoint $\bar{\partial}_b^*$, and pass to the corresponding boundary Laplacian

$$\Box_b = \bar{\partial}_b^*\bar{\partial}_b + \bar{\partial}_b\bar{\partial}_b^*.$$

Inverting $\Box_b$ is closely related to solving the $\bar{\partial}$-Neumann problem, but is not identical to it. Note that, for $\Box_b$, there are no "boundary conditions" like (10); however the nonellipticity of the $\bar{\partial}$-Neumann problem is reflected here in that the second-order operator $\Box_b$ is not elliptic: it involves second-order differentiation only in the directions of the real and imaginary parts of the tangential Cauchy-Riemann operators, so there is always one "missing" direction.

1.5 Pseudoconvexity. Another significant feature of the higher-dimensional situation is the relevance of the pseudoconvexity of the domain $\mathcal{D}$. This property, which is a geometrical limitation on the nature of the boundary, is essential if one wishes to solve the inhomogeneous Cauchy-Riemann equations (1). It can be expressed in a variety of ways, but for our purposes it is most usefully formulated in terms of the tangential Cauchy-Riemann vector fields on $\mathcal{D}$.

Suppose that $\overline{W}_1, \ldots, \overline{W}_n$ is a basis for these vector fields (near some boundary point). Their real and imaginary parts span a (real) $2n$-dimensional subspace of the tangent space (which has real dimension $2n+1$) at each point. Let us choose a vector field S tangent to $b\mathcal{D}$ that generates the "missing" direction, and so that the vector field $i \cdot S$ points inward.

We can always write the commutator identity

$$[\overline{W}_j, W_k] = i a_{jk} S \qquad \text{modulo } \{W_j, \overline{W}_j\}_{j=1}^n. \tag{11}$$

This defines the $n \times n$ Hermitian matrix $\{a_{jk}\}$, which is called the *Levi matrix*. In its weakest form, *pseudoconvexity* is the requirement that this matrix be positive semidefinite at each point. In its strongest form, *strict pseudoconvexity*, we require that the Levi matrix be strictly positive definite at each point. These conditions do not depend on the choice of vector fields $\overline{W}_1, \ldots, \overline{W}_n, S$, and are holomorphic invariants.

The striking thing to notice is the transparent form that these conditions take when our domain $\mathcal{D}$ happens to be the upper half-space $\mathcal{U}^n$ treated in the previous chapter. In that case, when we choose $\overline{W}_1, \dots, \overline{W}_n, S$ to be the left-invariant tangential vector fields

$$\bar{Z}_1, \dots, \bar{Z}_n, T,$$

the condition (11) is then a consequence of (44) in §2.6 of that chapter.

The Heisenberg group thus gives the simplest model of the boundary of a strongly pseudoconvex domain; it is from this point of view that we shall study it in this chapter. Moreover, this analysis already contains the fundamental ideas needed to treat the $\bar{\partial}$-complex, the $\bar{\partial}$-Neumann problem, $\Box_b$, etc., for general strongly pseudoconvex domains; for these matters, see also §7.13 below.

2. The operators $\bar{\partial}_b$ and $\Box_b$ on the Heisenberg group

2.1 The $\bar{\partial}_b$ formalism. Motivated by what has been set forth in §1, and recalling the identification of the Heisenberg group with the boundary of the domain $\mathcal{U}^n$, we now define the $\bar{\partial}_b$-complex on $\mathbf{H}^n$. We continue to use the notation that labels points in $\mathbf{H}^n$ by $[z,t]$, with $z = (z_1, \dots, z_n) \in \mathbf{C}^n$ and $t \in \mathbf{R}$.

To begin with, we have the basic 1-forms on $\mathbf{H}^n$ given by

$$d\bar{z}_1, \dots, d\bar{z}_n,$$

with $z_j = x_j + iy_j$. For any ordered q-tuple $I = (j_1, \dots, j_q)$, where $1 \le j_1 < j_2 < \cdots < j_q \le n$, we write

$$d\bar{z}^I = d\bar{z}_{j_1} \wedge \cdots \wedge d\bar{z}_{j_q}.$$

A q-form on $\mathbf{H}^n$ is a sum

$$f = \sum_{|I|=q} f_I d\bar{z}^I,$$

where the f_I are complex-valued functions on $\mathbf{H}^n$.

To define $\bar{\partial}_b$, we do it first when f is a function (i.e., a 0-form) by setting

$$\bar{\partial}_b f = \sum_{j=1}^{n} \bar{Z}_j(f)\, d\bar{z}_j,$$

where $\bar{Z}_j$ are the left-invariant Cauchy-Riemann operators

$$\bar{Z}_j(f) = \frac{\partial f}{\partial \bar{z}_j} - iz_j \frac{\partial f}{\partial t}$$

that were considered in §2.6 of the previous chapter. When f is a q-form, we extend this by taking

$$\bar{\partial}_b\Big(\sum_{|I|=q} f_I\, d\bar{z}^I\Big) = \sum_{j=1}^{n} \sum_{|I|=q} \bar{Z}_j(f_I)\, d\bar{z}_j \wedge d\bar{z}^I. \tag{12}$$

Recalling that $[\bar{Z}_j, \bar{Z}_k] = 0$ for $j \neq k$, we see that $\bar{\partial}_b^2 = 0$.

Note that the definition (12) is in fact determined by two requirements. First, if f is a function, then $\langle \bar{\partial}_b f, \overline{W} \rangle = \overline{W}(f)$, whenever $\overline{W}$ is a tangential Cauchy-Riemann operator; here $\langle \cdot, \cdot \rangle$ denotes the dual pairing between 1-forms and vector fields. Second, when f is a p-form and g is a q-form, then

$$\bar{\partial}_b(f \wedge g) = (\bar{\partial}_b f) \wedge g + (-1)^{p+q} f \wedge (\bar{\partial}_b g).$$

On the space of q-forms whose coefficients are in $L^2(\mathbf{H}^n)$, we have the inner product

$$\langle f, g \rangle_q = \sum_{|I|=q} \int_{\mathbf{H}^n} f_I\, \bar{g}_I\, dz\, dt$$

where $f = \sum f_I\, d\bar{z}^I$, $g = \sum g_I\, d\bar{z}^I$. We then define the formal adjoint $\bar{\partial}^*$ of $\bar{\partial}$ by the identity

$$\langle \bar{\partial}_b^* f, g \rangle_{q-1} = \langle f, \bar{\partial}_b g \rangle_q, \tag{13}$$

for all $(q-1)$-forms g that are smooth and have compact support.

To give an explicit formula for $\bar{\partial}_b^*$, it is convenient to define the "interior" product. It lowers the order of a form by 1, is denoted by $\lrcorner$, and is given by

$$d\bar{z}_j \lrcorner d\bar{z}_I = 0,$$

if $j \notin I$, and

$$d\bar{z}_j \lrcorner d\bar{z}_I = (-1)^{\ell-1} d\bar{z}_{j_1} \wedge \cdots \wedge d\bar{z}_{j_{\ell-1}} \wedge d\bar{z}_{j_{\ell+1}} \wedge \cdots \wedge d\bar{z}_{j_q}, \tag{14}$$

if $j = j_\ell \in I$. The formula for $\bar{\partial}_b^*$ is then

$$\bar{\partial}_b^*\Big(\sum_{|I|=q} f_I\, d\bar{z}^I\Big) = -\sum_{j=1}^{n} \sum_{|I|=q} Z_j(f_I)\, d\bar{z}_j \lrcorner d\bar{z}^I. \tag{15}$$

This follows directly from (14) and the observation that

$$\int_{\mathbf{H}^n} \bar{Z}_j(F)\, \bar{G}\, dz\, dt = -\int_{\mathbf{H}^n} F(\overline{Z_j G})\, dz\, dt$$

whenever F and G are smooth functions on $\mathbf{H}^n$, with G having compact support. Note that when f is a 1-form, $f = \sum f_j\, d\bar{z}_j$,

$$\bar{\partial}_b^*(f) = -\sum_{j=1}^{n} Z_j(f_j).$$

2.2 $\Box_b$ and $\mathfrak{L}_\alpha$. The operator

$$\Box_b = \bar\partial_b^*\bar\partial_b + \bar\partial_b\bar\partial_b^*,$$

which maps q-forms to q-forms, will be the focus of our interest.

We shall see that $\Box_b$ can be expressed in terms of a family of operators that occur as analogues (for the Heisenberg group) of the Laplacian $\Delta = \sum_{j=1}^n \partial^2/\partial x_j^2$ on $\mathbf{R}^n$. The operators we have in mind are denoted by $\mathfrak{L}_\alpha$, where α is a parameter, and are given by

$$\mathfrak{L}_\alpha = -\frac{1}{2}\sum_{j=1}^n (Z_j\bar Z_j + \bar Z_j Z_j) + i\alpha T. \tag{16}$$

We shall see that $\Box_b$, when restricted to q-forms, equals $\mathfrak{L}_{n-2q}$.

PROPOSITION. *We have the identity*

$$\Box_b\Big(\sum_{|I|=q} f_I\, d\bar z^I\Big) = \sum_{|I|=q} \mathfrak{L}_\alpha(f_I)\, d\bar z^I, \tag{17}$$

where

$$\alpha = n - 2q. \tag{18}$$

2.2.1 Let us remark that the $\mathfrak{L}_\alpha$ satisfy symmetry properties analogous to those of Δ on $\mathbf{R}^n$. Indeed, the latter operator can be characterized (up to a constant multiple) by the following invariance properties. It is a differential operator on $\mathbf{R}^n$ that:

(i) is translation-invariant (i.e., has constant coefficients),

(ii) has degree 2 with respect to dilations, and

(iii) is rotation-invariant.

Similarly, we have that $\mathfrak{L}_\alpha$:

(i) is left-invariant on $\mathbf{H}^n$,

(ii) has degree 2 with respect to the dilation automorphisms of $\mathbf{H}^n$ (given by (4) in Chapter 12), and

(iii) is invariant under the unitary rotations (5) of Chapter 12.

It can be shown that any differential operator that enjoys these symmetry properties is a constant multiple of one of the $\mathfrak{L}_\alpha$; see §7.3 below.

We note here that (see §2.6 of the previous chapter)

$$\bar Z_j = \frac{1}{2}(X_j + iY_j) \qquad \text{and} \qquad Z_j = \frac{1}{2}(X_j - iY_j),$$

so $\mathfrak{L}_\alpha$ can be rewritten as

$$\mathfrak{L}_\alpha = -\frac{1}{4}\sum_{j=1}^n (X_j^2 + Y_j^2) + i\alpha T. \tag{19}$$

2.2.2 The proof of the proposition is a simple exercise involving the formulae in §2.1 above. Indeed, using first (12) and then (15), we see that if $f = f_I\, d\bar z^I$ then

$$\bar\partial_b^*\bar\partial_b(f) = -\sum_{k=1}^n\sum_{\ell=1}^n Z_k\bar Z_\ell(f_I)\, d\bar z_k \lrcorner (d\bar z_\ell \wedge d\bar z^I).$$

Similarly, using (12) and (15) in the reverse order, we get

$$\bar\partial_b\bar\partial_b^*(f) = -\sum_{\ell=1}^n\sum_{k=1}^n \bar Z_\ell Z_k(f_I)\, d\bar z_\ell \wedge (d\bar z_k \lrcorner d\bar z^I).$$

Now observe that when $\ell \neq k$, $\bar Z_\ell$ and Z_k commute: $Z_k\bar Z_\ell = \bar Z_\ell Z_k$; moreover, in that case we have

$$d\bar z_k \lrcorner (d\bar z_\ell \wedge d\bar z^I) = -d\bar z_\ell \wedge (d\bar z_k \lrcorner d\bar z^I).$$

Also, when $k = \ell$, $d\bar z_k \lrcorner (d\bar z_k \wedge d\bar z^I) = d\bar z^I$ or 0, depending on whether $k \notin I$ or $k \in I$. Similarly $d\bar z_k \wedge (d\bar z_k \lrcorner d\bar z^I) = d\bar z^I$ or 0, depending on whether $k \in I$ or $k \notin I$. As a result,

$$\Box_b(f_I\, d\bar z^I) = -\Big(\sum_{k\in I} \bar Z_k Z_k + \sum_{k\notin I} Z_k\bar Z_k\Big)(f_I)\, d\bar z^I. \tag{20}$$

However, by the commutation relations $[\bar Z_k, Z_k] = 2iT$, which hold for each k (see (44) in the previous chapter), we see that

$$\bar Z_k Z_k = \frac{1}{2}(\bar Z_k Z_k + Z_k\bar Z_k) + iT.$$

Similarly, $Z_k\bar Z_k = (1/2)(\bar Z_k Z_k + Z_k\bar Z_k) - iT$. Therefore the operator

$$-\Big(\sum_{k\in I} \bar Z_k Z_k + \sum_{k\notin I} Z_k\bar Z_k\Big)$$

appearing in (20) can be rewritten as

$$-\frac{1}{2}\Big(\sum_{k=1}^n \bar Z_k Z_k + Z_k\bar Z_k\Big) + i(-q + n - q)T.$$

This gives us $\Box_b(f_I\, d\bar z^I) = \mathfrak{L}_\alpha(f_I)\, d\bar z^I$ with $\alpha = n - 2q$, and the proposition is proved.

2.3 Fundamental solution for $\mathfrak{L}_\alpha$. We now turn to the question of the invertibility of $\Box_b$ and $\mathfrak{L}_\alpha$, and more precisely to the existence of fundamental solutions. Our considerations here will be divided into two parts. First, using heuristic arguments we will gain a clearer idea of what to expect. With this in hand we will then derive the actual results in question.

2.3.1 For each $\lambda \neq 0$, we have (see Chapter 12, §6.1) the irreducible unitary representation R^λ on $\mathbf{H}^n$ given by

$$(R^\lambda[z,t]f)(x) = e^{2\pi i\lambda[u\cdot x + u\cdot v/2 + t/4]}\, f(x+v)$$

where $z = u + iv$. If one passes to the corresponding representation of the Lie algebra $\mathfrak{h}^n$, using the basis X_j, Y_j, T, we get that these basis elements are represented by

$$\frac{\partial}{\partial u_j} R^\lambda[z,t]\big|_{[z,t]=[0,0]}, \qquad \frac{\partial}{\partial v_j} R^\lambda[z,t]\big|_{[0,0]}, \qquad \text{and} \quad \frac{\partial}{\partial t} R^\lambda[z,t]\big|_{[0,0]},$$

respectively. That is, X_j, Y_j, and T are represented by

$$f \mapsto 2\pi i\lambda x_j f(x), \qquad \frac{\partial f}{\partial x_j}, \qquad \text{and} \quad \frac{\pi i\lambda}{2} f,$$

respectively. This means that $\mathfrak{L}_\alpha$ (see (19)) is represented by

$$-\frac{1}{4}(\Delta f - 4\pi^2\lambda^2|x|^2 f) - \frac{\pi\alpha\lambda}{2} f. \tag{21}$$

Conjugating this operator with the change of scale $x \mapsto |2\pi\lambda|^{1/2}x$ gives

$$(-\pi/2)\,|\lambda|\,(\Delta f - |x|^2 f - \alpha\,\mathrm{sign}(\lambda) f). \tag{22}$$

This is essentially the n-dimensional Hermite operator. Knowledge of its spectrum (see Chapter 12, §6.4) tells us that (22) is invertible exactly when $\alpha\,\mathrm{sign}(\lambda) \neq -n, -n-2, \ldots$. So we may expect that $\mathfrak{L}_\alpha$ is invertible when $\pm\alpha \neq n, n+2, \ldots$.

2.3.2 We refer to the values $\alpha = \pm n, \pm(n+2), \ldots$ as the *forbidden* values for $\mathfrak{L}_\alpha$. Further confirmation that $\mathfrak{L}_\alpha$ cannot be expected to be invertible for these values is seen when $\alpha = n$. Then, according to the proposition in §2.2, $\mathfrak{L}_\alpha = \Box_b$ on functions, i.e.,

$$\mathfrak{L}_n f = \bar\partial_b^*\bar\partial_b f,$$

since $\bar\partial_b^* f = 0$ automatically. Alternatively, by the commutation relations,

$$\mathfrak{L}_n = -\sum_{j=1}^n Z_j\bar Z_j.$$

From either representation of $\mathfrak{L}_n$ it is clear that it annihilates those f on $\mathbf{H}^n$ that occur as boundary values of holomorphic functions on $\mathcal{U}^n$; hence $\mathfrak{L}_n$ has a large null-space, even when restricted to L^2 (see also §4.1.1 below).

2.3.3 Turning our attention to those α that are not forbidden, we seek a *fundamental solution* of $\mathfrak{L}_\alpha$, that is, a function F_α, smooth away from the origin, so that

$$\mathfrak{L}_\alpha(F_\alpha) = \delta_0 \tag{23}$$

in the sense of distributions, where δ_0 is the Dirac delta function at the origin.

In the situation of the Laplacian Δ on $\mathbf{R}^n$, $n \geq 3$, we have (as is well known) that $\Delta F = \delta_0$ for $F(x) = c_n|x|^{2-n}$. Notice that $|x|^{2-n}$ reflects the characteristic symmetry properties that were discussed in §2.2.1 above. Similarly, we might expect F_α to be homogeneous of degree $-2n$, with respect to the dilations on the Heisenberg group (note that δ_0 is homogeneous of degree $-2n-2$ with respect to these dilations). Moreover, we might also expect that F_α is invariant under the unitary rotations of $\mathbf{H}^n$. Thus it could be surmised that F_a should be expressible in terms of the function $t+i|z|^2$, which occurs in the Cauchy-Szegö kernel (see §2.4 of the previous chapter), since that kernel was characterized by similar invariance considerations.

2.3.4 Formula for the fundamental solution. Our heuristics above are vindicated by the following result.

We define the function ϕ_α on $\mathbf{H}^n$ by

$$\phi_\alpha(z,t) = (|z|^2 - it)^{-(n+\alpha)/2}\,(|z|^2 + it)^{-(n-\alpha)/2}. \tag{24}$$

Observe that $|z|^2 - it$ and $|z|^2 + it$ always lie in the right half-plane, so we can use the principal branch of the power function in the above definition.

Clearly, ϕ_α is homogeneous of degree $-2n$ and satisfies

$$|\phi_\alpha(z,t)| \leq c\max(|z|, |t|^{1/2})^{-2n},$$

which guarantees that ϕ_α is locally integrable (see also Chapter 12, §2.5.1).

THEOREM 1. *We have*

$$\mathfrak{L}_\alpha\phi_\alpha = \gamma_\alpha\delta_0 \tag{25}$$

in the sense of distributions, where the constant γ_α has the value

$$\gamma_\alpha = \frac{2^{2-n}\,\pi^{n+1}}{\Gamma([n+\alpha]/2)\Gamma([n-\alpha]/2)}. \tag{26}$$

Remark. Given that $\mathfrak{L}_\alpha(\phi_\alpha)$ is automatically a distribution that is homogeneous of degree $-2n-2$, the theorem can be seen to be made up of two assertions. The first is that $\mathfrak{L}_\alpha\phi_\alpha(z,t) = 0$, whenever $(z,t) \neq (0,0)$, which means that this distribution is supported at the origin; by the homogeneity, it must be a constant multiple of the delta function.† The second is the determination of the constant γ_α. The significance of this computation can be appreciated when we recall that the function $1/\Gamma(s)$ vanishes when $s = 0, -1, -2, \ldots$; thus γ_α is zero exactly when α is a forbidden value.

2.3.5 To prove the theorem, we set

$$\phi_{\alpha,\varepsilon}(z,t) = (|z|^2 + \varepsilon^2 - it)^{-(n+\alpha)/2} \cdot (|z|^2 + \varepsilon^2 + it)^{-(n-\alpha)/2},$$

with $\varepsilon > 0$. Note that $\phi_{\alpha,\varepsilon}$ is everywhere smooth and that $\phi_{\alpha,\varepsilon} \to \phi_\alpha$ (as $\varepsilon \to 0$) dominatedly, and hence in the sense of distributions.

Let $\psi_{\alpha,\varepsilon} = \mathfrak{L}_\alpha(\phi_{\alpha,\varepsilon})$. We will show that

$$\psi_{\alpha,\varepsilon} = \varepsilon^{-2n-2}\psi_{\alpha,1}(z/\varepsilon, t/\varepsilon^2), \tag{27}$$

where, in fact,

$$\psi_{\alpha,1}(z,t) = (n^2 - \alpha^2)(|z|^2 + 1 - it)^{-(n+\alpha+2)/2}\,(|z|^2 + 1 + it)^{-(n-\alpha+2)/2}.$$

Since $\psi_{\alpha,\varepsilon}(z,t) \to 0$ as $\varepsilon \to 0$, for $(z,t) \neq (0,0)$, and $\psi_{\alpha,1}$ is integrable, the family $\{\psi_{\alpha,\varepsilon}\}_{\varepsilon>0}$ forms a constant multiple of an approximation to the identity. Therefore, our theorem will follow with

$$\gamma_\alpha = \int_{\mathbf{H}^n} \psi_{\alpha,1}(z,t)\,dz\,dt. \tag{28}$$

Let us abbreviate the quantity $|z|^2 + \varepsilon^2 - it$ as μ. Then

$$Z_j(\mu^a) = 2a\bar{z}_j\mu^{a-1} \qquad \text{and} \qquad \bar{Z}_j(\mu^a) = 0.$$

By conjugation,

$$Z_j(\bar{\mu}^a) = 0 \qquad \text{and} \qquad \bar{Z}_j(\bar{\mu}^a) = 2az_j\bar{\mu}^{a-1}.$$

Also

$$T(\mu^a) = -ia\mu^{a-1} \qquad \text{and} \qquad T(\bar{\mu}^a) = ia\bar{\mu}^{a-1}.$$

It follows that

$$\sum_{j=1}^n Z_j\bar{Z}_j(\mu^a\bar{\mu}^b) = \mu^{a-1}\bar{\mu}^{n-1}(2nb\mu + 4ab|z|^2)$$

† See also Chapter 1, §8.19–§8.20 and Chapter 6, §7.5.

and

$$T(\mu^a\bar{\mu}^b) = i\mu^{a-1}\bar{\mu}^{b-1}(b\mu - a\bar{\mu}).$$

As a result, if we set $a = -(n+\alpha)/2$, $b = -(n-\alpha)/2$, and note that

$$\mathfrak{L}_\alpha = -\sum_{j=1}^n Z_j\bar{Z}_j + i(\alpha - n)T,$$

we find, after some simplifications, that

$$\psi_{\alpha,\varepsilon}(z,t) = \varepsilon^2(n^2 - \alpha^2)\mu^{-(n+\alpha+2)/2}\bar{\mu}^{-(n-\alpha+2)/2}.$$

This is exactly the assertion (27), with

$$\psi_{\alpha,1}(z,t) = (n^2 - \alpha^2)(|z|^2 + 1 - it)^{-(n+\alpha+2)/2}\,(|z|^2 + 1 + it)^{-(n-\alpha+2)/2}.$$

It remains to evaluate (28). This calculation can be thought of as an elaboration of the one used to obtain the constant in the Cauchy-Szegö kernel (see §2.3.3 of the previous chapter).

Let us define

$$I = \int_{\mathbf{H}^n} (|z|^2 + 1 - it)^{-\beta}\,(|z|^2 + 1 + it)^{-\gamma}\,dz\,dt$$

where $\beta = (n + \alpha + 2)/2$, $\gamma = (n - \alpha + 2)/2$, so $\beta + \gamma = n + 2$. Upon rescaling, we see that

$$I = \int_{\mathbf{C}^n} (|z|^2 + 1)^{-n-1}\,dz \cdot J, \tag{29}$$

where

$$J = \int_{-\infty}^{\infty} (1 - it)^{-\beta}\,(1 + it)^{-\gamma}\,dt.$$

To evaluate J, we make a preliminary restriction that $-n \le \alpha \le n$, which means that $\beta \ge 1$ and $\gamma \ge 1$. Recall that, when $\gamma > 0$,

$$(1 + it)^{-\gamma}\Gamma(\gamma) = \int_{-\infty}^{\infty} e^{-ixt} f(x)\,dx,$$

where $f(x) = e^{-x}x^{\gamma-1}$ for $x > 0$, and $f(x) = 0$ for $x < 0$. Similarly, if $\beta > 0$,

$$(1 - it)^{-\beta}\Gamma(\beta) = \int_{-\infty}^{\infty} e^{-ixt} g(x)\,dx,$$

where $g(x) = e^{-|x|}|x|^{\beta-1}$ for $x < 0$, and $g(x) = 0$ for $x > 0$.

Thus by Plancherel's theorem

$$\Gamma(\beta)\Gamma(\gamma)\int_{-\infty}^{\infty}(1-it)^{-\beta}(1+it)^{-\gamma}\,dt = 2\pi\int_0^{\infty} f(x)\,g(-x)\,dx$$
$$= 2\pi\int_0^{\infty} e^{-2x}\,x^{\beta+\gamma-2}\,dx = \frac{\pi n!}{2^n},$$

since $\beta+\gamma-2=n$. The result is that

$$J = \frac{2^{-n}\,\pi\,n!}{\Gamma(\beta)\,\Gamma(\gamma)}.$$

Next, $\int_0^\infty e^{-At}\,t^n\,dt = n!A^{-n-1}$ if $A>0$; therefore

$$\int_{\mathbf{C}^n}(1+|z|^2)^{-n-1}\,dz = \frac{\pi^{n+1}}{n!}\int_{\mathbf{C}^n}\left(\int_0^\infty e^{-\pi t(1+|z|^2)}\,t^n\,dt\right)dz$$
$$= \frac{\pi^{n+1}}{n!}\int_0^\infty e^{-\pi t}\,t^{-n}\cdot t^n\,dt = \frac{\pi^{n+1}}{n!}\cdot\pi^{-1} = \frac{\pi^n}{n!}.$$

Combining this with the value for J gives

$$I = \frac{2^{-n}\pi^{n+1}}{\Gamma(\beta)\Gamma(\gamma)} = \frac{2^{-n}\pi^{n+1}}{\Gamma(1+[n+\alpha]/2)\Gamma(1+[n-\alpha]/2)}$$
$$= \frac{2^{2-n}\pi^{n+1}}{\Gamma([n+\alpha]/2)\Gamma([n-\alpha]/2)}\cdot(n^2-\alpha^2)^{-1}.$$

However, according to (28), $\gamma_\alpha = (n^2-\alpha^2)I$, and thus (26) is established, when $-n\le\alpha\le n$. It then follows for all values of α by analytic continuation. This concludes the proof of Theorem 1.

2.4 Solvability and regularity of $\mathfrak{L}_\alpha$. When α is not a forbidden value, that is, when $\pm\alpha\ne n, n+2,\ldots$, we set $F_\alpha = \gamma_\alpha^{-1}\phi_\alpha$. We see that F_α is a fundamental solution of $\mathfrak{L}_\alpha$: one has that $\mathfrak{L}_\alpha(F_\alpha)=\delta_0$. We then define the solution operator S_α by

$$S_\alpha(f) = f * F_\alpha. \tag{30}$$

By the usual properties of convolutions, $S_\alpha(f)$ is well-defined in the following circumstances (among others): first, if f is a test function, then $S_\alpha(f)$ is a C^∞ function; second, if f is a distribution of compact support, then $S_\alpha(f)$ is a tempered distribution. In both senses, S_α is the inverse of $\mathfrak{L}_\alpha$. This is the thrust of the following proposition.

PROPOSITION. *We have that*

$$\mathfrak{L}_\alpha S_\alpha(f) = S_\alpha\mathfrak{L}_\alpha(f) = f \tag{31}$$

if either f is in $\mathcal{S}$, or f is a distribution of compact support.

2.4.1 To prove the proposition, we note first that whenever V is a left-invariant vector field

$$V(f*K) = f*(VK). \tag{32}$$

Let us check this formula first when (say) both f and K are in $\mathcal{S}$. We know that

$$(Vg)(x) = \frac{d}{ds}g(x\cdot\gamma(s))|_{s=0}$$

for an appropriate curve $\gamma(s)$ in $\mathbf{H}^n$. However

$$(f*K)(x\cdot\gamma(s)) = \int_{\mathbf{H}^n} f(y)\,K(y^{-1}x\cdot\gamma(s))\,dy,$$

and (32) follows for these f and K. A simple limiting argument then extends (32) to the situation where K is a tempered distribution and f is either in $\mathcal{S}$ or is a distribution of compact support.

If we apply this formula when $K=F_\alpha$ and V is Z_j, $\bar Z_j$, or T, and iterate this, we see that

$$\mathfrak{L}_\alpha(f*F_\alpha) = f*(\mathfrak{L}_\alpha F_\alpha) = f*\delta_0 = f.$$

Thus we obtain that

$$\mathfrak{L}_\alpha S_\alpha(f) = f.$$

We deduce from this the second identity: $S_\alpha\mathfrak{L}_\alpha(f)=f$. Observe that if $x=[z,t]$, then $x^{-1}=[-z,-t]$, and therefore $F_\alpha(x^{-1}) = F_{-\alpha}(x)$. Now if $f\in\mathcal{S}$, then

$$S_\alpha\mathfrak{L}_\alpha(f)(x) = \int_{\mathbf{H}^n}(\mathfrak{L}_\alpha f)(y)\,F_\alpha(y^{-1}x)\,dy$$
$$= \int_{\mathbf{H}^n}(\mathfrak{L}_\alpha f)(y)\,F_{-\alpha}(x^{-1}y)\,dy$$
$$= \int_{\mathbf{H}^n} f(y)\,(\mathfrak{L}_{-\alpha}F_{-\alpha})(x^{-1}y)\,dy.$$

This holds since an integration by parts implies that

$$\int_{\mathbf{H}^n}(\mathfrak{L}_\alpha f)(y)\,g(y)\,dy = \int_{\mathbf{H}^n} f(y)\,(\mathfrak{L}_{-\alpha}g)(y)\,dy.$$

As a consequence of the fact that $\mathfrak{L}_{-\alpha}(F_{-\alpha})=\delta_0$, we have then

$$S_\alpha\mathfrak{L}_\alpha(f) = f.$$

This identity extends to distributions of compact support, again by a simple limiting argument. The proposition is therefore proved.

2.4.2 The existence and regularity of solutions of $\mathfrak{L}_\alpha$ (and hence of $\Box_b$) will be expressed in terms of the notions of local solvability and hypoellipticity. These we now formulate. The operator $\mathfrak{L}_\alpha$ will be said to be *locally solvable* if, given any distribution f defined in a neighborhood of some point in $\mathbf{H}^n$, there exists a distribution u, defined in a neighborhood of the same point, so that

$$\mathfrak{L}_\alpha(u) = f. \tag{33}$$

The operator $\mathfrak{L}_\alpha$ is said to be *hypoelliptic* if, given a distribution u in the neighborhood of a point, when f is defined by (33) and is C^∞ in that neighborhood, then u is also C^∞ in that neighborhood.

PROPOSITION. *If $\pm\alpha \neq n, n+2, \ldots$, the operator $\mathfrak{L}_\alpha$ is both locally solvable and hypoelliptic.*

COROLLARY. *When $0 < q < n$, the operator $\Box_b$ is both locally solvable and hypoelliptic.*

The corollary follows from the proposition in §2.2 (the relevant α is $n - 2q$).

2.4.3 Starting with a distribution f defined near a given point, we can choose a suitable $\eta \in C_0^\infty$, so that $F = \eta f$ is a compactly supported distribution on $\mathbf{H}^n$ that agrees with f near our given point. Then we need only take $u = S_\alpha(f)$. Indeed, $\mathfrak{L}_\alpha(u) = \mathfrak{L}_\alpha S_\alpha(F) = F$, and we have solved (33) in a neighborhood of our given point.

Next suppose that u is given in a neighborhood N and $\eta \in C_0^\infty(N)$, with $\eta \equiv 1$ in a smaller neighborhood N_1. Let $U = \eta u$. If we assume that $\mathfrak{L}_\alpha(u)$ is C^∞ on N, then $\mathfrak{L}_\alpha(U) = F = f_1 + f_2$; here f_1 is a C_0^∞ function that agrees with f on N_1 and f_2 is a distribution of compact support that vanishes on N_1. Now $U = S_\alpha(F) = S_\alpha(f_1) + S_\alpha(f_2)$. The first term, $S_\alpha(f_1)$, is C^∞ everywhere. The second term can be written formally as

$$(S_\alpha f_2)(x) = \int_{\mathbf{H}^n} f_2(y)\, F_\alpha(x \cdot y^{-1})\, dy.$$

Note that $F_\alpha(x \cdot y^{-1})$ is smooth in x and y when $x \in N_1$ and $y \in \operatorname{supp}(f_2) \subset \overline{{}^c N_1}$, which shows that u is smooth in N_1. The proposition is therefore proved.

We remark that when α is a forbidden value, $\pm\alpha = n, n+2, \ldots$, the operator $\mathfrak{L}_\alpha$ is *not* hypoelliptic. This follows immediately from (25) and (26), because then $\mathfrak{L}_\alpha \phi_\alpha \equiv 0$, and ϕ_α is obviously not smooth at the origin. For these values of α, the question of local solvability is also of interest. We shall take this up in §4 below, when we deal with the unsolvable Lewy operator.

3. Applications of the fundamental solution

The further study of the regularity properties of $\mathfrak{L}_\alpha$ involves the boundedness of operators made up of products of the solution operator S_α with left-invariant vector fields on $\mathbf{H}^n$. Here we need to come to grips with a characteristic feature of the Heisenberg group, namely the role of the preferred "good directions" determined by the vector fields Z_j and $\bar{Z}_j$. From this perspective, the main conclusion is that $\mathfrak{L}_\alpha$, when $\pm\alpha \neq n, n+2, \ldots$, behaves in many ways like Δ does on $\mathbf{R}^n$ (where all directions are equally "good"). Exploitation of this principle leads us to the analogue of the L^p Sobolev spaces in this context, whose nonisotropic nature reflects the smoothing properties of the solution operator S_α.

3.1 We consider first the analogues of the classical Riesz transforms in $\mathbf{R}^n$. What has to be taken into account here is that the Z_j and $\bar{Z}_j$ (or their real and imaginary parts X_j and Y_j) occur quadratically in $\mathfrak{L}_\alpha$, while T occurs only linearly. This is of course consistent with the automorphic dilations of $\mathbf{H}^n$ and leads us to formulate the following proposition.

PROPOSITION 1. *Suppose $P(Z, \bar{Z})$ is a quadratic (noncommutative) polynomial in the Z_j and $\bar{Z}_j$, $j = 1, \ldots, n$, and that $\pm\alpha \neq n, n+2, \ldots$. Then the operator*

$$R = P(Z, \bar{Z}) S_\alpha, \tag{34}$$

initially defined on $\mathcal{S}$, has a bounded extension mapping $L^p(\mathbf{H}^n)$ to itself, for $1 < p < \infty$.

3.1.1 Proof. Suppose V_1 and V_2 are a pair of left-invariant vector fields. Then, according to (32),

$$V_1 V_2 S_\alpha(f) = f * K,$$

with $K = V_1 V_2(F_\alpha)$. However, F_α is a homogeneous function of degree $-2n$ that is smooth away from the origin. It follows from §5.2.1 of the previous chapter that, if V_1 and V_2 are of the type Z_j or $\bar{Z}_j$, then the distribution K satisfies the assumptions (80)–(83) of the theorem in §5.2 of that chapter. The proposition is therefore proved.

Remarks. (i) The analogous result holds for the operator TS_α. This is because the commutator relations give $[\bar{Z}_j, Z_j] = 2iT$, which expresses the vector field T as a quadratic polynomial in the Z_j and $\bar{Z}_j$.

(ii) A similar result is valid for $S_\alpha P(Z, \bar{Z})$. Also an analogue holds when the quadratic polynomial is split so that one factor appears to the left of S_α, and another to the right (e.g., the operator $Z_j S_\alpha \bar{Z}_j$). The proofs for these variants require the identity (38) below, but are otherwise similar to that of Proposition 1.

3.2 The nonisotropic Sobolev spaces. We come now to the appropriate versions of the Sobolev spaces for the Heisenberg group. Following the indications above, we formulate these so as to take into account that the differentiations Z_j and $\bar{Z}_j$ are to have order 1, while T is to have order 2.

The "nonisotropic" Sobolev space NL_k^p is the set of all $f \in L^p(\mathbf{H}^n)$ so that

$$P(Z, \bar{Z})f \in L^p(\mathbf{H}^n)$$

for all (noncommutative) polynomials P in Z_j and $\bar{Z}_j$ of degree at most k. To define the norm, let B_k be any (fixed) basis for the linear space of these polynomials and set

$$\|f\|_{NL_k^p} = \sum_{P \in B_k} \|P(Z, \bar{Z})f\|_{L^p(\mathbf{H}^n)}. \tag{35}$$

We make some preliminary comments about the spaces $NL_k^p(\mathbf{H}^n)$, touching on their relation with the standard Sobolev spaces $L_k^p(\mathbf{H}^n)$ (here we identify $\mathbf{H}^n$ with $\mathbf{R}^{2n+1}$) that were considered in §5 of Chapter 6. In view of the fact that the expressions $P(Z, \bar{Z})$ involve differentiations with polynomial coefficients that can be large at infinity, it is not really possible to make a comparison of NL_k^p with L_k^p that would be valid in unbounded regions of $\mathbf{H}^n$. For this reason the assertions below are limited to functions having compact support.

(a) First we have the obvious inclusion: If $f \in L_k^p$ then $f \in NL_k^p$, provided f has compact support.

(b) In the converse direction, we have a loss of half of the nonisotropic smoothness: If $f \in NL_k^p$, then $f \in L_{k/2}^p$, as long as f has compact support. This is evident when k is an even integer, because the "missing" direction T occurs as a quadratic polynomial in Z_j and $\bar{Z}_j$. It is also valid for all integers k (and is limited to $1 < p < \infty$), but that requires a subtler argument. Moreover, this inclusion relation is sharp: one cannot have $L_{k'}^p \subset NL_k^p$ (restricted to functions of compact support) if $k' > k/2$; see §7.5.

Note that we have defined the spaces NL_k^p for positive integral k only, while the L_k^p have been defined for all real k. We can now formulate the extension of Proposition 1.

PROPOSITION 2. *The operator R, given by* (34), *extends to a bounded mapping from NL_k^p to itself when $1 < p < \infty$, and $k = 0, 1, 2, \ldots$.*

3.2.1 The main difficulty involved in the proof is already evident when $k = 1$, and we shall consider this case in detail. Suppose then that $f \in NL_1^p$, and let $R(f) = f * K$. We need to show that $V(f*K) \in L^p$, where V is any one of the left-invariant vector fields Z_j or $\bar{Z}_j$, $1 \leq j \leq n$. Now according to (32)

$$V(f * K) = f * (VK). \tag{36}$$

However, in the above, the differentiation falls on the kernel K and not on the function f where it can be exploited; note that in general $f*(VK) \neq (Vf)*K$, because the Heisenberg group is not commutative.

The way around this obstacle is to consider, together with the left-invariant vector fields, the corresponding right-invariant vector fields. In this connection it is useful to adopt the following notation: whenever V is a left-invariant vector field, let V^R denote the right-invariant vector field that agrees with V at the origin. Since

$$(Vf)(x) = \frac{d}{ds} f(x \cdot \gamma(s))|_{s=0},$$

for an appropriate curve $s \mapsto \gamma(s)$, with $\gamma(0) = 0$ (see (36) of the previous chapter), we see that V^R comes about by a similar differentiation, where the x multiplication now occurs on the right:

$$(V^R f)(x) = \frac{d}{ds} f(\gamma(s) \cdot x)|_{s=0}.$$

Now the left and right actions in a group are interchanged by the inverse $x \mapsto x^{-1}$; this means that if we set $(Jf)(x) = f(x^{-1})$, then $J^{-1}VJ = -V^R$. Since $x^{-1} = [-z, -t]$, if $x = [z, t]$, we have, because of the formulas (39), (42), and (43) of the previous chapter, that

$$Z_j^R = \frac{\partial}{\partial z_j} - i\bar{z}_j \frac{\partial}{\partial t}, \quad \bar{Z}_j^R = \frac{\partial}{\partial \bar{z}_j} + iz_j \frac{\partial}{\partial t}, \quad T^R = \frac{\partial}{\partial t}. \tag{37}$$

Observe next that, as a counterpart of (36), we have

$$(Vf) * K = f * V^R(K). \tag{38}$$

Indeed, assuming first that f and K are test functions, we see that

$$\int_{\mathbf{H}^n} f(y \cdot \gamma(s))\, K(y^{-1} \cdot x)\, dy = \int_{\mathbf{H}^n} f(y)\, K(\gamma(s) \cdot y^{-1} \cdot x)\, dy,$$

if we make the appropriate change of y-variable. Differentiating both sides of the above identity with respect to s, and setting $s = 0$, establishes (38) for $f, K \in \mathcal{S}$. The general case of (38) follows by a limiting argument.

3.2.2 The key point of the argument is then to realize left differentiations as right differentiations, when these act on kernels of the type that appear in singular integrals, such as R. The kinds of kernels we are dealing with are those K that satisfy the following two conditions:

(i) K is a distribution that, away from the origin, agrees with a smooth homogeneous function of degree $-2n - 2$.

(ii) K satisfies the cancellation condition $|K(\Phi^R)| \leq A$ for all $R > 0$, whenever Φ is a normalized bump function (see also the broader conditions (80)–(83) in the previous chapter).

LEMMA. *Let V denote one of the Z_j or $\bar{Z}_j$. Suppose that K is a kernel that satisfies conditions (i) and (ii) above. Then there are kernels K_j' and K_j'', $1 \leq j \leq n$, each of which also satisfies these conditions, so that*

$$V(K) = \sum_{j=1}^{n} Z_j^R(K_j') + \bar{Z}_j^R(K_j''). \tag{39}$$

To prove the lemma, suppose for example that $V = Z_1$. Comparing the formula (37) with the corresponding formula (43) in the previous chapter, we see that

$$Z_1 - Z_1^R = 2i\bar{z}_i\frac{\partial}{\partial t}.$$

Moreover, $T^R = \partial/\partial t = (2i)^{-1}[Z_1^R, \bar{Z}_1^R]$, thus

$$Z_1(K) = Z_1^R(K) + Z_1^R\bar{Z}_1^R(\bar{z}_1 K) - \bar{Z}_1^R Z_1^R(\bar{z}_1 K).$$

Now if K satisfies our hypotheses it is easy to see that $Z_1^R(\bar{z}_1 K)$ does also. Indeed, the homogeneity and smoothness of the function corresponding to $Z_1^R(\bar{z}_1 K)$ are clear. Moreover

$$\langle Z_1^R(\bar{z}_1 K), \phi\rangle = -\langle \bar{z}_1 K, Z_1^R\phi\rangle = -K(\bar{z}_1 Z_1^R\phi),$$

and if ϕ is a normalized bump function, then $\bar{z}_1 Z_1^R\phi$ is a bounded multiple of a normalized bump function.

A similar argument works for the kernel $\bar{Z}_1^R(\bar{z}_1 K)$. Thus if $V = Z_1$, we obtain the desired conclusion (39) with

$$K_1' = K - \bar{Z}_1^R(\bar{z}_1 K), \qquad K_1'' = Z_1^R(\bar{z}_1 K),$$

and $K_j' = K_j'' = 0$ for $2 \leq j \leq n$. The proof for the other V's is of course similar, and the lemma is established.

To conclude the proof of Proposition 2, we again take up (36), namely that $V(f * K) = f * (VK)$. Because of (39), we write this as

$$\sum_j f * (Z_j^R K_j') + f * (\bar{Z}_j^R K_j'').$$

Next we invoke (38), which allows us to rewrite the sum as

$$\sum_j Z_j(f) * K_j' + (\bar{Z}_j f) * K_j''.$$

Notice that $f \in NL_k^p$ implies that $Z_j f$ and $\bar{Z}_j f$ are in L^p; also, since the kernels K_j' and K_j'' satisfy (i) and (ii) above, they also satisfy the hypotheses (80)–(83) of Chapter 12. As a result, $V(f * K) \in L^p$, and we have verified our assertion when $k = 1$. Iterating this argument establishes the proposition for general k.

The solution operator S_α has a defect: Although, morally speaking, it is smoothing of order 2 (in the sense above), it is not correct to say that it maps NL_k^p to NL_{k+2}^p. In fact, S_α is not bounded from L^p to itself, because the kernel F_α is not integrable at infinity. One way to remedy this shortcoming is as follows.

3.2.3 COROLLARY. *Suppose η and η_1 are two fixed C^∞ functions of compact support. Then the mapping*

$$f \mapsto \eta_1 \cdot S_\alpha(\eta f) \tag{40}$$

extends to a bounded operator from NL_k^p to NL_{k+2}^p.

This follows from Proposition 2 and two easy observations. First, the mapping $f \mapsto \eta f$ is bounded from NL_k^p to itself. From this we see that Proposition 2 already guarantees control of derivatives of the form

$$\eta_1 \cdot P(Z, \bar{Z}) \cdot S_\alpha(\eta f)$$

for P homogeneous of degree m, with $2 \leq m \leq k+2$.

We are therefore reduced to proving that the mapping (40) (with a possibly altered η_1) takes L^p to NL_1^p; that means we are required to show that operators of the form

$$f \mapsto \eta_1' Z_j(S_\alpha \eta f), \qquad f \mapsto \eta_1' \bar{Z}_j(S_\alpha \eta f), \qquad \text{and} \quad f \mapsto \eta_1' S_\alpha(\eta f)$$

are bounded from L^p to itself. However, each of these can be written as

$$f \mapsto \int f(y)\, L(x,y)\, dy,$$

where $L(x,y) = \eta_1'(x)\,\eta(y)\,M(y^{-1}x)$ with $M \in L^1(\mathbf{H}^n)$, from which their boundedness on L^p is clear.

3.3 Sharp regularity for $\mathfrak{L}_\alpha$. The spaces NL_k^p allow us to formulate in a precise sense the hypoellipticity of $\mathfrak{L}_\alpha$, which was already guaranteed by the proposition in §2.4.2. For this purpose it is convenient to use the terminology that a distribution u, defined in an open set U, is *locally in* NL_k^p if, for each $\eta \in C_0^\infty(U)$, $\eta \cdot u \in NL_k^p(\mathbf{H}^n)$. The regularity result below is the analogue of the interior regularity for elliptic equations (see Chapter 6, §7.13) except that here the maximum gain is restricted to the "good" directions.

THEOREM 2. *Assume $\pm\alpha \neq n, n+2, \ldots$ and suppose that u is a distribution given in an open set U.*

(a) *If $\mathfrak{L}_\alpha(u)$ is locally in NL_k^p, then u is locally in NL_{k+2}^p.*

(b) *In particular, if* $\eta_1, \eta_2 \in C_0^\infty(U)$ *and* $\eta_2 \equiv 1$ *in a neighborhood of* $\operatorname{supp}\eta_1$, *then*

$$\|\eta_1 u\|_{NL^p_{k+2}} \le c_{\eta_1,\eta_2}(\|\eta_2 \mathfrak{L}_\alpha(u)\|_{NL^p_k} + \|\eta_2 u\|_{L^p}). \tag{41}$$

3.3.1 For the proof of the theorem, it is convenient to use the notation $\eta_1 \prec \eta_2$ to mean that $\eta_2 \equiv 1$ in a neighborhood of $\operatorname{supp}\eta_1$. Given that $\eta_1 \prec \eta_2$, we can find an $\eta \in C^\infty$, so that $\eta_1 \prec \eta \prec \eta_2$. Next write $u_0 = \eta u$. Then

$$\mathfrak{L}_\alpha(u_0) = \eta\mathfrak{L}_\alpha(u) + \sum_{j=1}^{2n} \zeta_j V_j(u) + \zeta u,$$

where $V_j = Z_j$ for $1 \le j \le n$, $V_j = \bar{Z}_{j-k}$ for $n+1 \le j \le 2n$, and ζ, ζ_j are various combinations of derivatives of η.

Now by (31), $u_0 = S_\alpha \mathfrak{L}_\alpha(u_0)$, and therefore

$$\eta u = S_\alpha(\eta\mathfrak{L}_\alpha(u)) + \sum_{j=1}^{2n} S_\alpha(\zeta_j V_j(u)) + S_\alpha(\zeta u).$$

We multiply the above by η_1, noting that $\eta \cdot \eta_1 = \eta_1$. Thus $\eta_1 u$ is expressed as the sum of three kinds of terms.

The first is $\eta_1 S_\alpha(\eta\mathfrak{L}_\alpha(u)) = \eta_1 S_\alpha(\eta\eta_2\mathfrak{L}_\alpha(u))$ (since $\eta\cdot\eta_2 = \eta$). Here one can apply the corollary to assert that it belongs to $NL^p_{k+2}(\mathbf{H}^n)$, with norm at most $c\|\eta_2\mathfrak{L}_\alpha(u)\|_{NL^p_k}$.

The others are typified by the term $\eta_1 S_\alpha(\zeta_1 V_1(u))$. Observe here that $\operatorname{supp}\eta_1 \cap \operatorname{supp}\zeta_1 = \emptyset$, since ζ_1 is a derivative of η. This term can be written formally as

$$\eta_1(x) \int \zeta_1(y)\,(V_1 u)(y)\, F_\alpha(x\cdot y^{-1})\, dy. \tag{42}$$

Taking into account the restrictions on x and y guarantees that $x \cdot y^{-1}$ stays away from the origin, and hence that $F_\alpha(x\cdot y^{-1})$ is jointly smooth there; we see, therefore, that (42) represents a C^∞ function. Thus conclusion (a) of the theorem is proved. To prove conclusion (b), note that any number of derivatives of (42) with respect to x can be written as sums of terms of the form

$$-\eta_1'(x) \int u(y)\, V_1^y[\zeta(y)\, F_\alpha'(x\cdot y^{-1})]\, dy,$$

as an integration by parts shows. In the above, we can always insert the factor $\eta_2(y)$ (since $\zeta \prec \eta_2$), and thus any derivative of (42) is majorized by $\|\eta_2 u\|_{L^p}$. This establishes (41) and concludes the proof of the theorem.

4. The Lewy operator

Our final application of the Heisenberg group to complex analysis will be to elucidate the properties of the unsolvable Lewy operator. This operator, which was initially considered without reference to the Heisenberg group, is in fact the complex vector field

$$Z_1 = \frac{\partial}{\partial z_1} + i\bar{z}_1\frac{\partial}{\partial t},$$

on the group $\mathbf{H}^1$ (which we identify with $\mathbf{R}^3$); see (43) in the previous chapter.

We will try to find the conditions to be imposed on a distribution f so that the equation

$$Z_1(u) = f \tag{43}$$

is locally solvable. That is, we suppose f to be a distribution given near a point $x^0 \in \mathbf{H}^1$ and ask: What are the conditions on f, near x^0, that guarantee the existence of a distribution u, defined near x^0, so that (43) holds?

Recall our identification of $\mathbf{H}^1$ with the boundary $b\mathcal{U}^1$ of the domain $\mathcal{U}^1$; we then obtain a corresponding point on this boundary, which we also denote by x^0. The basic condition to be imposed will be a precise formulation of the following hypothesis:

$$\textit{The Cauchy-Szegö integral of } f \textit{ continues past } x^0 \in b\mathcal{U}^1. \tag{44}$$

To begin with, let f^0 be any compactly supported distribution that agrees with f near x^0. For any $z \in \mathcal{U}^1$, we wish to define the Cauchy-Szegö integral of f^0, which is formally given by

$$\int_{b\mathcal{U}^1} S(z,w)\, f^0(w)\, d\beta(w).$$

In fact, we define $\tilde{C}(f^0)$ to be the result of testing the distribution f^0 against the C^∞ function $w \mapsto S_z(w) = S(z,w)$, i.e.,

$$(\tilde{C}f^0)(z) = f^0(S_z). \tag{45}$$

Since the C^∞ function S_z depends analytically on $z \in \mathcal{U}^n$ (and the distribution f^0 has compact support), it is clear that $(\tilde{C}f^0)(z)$ is itself analytic in $z \in \mathcal{U}^n$. Note that if f^0 happens to be an L^2 function (of compact support), then $(\tilde{C}f^0)(z) = (Cf)(z)$, where Cf is as defined for L^2 functions in §2.3.1 of the previous chapter. The condition (44) is then interpreted to mean that

$$(\tilde{C}f^0)(z) \textit{ can be analytically continued past } x^0. \tag{46}$$

For this condition to be an intrinsic property of f, it must be independent of the particular distribution f^0 we have chosen; we require that f^0 has compact support and agrees with f near x^0. In fact, consider two such distributions, f^0 and f^1. Their difference vanishes near x^0 and has compact support. But if $S(z,w)$ is restricted to w in a fixed compact set that excludes $w = x^0$, then by the formula (20) of Chapter 12, it is clear that $z \mapsto S(z,w)$ continues analytically beyond x_0. Thus, if (46) holds then it also holds when f^0 is replaced by f^1. Our considerations also show that the condition (44) reflects only the behavior of f near x^0.

THEOREM 3. *The equation $Z_1(u) = f$ is solvable near x^0 if and only if condition* (44) *holds.*

4.1 From the point of view adopted above, the proof of the necessity is easy. Suppose that u is a solution of (43) near x^0, and let $\eta \in C_0^\infty$ with $\eta \equiv 1$ near x^0. Let $U = \eta u$. Then $f^0 = Z_1(U)$ is a compactly supported distribution that agrees with f in a neighborhood of x^0.

Now consider $(\tilde{C}f^0)(z)$. By definition, it is

$$f^0(S_z) = Z_1(U)(S_z).$$

Using the rule for differentiating distributions, we see that the latter is $-U(Z_1^w S_z)$; here $Z_1^w S_z$ denotes the Z_1 derivative of $S_z(w) = S(z,w)$ with respect to w. However, $w \mapsto S(z,w)$ is conjugate-analytic in w, and hence $Z_1^w S_z \equiv 0$,‡ from which it follows that $(\tilde{C}f^0)(z) \equiv 0$ in $\mathcal{U}^1$. Thus (46), and hence (44), is satisfied.

4.1.1 Some examples. Before giving the proof of the sufficiency of condition (44), we give some examples in which (44) is violated (and hence for which the equation (43) cannot be solved).

(a) For all positive integers k and N, define

$$F(z) = (z_2)^{k+1/2}(z_2+i)^{-N}.$$

Note that F is analytic in $\mathcal{U}^1$ and continuous up to the boundary. We let $f = F^b$ denote its restriction to $b\mathcal{U}^1$. If $N > k+3/2$ then $f \in L^2(b\mathcal{U}^1)$ (and $F \in \mathcal{H}^2(\mathcal{U}^1)$); hence $(\tilde{C}f)(z) = F(z)$. Note that $f \in C^k$, but F does not continue past the origin, because of the singularity of $(z_2)^{k+1/2}$ there. Thus we have constructed an $f \in C^k$ so that $Z_1(u) = f$ has no solution near the origin.

(b) A slight variation of (a) gives us an f that is C^∞ (in fact, we have $f \in \mathcal{S}$), for which (43) is not solvable near the origin. We need only take f to be the restriction to $b\mathcal{U}^1$ of the function

$$F(z) = e^{-(z_2/i)^{-1/2}}\, e^{-(1+z_2/i)^{1/2}}.$$

‡ See Chapter 12, §2.6.3.

The factor $e^{-(z_2/i)^{-1/2}}$ has a zero of infinite order at the origin, and is smooth in $\overline{\mathcal{U}}^1$. The factor $e^{-(1+z_2/i)^{1/2}}$ guarantees the rapid decrease of F on $\overline{\mathcal{U}}^1$.

(c) As a last example, we describe a further modification of the above that yields an $f \in \mathcal{S}$ so that $Z_1(u) = f$ is not solvable near *any* point. For this purpose, we construct an F that is holomorphic in $\mathcal{U}^1$, is smooth and rapidly decreasing on $\overline{\mathcal{U}}^1$, and which does not continue past any boundary point of $\mathcal{U}^1$.

We begin with a further examination of the function

$$E(w) = e^{-(w/i)^{-1/2}},$$

which is defined for complex numbers w lying in any proper obtuse sector S that contains the upper half-plane $\mathrm{Im}(w) > 0$ (see Figure 1).

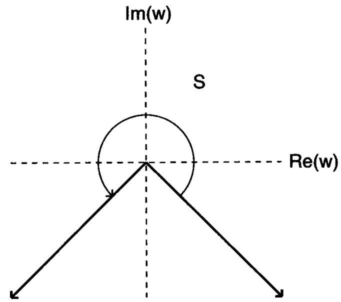

Figure 1. The open sector S.

Note that E is holomorphic and bounded in S and is smooth at the origin. Let us define the constants M_m by

$$M_m = \sup\{|E^{(m)}(w)| : \mathrm{Im}(w) > 0\}/m!\,. \tag{47}$$

Similarly, for $r > 0$, define

$$M_m(r) = \sup\{|E^{(m)}(w)| : \mathrm{Im}(w) > 0, |w| \geq r\}/m!\,. \tag{48}$$

Since E is bounded in S, Cauchy's inequality gives

$$M_m(r) \leq A^{-m} r^{-m},$$

for some constant $A > 0$. Next, since E does not continue past the origin,

$$\limsup_{m\to\infty} (M_m)^{1/m} = \infty,$$

again by Cauchy's inequality. Thus there exists a sequence m_j with $(M_{m_j})^{1/m_j} \to \infty$ and, in view of (48), there is a sequence $w_j \to 0$ in the upper half-plane so that

$$|E^{(m_j)}(w_j)| \geq \frac{3}{4} \cdot m_j!\, M_{m_j}.$$

Now set

$$G(z) = \sum_{k=0}^{\infty} 3^{-k}\, g(h_k(z)) \tag{49}$$

where $g(z) = g(z_1, z_2) = e^{-(z_2/i)^{-1/2}}$, and $\{h_k\}$ is a sequence of distinct points in $\mathbf{H}^1$ that is dense.

Observe that G is automatically holomorphic in $\mathcal{U}^1$, and smooth and bounded in $\overline{\mathcal{U}}^1$. We now show that it does not continue past any point of $b\mathcal{U}^1$. It suffices to show that G is not holomorphic in a fixed neighborhood of $h_{k_0}^{-1}(0)$, for each k_0. So fix k_0. Observe that

$$\sum_{k<k_0} 3^{-k}\, g(h_k(z))$$

is analytic near $h_{k_0}^{-1}(0)$ since g is analytic away from the origin in $b\mathcal{U}^1$.

Therefore, it suffices to consider

$$G_{k_0}(z) = 3^{-k_0}\, g(h_{k_0}(z)) + \sum_{k>k_0} 3^{-k}\, g\big(h_k(z)\big). \tag{50}$$

We take the derivative $(\partial/\partial z_2)^{m_j}$ of G_{k_0} in a (fixed) small neighborhood of $h_{k_0}^{-1}(0)$ in $\overline{\mathcal{U}}^1$, and consider in particular its value at the points $z = h_{k_0}^{-1}(0, w_{m_j})$. For large j, the absolute value of the first term in (50) is essentially

$$m_j!\, M_{m_j}\, 3^{-k_0}.$$

The sum of the absolute values of the other terms in (50) is at most

$$m_j!\, M_{m_j} \sum_{k>k_0} 3^{-k} = \frac{m_j!\, M_{m_j}\, 3^{-k_0}}{2}$$

and hence

$$\left| \left(\frac{\partial}{\partial z_2}\right)^{m_j} G_{k_0}(z) \right| \geq \frac{3^{-k_0}}{4} \cdot m_j!\, M_{m_j},$$

if $z = h_{k_0}^{-1}(0, w_{m_j})$. Consequently, $G_{k_0}(z)$ does not continue past $h_{k_0}^{-1}(0)$.

Thus $G(z)$ does not continue past any boundary point of $\mathcal{U}^1$, and we may take

$$F(z) = G(z)\, e^{-(1+z_2/i)^{1/2}}.$$

4.2 The relative fundamental solution. For the proof of the sufficiency part of Theorem 3, we return to the study of $\Box_b$ on $\mathbf{H}^n$, and analyze further the situation in the "forbidden" case corresponding to the action of $\Box_b$ on functions (or 0-forms); to emphasize this point, we write ${\Box_b}^{(0)}$ for $\Box_b$ acting on functions. Recall that, in this case, according to §2.2 we have

$${\Box_b}^{(0)} f = \mathfrak{L}_n f.$$

The relation with Z_1, when $n = 1$, comes about because

$$\mathfrak{L}_1 = -Z_1 \bar{Z}_1$$

on $\mathbf{H}^1$.

While the operator $\mathfrak{L}_n$ on $\mathbf{H}^n$ does not have a fundamental solution, it does have one "relative to its null space". We shall first state the results related to this, and then will interpret them.

Consider first the function Φ on $\mathbf{H}^n$, defined by

$$\Phi(z,t) = \frac{2^{n-2}(n-1)!}{\pi^{n+1}} \cdot \log\left(\frac{|z|^2 - it}{|z|^2 + it}\right)(|z|^2 - it)^{-n}. \tag{51}$$

Here

$$\log\left(\frac{|z|^2 - it}{|z|^2 + it}\right) = \log(|z|^2 - it) - \log(|z|^2 + it),$$

with the logarithms taken to be their principal branches in the right half-plane. It is easily verified that Φ is homogeneous of degree $-2n$ and is smooth away from the origin.

PROPOSITION.

$$\mathfrak{L}_n \Phi = \delta_0 - K, \tag{52}$$

where K is the convolution kernel of the Cauchy-Szegö projection, given by (29) of the previous chapter.

Let us define the relative solution operator $\tilde{S}$ by

$$\tilde{S}(f) = f * \Phi;$$

similarly, we have the Cauchy-Szegö projection operator C, given by

$$C(f) = f * K.$$

COROLLARY. *We have that*

$$\mathfrak{L}_n \tilde{S}(f) = \tilde{S}\mathfrak{L}_n(f) = f - C(f), \tag{53}$$

if either $f \in \mathcal{S}$, or f is a compactly supported distribution.

Note that (53) is the substitute for (31) when $\alpha = n$, and that, in the same sense, (52) is the substitute for (25). What it gives us is the inverse of $\mathfrak{L}_n$, not on all functions, but for those orthogonal to its "null" space, when properly restricted.

To see this, observe that if f is a test function, then $\mathfrak{L}_n f = 0$ if and only if f is the boundary restriction of a holomorphic function in $\mathcal{U}^n$, that is, when $f = C(f)$. Indeed, if $\mathfrak{L}_n(f) = 0$ then $\langle \mathfrak{L}_n f, f\rangle = 0$; since $\mathfrak{L}_n = -\sum_{j=1}^{n} Z_j \bar{Z}_j$, we obtain

$$\sum_{j=1}^{n} \langle \bar{Z}_j f, \bar{Z}_j f\rangle = 0,$$

which implies $\bar{Z}_j(f) = 0$ for $j = 1, \ldots, n$. Thus f arises as the boundary values of a holomorphic function. Conversely, it is clear that if $\bar{Z}_j(f) = 0$, $j = 1, \ldots, n$, then $\mathfrak{L}_n(f) = 0$.

Further motivation for (53), and why in fact it should be derivable from (31) for α near n, may be found in §7.4 below.

4.2.1 Let us go back to the identity in §2.5.4, namely

$$\mathfrak{L}_\alpha \phi_\alpha = \gamma_\alpha \delta_0, \qquad \gamma_\alpha = \frac{2^{2-n}\, \pi^{n+1}}{\Gamma([n+\alpha]/2)\Gamma([n-\alpha]/2)}, \tag{54}$$

where $\phi_\alpha(z,t) = (|z|^2 - it)^{-(n+\alpha)/2}\, (|z|^2 + it)^{-(n-\alpha)/2}$.

We note two facts. First, $\alpha \mapsto \phi_\alpha$ is a smooth mapping to functions that are homogeneous of degree $-2n$, and hence we can consider ϕ_α to be a distribution-valued function of α, depending smoothly on α. Next, since $s\Gamma(s) = \Gamma(s+1)$, we have that $\Gamma(s)^{-1} \sim s$ as $s \to 0$, and therefore

$$\left.\frac{d\gamma_\alpha}{d\alpha}\right|_{\alpha=n} = -\frac{2^{1-n}\pi^{n+1}}{\Gamma(n)} = -\frac{2^{1-n}\pi^{n+1}}{(n-1)!}.$$

Let a denote this nonzero constant.

We now differentiate both sides of (54) in α and set $\alpha = n$. Recalling that $\mathfrak{L}_\alpha = \sum_{j=1}^{n} (Z_j \bar{Z}_j + \bar{Z}_j Z_j) + i\alpha T$, we see that $\partial \mathfrak{L}_\alpha / \partial \alpha = iT$ and hence

$$\mathfrak{L}_n \left(\left.\frac{\partial \phi_\alpha}{\partial \alpha}\right|_{\alpha=n} \right) + iT(\phi_n) = a \cdot \delta_0.$$

This turns out to be the identity (52). To see this, note that

$$\left.\frac{d\phi_\alpha}{d\alpha}\right|_{\alpha=n} = \frac{(|z|^2 - it)^{-n}}{2} \cdot \left[-\log(|z|^2 - it) + \log(|z|^2 + it) \right],$$

and so $a^{-1} \cdot \partial\phi_\alpha/\partial\alpha|_{\alpha=n} = \Phi$, as given by (51).

Next,

$$\frac{-i}{a} T(\phi_n) = \frac{i2^{n-1}}{\pi^{n+1}} \cdot (n-1)! \frac{\partial}{\partial t} (|z|^2 - it)^{-n} = -K,$$

according to formula (29) of the previous chapter, in which

$$c = \frac{2^{n-1} i^{n-1} n!}{\pi^{n+1}}.$$

Therefore, (52) is established.

The identity (53) follows from (52) in the same way as the corresponding result for S_α was proved in §2.4.

4.2.2 We can now return to the setting of $\mathbf{H}^1$ and the sufficiency of the condition in Theorem 3. Assume therefore that (44) holds. Thus if f^0 agrees with f near x^0, and f^0 has compact support, then $(\tilde{C} f^0)(z)$ continues past x^0. Now write

$$f^0 = C(f^0) + (1-C)(f^0) = f_1 + f_2,$$

where $C(f^0)$ is the distribution $f^0 * K$ on $\mathbf{H}^1$. Since the function $(\tilde{C} f^0)(z)$ continues past x^0, it follows easily that the distribution $C(f^0)$ agrees (in a neighborhood of x^0 in $\mathbf{H}^1$) with the continuation of this function. As a result, $f_1 = C(f^0)$ is a real-analytic function near x^0.

The problem of solving $Z_1(u) = f$ can be reduced to the twin problem of solving

$$Z_1(u_1) = f_1 \tag{55}$$

and

$$Z_1(u_2) = f_2 = (1-C)f^0. \tag{56}$$

Since f^1 is real-analytic near x^0, equation (55) is solvable, by the Cauchy-Kovalevski theorem. A solution to (56) is given by $u_2 = -\bar{Z}_1 \tilde{S}(f^0)$. Indeed,

$$Z_1 u_2 = -Z_1 \bar{Z}_1 \tilde{S}(f^0) = \mathfrak{L}_1 \tilde{S}(f^0) = (1-C)(f^0) = f_2,$$

using (53). Since the question of solving $Z_1(u) = f$ near x^0 is the same as the problem for $Z_1(u) = f^0$, combining the above solutions gives a local solution to $Z_1(u) = f$.

4.2.3 Our discussion of the solvability of the Lewy equation in $\mathbf{H}^1$ has in reality dealt with the solvability of $\Box_b$ for 0-forms on $\mathbf{H}^n$. One can state the following: A necessary and sufficient condition for $\Box_b{}^{(0)}(u) = f$ to be solvable near a point $x^0 \in \mathbf{H}^n$ is that condition (44) holds. Indeed, on $\mathbf{H}^n$ we have $\Box_b{}^{(0)} = \mathfrak{L}_n$, and hence the assertion just made has a proof which is very close to that given for Theorem 3. The details may be left to the interested reader.

5. Homogeneous groups

The Heisenberg group is the simplest example (besides the usual additive group $\mathbf{R}^n$) of a class of groups which have acquired increasing importance in various problems in analysis. These are the groups that are "homogeneous" in the sense that they admit a natural family of scalings.

Our goal here, with regard to these groups, is by necessity a modest one: to present enough of the facts about homogeneous groups so as to be able to extend the basic L^p theory of singular integrals to this context.

5.1 Briefly put, a *homogeneous group* consists of $\mathbf{R}^n$ equipped with a Lie group structure, together with a family of dilations that are group automorphisms.

To be precise, we shall assume that we are given a pair of mappings:

$$[(x,y) \mapsto x \cdot y] : \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}^n \qquad \text{and} \qquad [x \mapsto x^{-1}] : \mathbf{R}^n \to \mathbf{R}^n$$

that are smooth and so that $\mathbf{R}^n$, together with these mappings, forms a group. We always assume that the origin is the identity. Next, we suppose that we are given an n-tuple of strictly positive exponents $a_1, \dots, a_n$, so that the *dilations*

$$x = (x_1, \dots, x_n) \mapsto \delta \circ x = (\delta^{a_1} x_1, \dots, \delta^{a_n} x_n)$$

are group automorphisms, for all $\delta > 0$.

5.1.1 As a first consequence of the definition, we observe that group multiplication is given by a polynomial; that is, $x \cdot y = P(x,y)$ for some polynomial $P : \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}^n$.

Indeed, if we write $P_k(x,y) = (x \cdot y)_k$, $P = (P_k)_{k=1}^n$, we have that

$$P_k(\delta \circ x, \delta \circ y) = \delta^{a_k} P_k(x,y), \qquad \text{for } \delta > 0,\ k = 1, \dots, n;$$

this is because $\delta \circ P(x,y) = P(\delta \circ x, \delta \circ y)$, in view of the fact that dilations are automorphisms. Thus each P_k is homogeneous of degree a_k

with respect to the dilations. Note next that, as a result, $(\partial/\partial x_j)P_k(x,y)$ is homogeneous of degree $a_k - a_j$, and similarly

$$\left(\frac{\partial}{\partial x}\right)^\alpha \left(\frac{\partial}{\partial y}\right)^\beta P_k(x,y)$$

is homogeneous of degree $a_k - \sum_{j=1}^n (\alpha_j + \beta_j) a_j$. Consequently, this degree is strictly negative whenever

$$(|\alpha| + |\beta|) \min_j \{a_j\} > a_k;$$

when this holds, we must have that $(D_x)^\alpha (D_y)^\beta P_k \equiv 0$, for otherwise $(D_x)^\alpha (D_y)^\beta P_k$ would be unbounded near the origin. Our assertion is therefore established.

Because the origin corresponds to the group identity, we also have that $P(0,0) = 0$, $P(x,0) = x$, and $P(0,y) = y$. As a result

$$x \cdot y = x + y + Q(x,y), \tag{57}$$

where Q is a polynomial with $Q(0,0) = Q(x,0) = Q(0,y) = 0$. Thus Q contains no pure monomials in x or y; all its terms are mixed and its degree (in the usual, isotropic, sense) is at least 2. Moreover, writing $Q = (Q_1, \dots, Q_n)$, we have that Q_k is homogeneous of degree a_k (with respect to the nonisotropic dilations specific to our group).

5.1.2 We can next assert that the Euclidean measure dx is both left and right invariant with respect to the group multiplication; i.e., it is Haar measure.

To see this, relabel the coordinates so that the exponents a_j are increasing, $a_1 \le a_2 \le \dots \le a_n$. Now consider right translations, that is, mappings of the form $x \mapsto x \cdot y = x + y + Q(x,y)$, where y is fixed. We need to compute the Jacobian matrix $\partial(x \cdot y)/\partial x$.

By the homogeneity argument above, $\partial Q_k(x,y)/\partial x_j$ has negative degree (and hence vanishes) whenever $j > k$; moreover, $\partial Q_k(x,y)/\partial x_k$ also vanishes when $j = k$, because Q_k does not contain any pure monomials in x. Therefore $\partial(x \cdot y)/\partial x$ is a triangular matrix with ones along the diagonal. Its determinant is then 1, which shows that dx is right-invariant. A similar argument shows that dx is also left-invariant.

5.1.3 We now describe the version of the real-variable structure, treated in Chapter 1, that is appropriate for our homogeneous group.

We first define the norm function ρ on $\mathbf{R}^n$ by

$$\rho(x) = \max_{1 \le j \le n} \{|x_j|^{1/a_j}\}.$$

Note that $\rho(x) \geq 0$, with $\rho(x) = 0$ if and only if $x = 0$. Also, ρ is homogeneous with respect to the dilations:

$$\rho(\delta \circ x) = \delta \cdot \rho(x), \qquad \delta > 0.$$

We next observe that there is a constant c so that

$$\rho(x \cdot y) \leq c(\rho(x) + \rho(y)), \qquad \text{and} \quad \rho(x^{-1}) \leq c\rho(x). \tag{58}$$

By homogeneity, it suffices to prove (58) when $\rho(x) = 1$ and $\rho(y) \leq 1$. In that case, they are obvious because the sets $\{x \cdot y\}$ and $\{x^{-1}\}$ are then bounded in $\mathbf{R}^n$.

We let $\rho(x, y) = \rho(y^{-1} \cdot x)$, and consider the collection of balls $\{B(x, \delta)\}$ given by

$$B(x, \delta) = \{y : \rho(y, x) < \delta\} = \{y : \rho(x^{-1} \cdot y) < \delta\}.$$

Since ρ is left-invariant, that is, $\rho(a \cdot x, a \cdot y) = \rho(x, y)$ for all $a \in \mathbf{R}^n$, so is the collection of balls, i.e., $B(x, \delta) = x \cdot B(0, \delta)$. Moreover by homogeneity $B(0, \delta) = \delta \circ B(0, 1)$, and hence

$$|B(x, \delta)| = c_1 \delta^a, \tag{59}$$

where $a = a_1 + \cdots + a_n$ is the *homogeneous dimension*, and $c_1 = |B(0, 1)|$.

In view of (58) and (59), the collection of balls $\{B(x, \delta)\}$ and the quasi-distance $\rho(x, y)$ satisfy the properties set forth in §1 of Chapter 1, when we take the underlying measure $d\mu$ to be the Lebesgue measure dx.

5.2 Examples. Before going further, we discuss some examples of homogeneous groups.

5.2.1 Consider $\mathbf{R}^n$ with its usual additive structure:

$$x \cdot y = x + y, \qquad x^{-1} = -x.$$

Then, regardless of the exponents $a_1, \ldots, a_n$ chosen, the mapping $x \mapsto \delta \circ x$ is an automorphism. From this, we see that the same underlying group may have many different homogeneous structures attached to it.

5.2.2 Another example arises in $\mathbf{R}^{2n+1}$, when it is identified with the Heisenberg group $\mathbf{H}^n$, and where the dilations are given by

$$\delta \circ x = [\delta z, \delta^2 t], \qquad \text{when} \quad x = [z, t] \in \mathbf{C}^n \times \mathbf{R}^1.$$

Here again, the underlying group may have a variety of homogeneous structures. For example, when $n = 1$, we can consider the dilations of $\mathbf{H}^1$ given by

$$\delta \circ [x + iy, t] = [\delta x + i\delta^a y, \delta^{a+1} t], \qquad \delta > 0,$$

for any $a > 0$. These mappings are group automorphisms and so give us alternate notions of dilations.

5.2.3 We realize $\mathbf{R}^n$, when $n = m(m-1)/2$, as the set of all real upper-triangular $m \times m$ matrices having ones along the diagonal. That is, we label coordinates in $\mathbf{R}^n$ by (x_{ij}) with $1 \leq i < j \leq m$; we can then identify points in $\mathbf{R}^n$ with $m \times m$ matrices $\{a_{ij}\}$ where: $a_{ij} = x_{ij}$ when $i < j$, $a_{ij} = 1$ when $i = j$, and $a_{ij} = 0$ when $i > j$. The group law on $\mathbf{R}^n$ is then inherited from matrix multiplication. Thus if $z = x \cdot y$, we have

$$z_{ij} = x_{ij} + y_{ij} + \sum_{i<k<j} x_{ik} y_{kj}.$$

From this it is clear that the mappings

$$\delta : (x_{ij}) \mapsto (\delta^{j-i} x_{ij})$$

give automorphisms of the group.

This example of a homogeneous group is but one instance of a large class that is of interest in the theory of semisimple Lie groups. This class arises in the Iwasawa decomposition of such groups, and, as a result, also as "boundaries" of their associated symmetric spaces. For further facts, see §7.10.

5.2.4 To each homogeneous group on $\mathbf{R}^n$, we can associate its Lie algebra $\mathfrak{l}$, consisting of left-invariant vector fields on $\mathbf{R}^n$ (in the same way as we associated $\mathfrak{h}^n$ to $\mathbf{H}^n$ in §2.6 of Chapter 12). As a vector space, $\mathfrak{l}$ is isomorphic to $\mathbf{R}^n$. Let X_j be the left-invariant vector field that agrees with $\partial/\partial x_j$ at the origin; $(X_j)_{1 \leq j \leq n}$ then forms a basis of $\mathfrak{l}$.

Because of (57), we have that

$$(X_j f)(x) = \left.\frac{\partial f(x \cdot y)}{\partial y_j}\right|_{y=0} = \frac{\partial f}{\partial x_j} + \sum_{j<k} q_j^k(x) \frac{\partial f}{\partial x_k},$$

where $q_j^k(x)$ is homogeneous of degree $a_k - a_j$. From this, we see that we can extend our dilations to $\mathfrak{l}$ by defining

$$\delta \circ \left(\sum_{j=1}^n c_j X_j \right) = \sum_{j=1}^n c_j \delta^{a_j} X_j,$$

when $\delta > 0$. These dilations then give an automorphism of the Lie algebra $\mathfrak{l}$, that is

$$\delta \circ [X, Y] = [\delta \circ X, \delta \circ Y].$$

In this sense, the Lie algebra $\mathfrak{l}$ is said to be *homogeneous*.

Next, define the chain of subalgebras

$$\mathfrak{l}^{(1)} \supset \mathfrak{l}^{(2)} \supset \cdots \supset \mathfrak{l}^{(j)} \supset \cdots$$

by $\mathfrak{l}^{(1)} = \mathfrak{l}$, $\mathfrak{l}^{(j)} = [\mathfrak{l}, \mathfrak{l}^{(j-1)}]$. We observe that $\mathfrak{l}^{(k)}$ is contained in the span of the X_j for which $a_j \geq k\min\{a_i\}$; therefore $\mathfrak{l}^{(k)} = \{0\}$ if $k\min\{a_i\} > \max\{a_i\}$. This means that $\mathfrak{l}$ is *nilpotent*. The least d for which $\mathfrak{l}^{(d+1)} = \{0\}$ is called the *step* of $\mathfrak{l}$. In §5.2.1 the group is abelian and the step is 1, in §5.2.2 the step is 2, and in §5.2.3 the step is $m-1$.

To summarize: the Lie algebra of a homogeneous group is itself homogeneous, and as a result is nilpotent. The converse holds in the following sense: not every (abstractly given) nilpotent Lie algebra corresponds to a homogeneous group, but every homogeneous Lie algebra does. Further details may be found in §7.7. Among the most important further examples of such Lie algebras (and their groups) are those which are "free" up to a given step d. In some sense, these play the same role that the space $\mathbf{R}^N$ did in Chapter 11, §2.1. Additional facts concerning these nilpotent Lie groups and their applications may be found in §7.9, §7.14, and §7.15 below.

5.3 Singular integrals. With these preliminaries behind us, we turn to the study of singular integrals on a homogeneous group. We shall be concerned with convolution operators, formally given by

$$(Tf)(x) = \int_{\mathbf{R}^n} K(y^{-1}\cdot x)\, f(y)\, dy, \tag{60}$$

where K is a distribution on $\mathbf{R}^n$. Generalizing the situation of $\mathbf{R}^n$ with its additive structure, considered in §4.5 of Chapter 6, and for $\mathbf{H}^n$, considered in §5.2 of the previous chapter, we make the following assumptions on K.

We suppose that K is a distribution on $\mathbf{R}^n$, that agrees with a function $K(x)$ for $x \neq 0$, with $K(x)$ satisfying the regularity conditions:

$$\begin{cases} |K(x)| & \leq A\rho(x)^{-a}, \\ \left|\dfrac{\partial}{\partial x_j}K(x)\right| & \leq A\rho(x)^{-a-a_j}, \quad j = 1,\dots,n. \end{cases} \tag{61}$$

Here a denotes the homogeneous dimension of $\mathbf{R}^n$.

We also assume that K enjoys the cancellation condition

$$|K(\Phi^R)| \leq A, \qquad \text{for all } R > 0, \tag{62}$$

where $\Phi^R(x) = \Phi(R^{-a_1}x_1,\dots,R^{-a_n}x_n)$, whenever Φ is a normalized bump function, that is, for Φ that are supported in the unit ball and satisfy

$$\left|\left(\frac{\partial}{\partial x}\right)^{\alpha}\Phi(x)\right| \leq 1, \qquad \text{for } |\alpha| \leq N,$$

for some fixed N.

THEOREM 4. *Assume that the distribution K satisfies* (61) *and* (62). *Then the convolution operator*

$$Tf = f * K, \tag{63}$$

defined initially for test functions f, has an extension to a bounded operator from $L^p(\mathbf{R}^n)$ to itself, $1 < p < \infty$.

5.3.1 We shall see that, in terms of the real-variable structure of the homogeneous group on $\mathbf{R}^n$ described in §5.1.3, all the results of Chapter 1 are in fact applicable to the present situation. What that theory requires is knowing the above result for a particular value of p, and we now address ourselves to the proof of Theorem 4 in the case $p = 2$ (the results of Chapter 1, together with duality, will then extend Theorem 4 to all p, $1 < p < \infty$). The method we shall use is that of almost orthogonal decompositions, as given in §2 of Chapter 7.

To carry this out, we choose a smooth cut-off function η so that $\eta(x) = 1$ for $\rho(x) \leq 1$, and $\eta(x) = 0$ for $\rho(x) \geq 2$. Then when $x \neq 0$, we have

$$1 = \sum_{j=-\infty}^{\infty} \psi(2^{-j}\circ x) = \sum_{j=-\infty}^{\infty} \eta(2^{-j}\circ x) - \eta(2^{-j+1}\circ x), \tag{64}$$

where we have defined $\psi(x) = \eta(x) - \eta(2\circ x)$. Clearly, $\psi(x)$ is supported in $1/2 \leq \rho(x) \leq 2$. Next, we write $K_j(x) = K(x)\psi(2^{-j}\circ x)$, so that formally $K = \sum_{-\infty}^{\infty} K_j$. We set $T_jf = f * K_j$. Our first main task is to prove that

$$\left\|\sum_{j=\ell_1}^{\ell_2} T_j\right\|_{L^2\to L^2} \leq A, \tag{65}$$

where the bound A does not depend on ℓ_1 or ℓ_2.

The regularity hypothesis (61) imposed on the kernel K will be manifested in the following inequalities for the K_j, which hold uniformly in j:

$$\int_{\mathbf{R}^n} \rho(x)^b\, |K_j(x)|\, dx \leq A2^{jb}, \qquad \text{for any real } b, \tag{66}$$

$$\int_{\mathbf{R}^n} |K_j(x\cdot u) - K_j(x)|\, dx \leq A2^{-ja_1}\, \rho(u)^{a_1}, \tag{67}$$

and

$$\int_{\mathbf{R}^n} |K_j(u\cdot x) - K_j(x)|\, dx \leq A2^{-ja_1}\, \rho(u)^{a_1}. \tag{68}$$

Here a_1 is the smallest of the exponents $a_1,\dots,a_n$.

The proof of (66)–(68) is facilitated by the exploitation of dilations. Observe that if $K(x)$ satisfies (61), then so does $K^{(j)}(x) = 2^{ja}K(2^j \circ x)$, uniformly in j. From this we see that in order to prove (66), it suffices to verify it in the special case $j = 0$, allowing all K for which (61) holds. Indeed, from

$$\int_{\mathbf{R}^n} \rho(x)^b \, |\psi(x) \, K^{(j)}(x)| \, dx \leq A,$$

we get (upon making the change of variable $x \mapsto 2^{-j} \circ x$) that (66) is valid, once we recall that

$$\rho(2^{-j} \circ x) = 2^{-j}\rho(x), \qquad d(2^{-j} \circ x) = 2^{-ja} \, dx,$$

and $\psi(2^{-j} \circ x)K(x) = K_j(x)$.

In the same way, the proof of (67) is reduced to the case $j = 0$, for all K that satisfy (61). Now $K_0(x) = \psi(x) \cdot K(x)$ has bounded first derivatives and is compactly supported. Therefore

$$\int_{\mathbf{R}^n} |K_0(x \cdot u) - K_0(x)| \, dx \leq A|u|,$$

since the mapping $(x, u) \mapsto x \cdot u$ is smooth. However when $\rho(u) \leq 1$, we have $|u| \leq c\rho(u)^{1/a_1}$, because

$$\rho(u) = \max_j \{|u_j|^{1/a_j}\}$$

and $a_1 = \min_j \{a_j\}$. On the other hand, when $\rho(u) \geq 1$, we may use the obvious fact that

$$\int_{\mathbf{R}^n} |K_0(x \cdot u) - K_0(x)| \, dx \leq A.$$

In either case, (67) is established for $j = 0$, and hence for all j. The proof of (68) is similar.

5.3.2 We continue with the proof of the L^2 boundedness of T by temporarily assuming a cancellation condition on K that is stronger than (62), namely that

$$\int_{\mathbf{R}^n} K_j(x) \, dx = 0, \qquad \text{for all } j. \tag{69}$$

With it, we shall prove that

$$\|T_i^* T_j\| + \|T_i T_j^*\| \leq A 2^{-a_1|i-j|}; \tag{70}$$

here $\| \cdot \|$ denotes the L^2 norm of an operator.

Now $T_i^* T_j f = f * L_{ij}$, where $L_{ij} = K_j * K_i^*(x)$, with $K^*(x) = \bar{K}(x^{-1})$; so

$$L_{ij}(x) = \int K_j(y) \, K_i^*(y^{-1}x) \, dy = \int K_j(y) \, \bar{K}_i(x^{-1} \cdot y) \, dy,$$

and $L_{ij}(x^{-1}) = \int K_j(y) \, \bar{K}_i(x \cdot y) \, dy$. By (69), we have

$$L_{ij}(x^{-1}) = \int K_j(y)[\bar{K}_i(x \cdot y) - \bar{K}_i(x)] \, dy;$$

using (66) and (67) then gives

$$\begin{aligned} \int |L_{ij}(x^{-1})| \, dx &\leq A \int |K_j(y)| \, 2^{-ia_1} \, \rho(y)^{a_1} \, dy \\ &\leq A \, 2^{-ia_i} \, 2^{ja_1} = A \, 2^{-a_1|i-j|}, \end{aligned}$$

if $j \leq i$.

Because $\|T_i^* T_j\| \leq \|L_{ij}(x^{-1})\|_{L^1} = \|L_{ij}(x)\|_{L^1}$, we see that

$$\|T_i^* T_j\| \leq A \, 2^{-a_1|i-j|},$$

when $j \leq i$. The case $j > i$ is treated similarly by rewriting L_{ij} as

$$\begin{aligned} L_{ij}(x) &= \int K_j(y) \, \bar{K}_i(x^{-1} \cdot y) \, dy = \int K_j(x \cdot y) \, \bar{K}_i(y) \, dy \\ &= \int [K_j(x \cdot y) - K_j(x)] \, \bar{K}_i(y) \, dy. \end{aligned}$$

The estimate $\|T_i T_j^*\| \leq A \, 2^{-a_1|i-j|}$ is the same as that for $\|T_i^* T_j\|$, except that $K_j(x)$ is replaced by $\bar{K}_j(x^{-1})$, and $\bar{K}_i(x^{-1})$ is replaced by $K_i(x)$.

Therefore, (70) has been established, and in view of the theorem in §2.1 of Chapter 7, we also have (65).

5.3.3 We now lift the extra cancellation requirement (69). We let $c_j = \int K_j(x) \, dx$, and set

$$\tilde{\psi}_j(x) = (\textstyle\int \psi \, dx)^{-1} \cdot 2^{-ja} \, \psi(2^{-j} \circ x),$$

so $\int \tilde{\psi}_j(x) \, dx = 1$ for all j. We also write

$$\tilde{K}_j = K_j - c_j \tilde{\psi}_j;$$

hence $\int \tilde{K}_j \, dx = 0$. Now

$$\sum_{\ell_1}^{\ell_2} K_j = \sum_{\ell_1}^{\ell_2} (\tilde{K}_j + c_j \tilde{\psi}_j).$$

But summation by parts gives

$$\sum_{\ell_1}^{\ell_2} c_j\psi_j = \sum_{\ell_1}^{\ell_2} s_j(\tilde{\psi}_j - \tilde{\psi}_{j+1}) + s_{\ell_2}\tilde{\psi}_{\ell_2+1},$$

where $s_j = \sum_{\ell_1}^{j} c_k$. Altogether then

$$\sum_{\ell_1}^{\ell_2} K_j = \sum_{\ell_1}^{\ell_2} [\tilde{K}_j + s_j(\tilde{\psi}_j - \tilde{\psi}_{j+1})] + s_{\ell_2}\tilde{\psi}_{\ell_2+1}. \tag{71}$$

We need to observe three facts about the terms in (71). First,

$$|s_j| \leq A \qquad \text{for all } j, \tag{72}$$

because

$$s_k = \sum_{\ell_1}^{k} \int [\eta(2^{-j}\circ x) - \eta(2^{-j+1}\circ x)]\, K(x)\, dx = K(\eta^{2^{\ell_1}}) - K(\eta^{2^{k+1}}),$$

which is bounded by assumption (62). Second, the terms

$$\tilde{K}_j + s_j(\tilde{\psi}_j - \tilde{\psi}_{j+1})$$

enjoy the cancellation condition (69), since

$$\int \tilde{K}_j\, dx = 0 \quad \text{and} \quad \int (\tilde{\psi}_j - \tilde{\psi}_{j+1})\, dx = 0.$$

Third, these terms also satisfy the inequalities (66)–(68) because the K_j do, and it is easy to verify that the $\tilde{\psi}_j$ also have this property.

Thus we can use $\tilde{K}_j + s_j(\tilde{\psi}_j - \tilde{\psi}_{j+1})$ instead of K_j in §5.3.2 above. Together with the observation that $\|\tilde{\psi}_{\ell_2+1}\|_{L^1} \leq A$, this lets us obtain the L^2 conclusion (65) in general.

5.3.4 We can now conclude the proof of the theorem, using by now familiar arguments (see §5.2.3 and §5.2.4 of the previous chapter). First, we find appropriate sequences $\ell_1^k \to -\infty$, $\ell_2^k \to +\infty$, so that $\sum_{\ell_1^k}^{\ell_2^k} K_j$ tends to a distribution K' with the following properties:

(a) K' agrees with K away from the origin.

(b) K' also has the cancellation property (62).

(c) If $T'(f) = f * K'$, then T' is bounded on L^2, because of (65).

Now (a) implies that

$$K - K' = \sum_{|\alpha|\leq M} c_\alpha(\partial_x)^\alpha \delta_0,$$

where δ_0 is the Dirac delta function, and so

$$K(\Phi^R) - K'(\Phi^R) = \sum_{|\alpha|\leq m} c_\alpha R^{-a\cdot\alpha}\Phi^{(\alpha)}(0),$$

with $a\cdot\alpha = a_1\alpha_1 + \cdots + a_n\alpha_n$. This is inconsistent with (62) and (b), unless $c_\alpha = 0$ for $\alpha \neq 0$. Thus $K = K' + c_0\delta_0$, and therefore T is bounded on L^2.

To obtain the L^p theory, we verify that condition (18′) of Chapter 1 holds. Indeed, $V(x,y) = |B(y,\delta)|$ with $\delta = \rho(x,y)$, thus

$$V(x,y) = c\rho(x,y)^a,$$

as we saw in §5.1. Next, since the kernel $K(x,y)$ equals $K(y^{-1}\cdot x)$, the condition

$$|K(x,y) - K(x,\bar{y})| \leq A\left(\frac{\rho(y,\bar{y})}{\rho(x,\bar{y})}\right)^{a_1} V(x,\bar{y})^{-1}, \quad \text{for} \quad \rho(x,\bar{y}) \geq c_1\rho(y,\bar{y})$$

is seen to be identical with

$$|K(u^{-1}\cdot x) - K(x)| \leq A\left(\frac{\rho(u)}{\rho(x)}\right)^{a_1}\cdot\rho(x)^{-a}, \qquad \text{for} \quad \rho(x) \geq c_1\rho(u). \tag{73}$$

We prove (73) in the same way as (67) and (68): A homogeneity argument reduces it to the case $\rho(x) = 1$. Then if c_1 is sufficiently large, (73) follows from the fact that K is C^1 away from the origin, and that $|u| \leq c\rho(u)^{a_1}$ for small u. In the same way, one proves the inequality for the kernel adjoint to $K(x,y)$, which arises in (29) of Chapter 1.

6. *Appendix: The $\bar{\partial}$-Neumann problem*

We shall discuss here the $\bar{\partial}$-Neumann problem (formulated in §1.3) in the setting of the model domain $\mathcal{U}^n$ and describe a derivation of the exact formula for the solution.

We begin with a few historical remarks. The $\bar{\partial}$-Neumann problem was formulated by Spencer. In the case of strictly pseudoconvex domains, it was first solved by Kohn, who obtained C^∞ results, using L^2 methods; see, for instance, the exposition in Folland and Kohn [1972]. The analysis relying on the Heisenberg group and leading to formulas for the solutions, together with sharp estimates, came later. It is the second approach that we sketch here.

6.1 Recall that

$$\mathcal{U}^n = \{z = (z', z_{n+1}) \in \mathbf{C}^{n+1} : z' \in \mathbf{C}^n, z_{n+1} \in \mathbf{C}, \operatorname{Im}(z_{n+1}) > |z'|^2\}.$$

Besides the ambient coordinates (z', z_{n+1}), it is useful to deal with the Heisenberg coordinates $[\zeta, t, r]$ given by

$$\zeta = z', \qquad t = \operatorname{Re} z_{n+1}, \qquad r = \operatorname{Im}(z_{n+1}) - |z'|^2,$$

as in §1.4.4 of Chapter 12.

When considering forms on $\mathcal{U}^n$ it will be natural to choose a basis

$$\omega_1, \dots, \omega_{n+1}$$

of $(1,0)$ forms so that

$$\omega_j = dz_j, \qquad j = 1, \dots, n \qquad \text{(the “tangential” forms)},$$

and

$$\omega_{n+1} = 2^{1/2}\partial r = -2^{1/2} \sum_{j=1}^{n} \bar{z}_j \partial z_j - i2^{1/2} dz_{n+1} \qquad \text{(the “normal” form)}.$$

The vector fields dual to these forms are then

$$Z_j = \frac{\partial}{\partial z_j} + 2i\bar{z}_j \frac{\partial}{\partial z_{n+1}}, \qquad j = 1, \dots, n,$$

and $Z_{n+1} = i2^{1/2}\partial/\partial z_{n+1}$. Note that in the Heisenberg coordinates this gives

$$Z_j = \frac{\partial}{\partial \zeta_j} + i\bar{\zeta}_j \frac{\partial}{\partial t}, \qquad j = 1, \dots, n,$$

and $Z_{n+1} = 2^{-1/2}\left(\frac{\partial}{\partial r} + i\frac{\partial}{\partial t}\right)$; also $\omega_j = d\zeta_j$, $j = 1, \dots, n$, $\omega_{n+1} = 2^{1/2}\partial r$. See also §2.6.3 of Chapter 12.

6.2 The adjoint $\bar{\partial}^*$ of $\bar{\partial}$ (and also the operator $\square$) are defined in the presence of a Hermitian metric. In the present context the appropriate metric to use is the one for which the basic forms $\omega_1, \dots, \omega_{n+1}$ are orthonormal at each point.

To describe $\square$ in this metric, split any $(0,1)$ form u as $u = u^\tau + u^\nu$, where $u^\tau = \sum_{j=1}^{n} u_j \overline{\omega}_j$ is the “tangential” part and $u^\nu = u_{n+1}\overline{\omega}_{n+1}$ is the “normal” component. An elementary calculation then shows that

$$\square u = \square u^\tau + \square u^\nu,$$

with

$$\square u^\tau = \sum_{j=1}^{n} \square^\tau(u_j)\overline{\omega}_j, \qquad \square u^\nu = \square^\nu(u_{n+1})\overline{\omega}_{n+1}$$

where, in Heisenberg coordinates,

$$\begin{aligned} \square^\tau &= \mathfrak{L}_{n-2} - \frac{1}{2}\left(\frac{\partial^2}{\partial t^2} + \frac{\partial^2}{\partial r^2}\right), \\ \square^\nu &= \mathfrak{L}_{n} - \frac{1}{2}\left(\frac{\partial^2}{\partial t^2} + \frac{\partial^2}{\partial r^2}\right). \end{aligned} \tag{$*$}$$

The operators $\mathfrak{L}_{n-2}$ and $\mathfrak{L}_n$ are given by (16).

If one takes into account the boundary conditions (10) for $\square$ then the $\bar{\partial}$-Neumann problem can be seen to split into a pair of problems (for scalar valued functions)

$$\square^\tau(U) = f \quad \text{in } \mathcal{U}^n, \text{ with } \quad \bar{Z}_{n+1}U|_{b\mathcal{U}^n} = 0, \tag{$**$}$$

and

$$\square^\nu(U) = f \quad \text{in } \mathcal{U}^n, \text{ with } \quad U|_{b\mathcal{U}^n} = 0.$$

The second problem, involving the normal component, is essentially the Dirichlet problem for the Laplacian; it can be treated by the more standard methods used in elliptic boundary value problems. We turn therefore to the first problem.

6.3 In order to deal with the problem $(**)$, we consider first a fundamental solution F for the operator $\square^\tau$. On the space

$$\mathbf{H}^n \times \mathbf{R} = \{(x, r) : x \in \mathbf{R}^n, r \in \mathbf{R}\}$$

we let $F(x, r)$ be the function determined by

$$\left[\mathfrak{L}_{n-2} - \frac{1}{2}\left(\frac{\partial^2}{\partial t^2} + \frac{\partial^2}{\partial r^2}\right)\right] F(x, r) = \delta_0,$$

with F vanishing at infinity; here δ_0 denotes the Dirac delta function at the origin. Such a fundamental solution exists and is unique (see, e.g., Nagel, Ricci, and Stein [1990]). The uniqueness of F and the fact that $\square^\tau$ is even in r show that $F(x, r)$ is even in r. We set

$$G(x, y, r, \rho) = F(y^{-1} \cdot x, r - \rho) - F(y^{-1} \cdot x, r + \rho).$$

Then the operator G, defined by

$$(Gf)(x, r) = \int_{\mathbf{H}^n \times \mathbf{R}} G(x, y, r, \rho)\, f(y, \rho)\, dy\, d\rho$$

is “Green's operator” for $\square^\tau$; that is

$$\square^\tau G(f) = f \quad \text{in } \mathcal{U}^n, \qquad \text{and} \qquad G(f)|_{b\mathcal{U}^n} = 0.$$

Next let

$$P(y^{-1}x, r) = \frac{\partial G}{\partial \rho}(x, y, r, \rho)|_{\rho=0} = -2\frac{\partial F}{\partial r}(y^{-1}x, r)$$

be the Poisson kernel, in the sense that if f_b is defined on $b\mathcal{U}^n = \mathbf{H}^n$, then

$$(Pf_b)(x, r) = \int_{\mathbf{H}^n} P(y^{-1} \cdot x, r)\, f_b(y)\, dy$$

satisfies

$$\square^\tau P(f_b) = 0, \qquad P(f_b)|_{b\mathcal{U}^n} = f_b.$$

6.4 We now isolate the basic boundary operator $\Box_+$ occurring in the $\bar\partial$-Neumann problem. We can write the solution of (**) as

$$U = G(f) + P(U_b)$$

where $U_b = U|_{b\mathcal{U}^n}$. In this way of looking at the problem, f is given and U_b is to be determined.

If we apply the boundary condition in (**) to the above, we find that

$$\bar Z_{n+1} P(U_b)|_{b\mathcal{U}^n} = -\bar Z_{n+1} G(f)|_{b\mathcal{U}^n}.$$

The basic operator $\Box_+$ is then defined by

$$\Box_+(U_b) = \bar Z_{n+1} P(U_b)|_{b\mathcal{U}^n}.$$

To calculate it, we note that $\bar Z_{n+1} = 2^{-1/2}\left(\frac{\partial}{\partial r} - i\frac{\partial}{\partial t}\right)$, and $T = \frac{\partial}{\partial t}$ is tangential. Thus

$$\Box_+ = 2^{-1/2}(N - iT)$$

where N is the "Dirichlet to Neumann" operator, given by

$$N(U_b) = \frac{\partial}{\partial r} P(U_b)|_{b\mathcal{U}^n}.$$

The operator N is a pseudo-differential operator of order 1, whose symbol can be computed to any degree of accuracy. However in the present case, its crucial property is contained in a simple and exact formula, namely

$$\frac{1}{2}N^2 = \mathfrak{L}_{n-2} - \frac{1}{2}T^2.$$

This identity follows because, in this situation, $P_{r_1}P_{r_2} = P_{r_1+r_2}$, and hence $N^2 = \frac{\partial^2}{\partial r^2} P_r|_{r=0}$.

Now let $\Box_- = 2^{-1/2}(N + iT)$. Using the fact that T commutes with N we have

$$\Box_+\Box_- = \frac{1}{2}(N^2 + T^2) = \mathfrak{L}_{n-2}.$$

For simplicity we now restrict our attention to the case $n \geq 2$. Then according to §2.4 we have an inverse S for the operator $\mathfrak{L}_{n-2}$, and thus we get

$$(\Box_+)^{-1} = \Box_- S. \qquad (***)$$

6.5 The above then determines U_b by the formula

$$U_b = -\Box_- S(\bar Z_{n+1} G(f)|_{b\mathcal{U}^n}),$$

and hence the solution of the $\bar\partial$-Neumann problem (**) is given by

$$U = G(f) - P\Box_- S R \bar Z_{n+1} G(f),$$

where R denotes the operator of restriction to the boundary $b\mathcal{U}^n$.

6.6 We make some additional remarks about the final formula stated in §6.5.

(i) When $\mathcal{U}^n$ is replaced by a more general strictly pseudoconvex domain, an "approximate" version of the exact formula is still valid, and can be derived following the same approach but using the calculus of pseudo-differential operators and the approximation by the Heisenberg group described in §7.13 below. The details of this are set out in Greiner and Stein [1977] in which the case $n = 1$ is also treated; this latter case requires the complex conjugate of the solving operator $\tilde S$ in place of S ($\tilde S$ is defined in §4.2).

(ii) The relatively explicit nature of the operators involved in the final formula, together with estimates of the kind arising in §3.2 for S (or $\tilde S$) allow one to obtain sharp estimates for solutions of the $\bar\partial$-Neumann problem. Broadly speaking, these assert that we can recover the full gain of 2 (as in the elliptic case) when we limit consideration to differentiation in the complex tangential directions. The theory along these lines has been extended to domains in $\mathbf{C}^2$ that are only *weakly* pseudoconvex (and of finite type); see D. Chang, Nagel, and Stein [1988]. In the case of weakly pseudoconvex domains of finite type in $\mathbf{C}^n$, $n > 2$, the C^∞ theory may be said to be in satisfactory shape (see Catlin [1983], [1986]); however, there is at present no analogue of the sharp results described above, except in some special instances.

(iii) If one takes $u = \bar\partial^* U$ one gets the solution u of $\bar\partial u = f$ that is orthogonal to holomorphic functions. In the case of strictly pseudoconvex domains, one can also solve the $\bar\partial u = f$ problem, without recourse to the $\bar\partial$-Neumann problem, by using the Cauchy-Fantappié-Leray formalism. This goes back to the work of Henkin, Ramirez, Grauert and Lieb, and Kerzman (among others); see the account in Range [1986].

(iv) An integral formula, in terms of an "explicit" kernel, has been given for the solution of the $\bar\partial$-Neumann problem for $\mathcal{U}^n$ by Phong [1979]; see also the exposition in M. Beals, C. Fefferman, and R. Grossman [1983]. Other formulas for the model domain are in Stanton [1981], Harvey and Polking [1985]. In order to recover the sharp estimates cited in (ii) using these formulas, one needs estimates for singular Radon transforms (as described in Chapter 11, §4.7); see Phong and Stein [1986b].

6.7 It is worthwhile to point out that if strict pseudoconvexity is not required, one cannot expect regularity for the $\bar\partial$-Neumann problem as above. This is indicated be consideration of the extreme case of the domain

$$\mathcal{D} = \{(z_1, z_2) \in \mathbf{C}^2 : \operatorname{Im} z_2 > 0\}.$$

The analogue of (**) in §6.2 can be taken to be the problem

$$\Box U = f, \qquad \text{with} \quad \frac{\partial}{\partial \bar z_2} U|_{\operatorname{Im} z_2 = 0} = 0.$$

Notice that we have a "null" solution $U(z_1, z_2) = u_1(z_1)u_2(z_2)$ whenever u_2 is holomorphic in z_2 for $\operatorname{Im} z_2 > 0$ and u_1 is harmonic in z_1; this solution need not have any pre-assigned degree of smoothness in $b\mathcal{D}$.

7. Further results

A. The operator $\mathfrak{L}_\alpha$ and the Heisenberg group

7.1 We describe several alternative approaches to the determination of the fundamental solution F_α of $\mathfrak{L}_\alpha$.

(a) The considerations of symmetry (with respect to dilations and rotations on $\mathbf{H}^n$) discussed in §2.3.3 prompt us to look for $F_\alpha(z,t)$ of the form

$$(|z|^4+t^2)^{-n/2}\psi\Big(\frac{t}{(|z|^4+t^2)^{1/2}}\Big),$$

where ψ is an undetermined function of one variable. The equation $\mathfrak{L}_\alpha F_\alpha = 0$ (which holds when $(z,t) \neq (0,0)$) then leads to an ordinary second-order differential equation in ψ. Making the substitution $\eta(\theta) = \psi(\cos\theta)$ yields (after some reduction) the equation

$$\Big(\sin\theta\frac{d}{d\theta}+n\cos\theta\Big)\Big(\frac{d}{d\theta}+i\alpha\Big)\eta(\theta)=0.$$

This last equation has $\eta(\theta) = ce^{-i\alpha\theta}$ as its bounded solution, giving

$$F_\alpha(z,t) = c(|z|^2-it)^{-(n+\alpha)/2}(|z|^2+it)^{-(n-\alpha)/2}.$$

(b) One can also proceed by first taking the Fourier transform in the t variable in the equation $\mathfrak{L}_\alpha F_\alpha = \delta_0$, assuming now that F_α has the form

$$|z|^{-2n}\tilde{\psi}\Big(\frac{t}{|z|^2}\Big).$$

Proceeding in this way, one is ultimately led to the Fourier transform of F_α in *all* variables.

Writing $\xi = [\zeta,\tau]$ with ξ, ζ, τ being dual to $x = [z,t]$, z, t respectively, one has the formula

$$\widehat{F_\alpha}(\zeta,\tau) = \frac{1}{2\pi\tau}\int_0^1 (1-u)^{\frac{n-\alpha}{2}-1}(1+u)^{\frac{n+\alpha}{2}-1}e^{-\pi u|\zeta|^2/2\tau}\,du$$

when $\tau > 0$, if $-n < \alpha < n$. The same formula holds with τ replaced by $-\tau$ and α by $-\alpha$ when $\tau < 0$.

(c) A third derivation uses the "heat" operator $e^{-s\mathfrak{L}_\alpha}$ on the Heisenberg group (see §7.2 below), and the formal identity

$$\mathfrak{L}_\alpha^{-1} = \int_0^\infty e^{-s\mathfrak{L}_\alpha}\,ds.$$

For (a), see Folland and Stein [1974]; (b) is in Greiner and Stein [1977]; (c) is in R. Beals and Greiner [1988].

7.2 We consider the initial value problem for the "heat" equation on the Heisenberg group

$$\mathfrak{L}_\alpha(u(x,s)) = -\frac{\partial u}{\partial s}(x,s), \qquad u(x,0) = f(x),$$

where f is a given function of $x \in \mathbf{H}^n$ and $s > 0$. When $-n < \alpha < n$, the operator $\mathfrak{L}_\alpha$ (on its natural domain) is a nonnegative self-adjoint operator on $L^2(\mathbf{H}^n)$, so the (bounded) operator $e^{-s\mathfrak{L}_\alpha}$ is well-defined when $s > 0$. The solution of the heat equation, for $f \in L^2$, is given by

$$u(x,s) = (e^{-s\mathfrak{L}_\alpha}f)(x).$$

It can be shown that $u(x,s) = (f * h_s^\alpha)(x)$, where the heat kernel satisfies

$$h_s^\alpha(x) = h_1^\alpha(z/s^{1/2}, t/s)s^{-n-1}, \qquad h_1^\alpha \in \mathcal{S}.$$

One also has the Fourier transform formula

$$\widehat{h_s^\alpha}(\zeta,\tau) = e^{2\pi\alpha\tau s}(\cosh 2\pi\tau s)^{-n}\,e^{-\pi|\zeta|^2\tanh(2\pi\tau s)/2\tau},$$

with $[\zeta,\tau]$ dual to $[z,t]$.

Hulanicki [1976], Gaveau [1977].

7.3 Let L be a differential operator on $\mathbf{H}^n$. We assume:

(i) L commutes with the left translations $f(x) \mapsto f(a\cdot x)$, for all $a \in \mathbf{H}^n$,

(ii) L commutes with the rotations $f(z,t) \mapsto f(u(z),t)$, for all $u \in \mathbf{U}(n)$, and

(iii) L has degree 2 in the sense that

$$[L(f^{\delta^{-1}})]^\delta = \delta^2 Lf, \qquad \text{for all } \delta > 0,$$

where $f^\delta(z,t) = f(\delta^{-1}z, \delta^{-2}t)$. Then one can conclude that

$$L = c\mathfrak{L}_\alpha \quad \text{for appropriate constants c and } \alpha.$$

For the proof of this assertion the following facts are useful.

(a) Any left-invariant differential operator on $\mathbf{H}^n$ is a (noncommutative) polynomial in the X_j, Y_j, $1 \le j \le n$, and T.

(b) Suppose we write $V_j = X_j$, $1 \le j \le n$, $V_j = Y_{j-n}$, $n+1 \le j \le 2n$, and $V_{2n+1} = T$. For any monomial $u^m = u_1^{m_1}\cdots u_{2n+1}^{m_{2n+1}}$ on $\mathbf{R}^{2n+1}$, we let V^m denote the corresponding (symmetrized) polynomial in $V_1,\dots,V_{2n+1}$, which is given as the coefficient of u^m in the expression

$$\frac{m!}{|m|!}\Big(\sum_{j=1}^{2n+1}u_jV_j\Big)^{|m|};$$

here $|m| = \sum_{j=1}^{2n+1} m_j$ and $m! = \prod_{j=1}^{2n+1} m_j!$.

Then $\{V^m\}$ forms a basis of the vector space of left-invariant differential operators on $\mathbf{H}^n$.

(c) If P is a homogeneous polynomial of degree 2 on $\mathbf{R}^{2n}$ that is invariant under the unitary group on $\mathbf{C}^n$, then $P(z) = c|z|^2$.

Assertions (a) and (b) are special cases of analogous statements that are valid for any Lie group (not just $\mathbf{H}^n$) and which are closely connected with the Birkhoff-Witt theorem for the universal enveloping algebra; this algebra is isomorphic to the algebra of left-invariant differential operators on the group in question. For these facts, see Jacobson [1967], Varadarajan [1974], Stein [1970a].

7.4 The motivation leading to the identity (53) for the relative fundamental solution of $\mathfrak{L}_\alpha$, $\alpha = n$, can be easily explained in the context of finite-dimensional operators.

Suppose L_α is, for each α, an $N \times N$ symmetric matrix, depending smoothly on α, and that the $\{L_\alpha\}$ are mutually commutative. Assume that there is a corresponding family $\{M_\alpha\}$ (also with smooth dependence on α) so that

$$L_\alpha M_\alpha = c_\alpha I.$$

We also suppose that c_α has a *simple* zero at $\alpha = \alpha_0$. Then

$$L_{\alpha_0}\tilde{M} = I - P$$

where P is the orthogonal projection on the null-space of L_{α_0} and $\tilde{M} = (c'_{\alpha_0})^{-1}M'_{\alpha_0}$; here $'$ denotes differentiation with respect to α. One also notes that $P = (c'_{\alpha_0})^{-1}L'_{\alpha_0}M_{\alpha_0}$.

This can be proved by using a simultaneous diagonalization of the L_α and M_α. If ℓ_α and m_α are corresponding diagonal entries, we have $\ell_\alpha m_\alpha = c_\alpha$. Differentiation gives $\ell'_{\alpha_0}m_{\alpha_0} = c'_{\alpha_0}$ if ℓ_α vanishes at $\alpha = \alpha_0$, or $\ell_{\alpha_0}m'_{\alpha_0} = c'_{\alpha_0}$ if m_α vanishes at α_0. Notice that ℓ_α and m_α cannot both vanish at α_0.

7.5 The nonisotropic Sobolev spaces NL^p_k considered in §3.2 can be compared to the standard (isotropic) Sobolev spaces treated in Chapter 6 §5, when $\mathbf{H}^n$ is identified with $\mathbf{R}^{2n+1}$.

(a) The inclusion relation $NL^p_k \subset L^p_{k/2,\mathrm{loc}}$ holds in the sense that, whenever $\phi \in C_0^\infty(\mathbf{H}^n)$, the mapping $f \mapsto \phi \cdot f$ is continuous from $NL^p_k(\mathbf{H}^n)$ to $L^p_{k/2}(\mathbf{R}^{2n+1})$, when $1 < p < \infty$ and k is a nonnegative integer.

(b) The inclusion is sharp in that $NL^p_k \not\subset L^p_{k/2+\varepsilon}$ for any $\varepsilon > 0$.

(c) There are also nonisotropic variants of the Lipschitz spaces Λ_γ considered in Chapter 6; these are called the Γ_γ spaces. Results analogous to those in §3.3 hold for these spaces. If $0 < \gamma < 1$, then Γ_γ consists of bounded functions f for which

$$|f(x \cdot y) - f(x)| \le A\rho(y)^\gamma;$$

when $\gamma = 1$, the condition is replaced by

$$|f(x \cdot y) + f(x \cdot y^{-1}) - 2f(x)| \le A\rho(y),$$

and when $\gamma > 1$, the spaces are defined by recursion. One has $\Gamma_\gamma \subset \Lambda_{\gamma/2,\mathrm{loc}}$; again this inclusion is sharp.

For further details concerning (a) and (c), the reader is referred to Folland and Stein [1974]. Various extensions may be found in Folland [1975], Rothschild and Stein [1976].

7.6 The reasoning used in §2.3.1, which led us to presume that $\mathfrak{L}_\alpha$ was invertible when $\pm\alpha \neq n, n+2, \ldots$, can be made precise and systematized as follows.

Suppose L is a left-invariant differential operator on $\mathbf{H}^n$ that is homogeneous of degree k (i.e., $[L(f^{\delta^{-1}})]^\delta = \delta^k Lf$ for all $\delta > 0$). Then the hypoellipticity of L can be determined by using the representations of $\mathbf{H}^n$ (for which, see the appendix to the previous chapter). The necessary and sufficient conditions are:

(a) $r_{\alpha,\beta}(L) \neq 0$, where $r_{\alpha,\beta}$ is the one-dimensional representation determined by

$$r_{\alpha,\beta}(z,t) = e^{i(\alpha\cdot x + \beta\cdot y)}, \qquad z = x + iy,$$

and the condition holds for $(\alpha,\beta) \neq (0,0)$.

(b) The operator $R^\lambda(L)$ is injective (when restricted to $\mathcal{S}$) for each $\lambda \neq 0$. Here R^λ is the representation determined by (91) in Chapter 12.

Note that the condition (a) is equivalent to the statement that L, when restricted to functions $f(z,t)$ that do not depend on t, is an elliptic operator on $\mathbf{C}^n \approx \mathbf{R}^{2n}$. The above theorem is in Rockland [1978].

When the conditions are satisfied, it can be shown that estimates akin to those in §3.3 hold for L. Moreover, L then has a "homogeneous" fundamental solution which in fact is analytic away from the origin. Thus L is also "analytically" hypoelliptic. Furthermore, if we assume only hypothesis (a), then there is a relative fundamental solution, akin to the situation described in §4.2 for $\mathfrak{L}_\alpha$ when $\alpha = n$. Again the kernels of these relative fundamental solutions (and the corresponding projection operators) are homogeneous of appropriate degree and are analytic away from the origin. For these conclusions see Geller [1980b], [1990], Métevier [1980], Rothschild and Tartakoff [1982].

Two additional remarks are in order. The use of representation theory to deal with questions of hypoellipticity had already occurred in Rothschild and Stein [1976]. An analogue of Rockland's criterion extends to all "graded" nilpotent groups; see Helffer and Nourrigat [1985].

B. Homogeneous groups

7.7 Suppose $\mathfrak{l}$ is a nilpotent Lie algebra of dimension n and step d. Choose a basis $X_1, \ldots, X_n$ of $\mathfrak{l}$, and, if $u \in \mathbf{R}^n$, abbreviate $u_1X_1 + \cdots + u_nX_n$ by $u \cdot X$. Then, according to the Campbell-Hausdorff formula (see §7.14 of the previous chapter and the references given there), we have that

$$e^{u\cdot X}\, e^{v\cdot X} = e^{P(u,v)\cdot X},$$

where $P : \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}^n$ is a polynomial map of degree at most d. This mapping gives a multiplication law of a Lie group whose underlying space is $\mathbf{R}^n$; this group is isomorphic to the simply connected nilpotent Lie group whose Lie algebra is $\mathfrak{l}$. The coordinates in $\mathbf{R}^n$ of this group are called *canonical coordinates*.

As a result, one should note that if $\mathfrak{l}$ is homogeneous then the corresponding group is also homogeneous; we say that $\mathfrak{l}$ is homogeneous if there are positive exponents $a_1, \ldots, a_n$ so that

$$\sum_{j=1}^{n} u_j X_j \mapsto \sum_{j=1}^{n} u_j \delta^{\alpha_j} X_j$$

is an automorphism of $\mathfrak{l}$, for all $\delta > 0$.

Further discussion concerning nilpotent Lie groups and algebras may be found in Chapter 6 of Helgason [1962] and Chapter 3 of Varadarajan [1974].

7.8 An interesting class of homogeneous groups (of step 2) arises as direct generalizations of the Heisenberg group. These groups are sometimes referred to as H-type groups. We describe them in terms of their Lie algebras $\mathfrak{l}$. We assume that we are given a positive-definite inner product $\langle , \rangle$ on $\mathfrak{l}$, and an orthogonal decomposition $\mathfrak{l} = \mathfrak{l}_1 \oplus \mathfrak{l}_2$ so that

(i) $\mathfrak{l}_2$ is in the center of $\mathfrak{l}$, and

(ii) for each $T \in \mathfrak{l}_2$, $|T| = 1$, the mapping $J_T : \mathfrak{l}_1 \to \mathfrak{l}_2$ given by $\langle J_T(X), Y\rangle = \langle [X,Y], T\rangle$ is orthogonal.

Assuming these properties, we choose orthogonal bases $\{X_j\}_{j=1}^{m_1}$ and $\{T_j\}_{j=1}^{m_2}$ for $\mathfrak{l}_1$ and $\mathfrak{l}_2$ respectively, and use canonical coordinates $(u,t) \in \mathbf{R}^{m_1} \times \mathbf{R}^{m_2}$ to represent the point

$$\exp\Big(\sum_{j=1}^{m_1} u_j X_j + \sum_{j=1}^{m_2} t_j T_j\Big)(0)$$

in the simply connected group corresponding to $\mathfrak{l}$. One notes that the mappings

$$(u,t) \mapsto (\delta u, \delta^2 t), \qquad \delta > 0$$

are automorphisms. We can then assert the following.

(a) The "sub-Laplacian" $\mathfrak{L} = \sum_{j=1}^{m_1} X_j^2$ has a fundamental solution given by

$$c(|u|^4 + 16|t|^2)^{-(m_1+2m_2-2)/4}$$

for an appropriate $c > 0$. There is however no analogue of the operator $\mathfrak{L}_\alpha$, $\alpha \neq 0$, except when $m_2 = 1$ (which reduces essentially to the Heisenberg group).

(b) The nilpotent groups that arise as "boundaries" of rank-one symmetric spaces of Cartan are H-type groups.

(c) The analogue of the commutativity of the algebra of radial functions (see Chapter 12, §7.13) holds for certain groups of this class.

For (a), see Kaplan [1980]. The results (b) are in Damek [1987]; Cowling, Dooley, Korányi, and Ricci [1991]. For (c), see Kaplan and Ricci [1983], Ricci [1985].

7.9 A basic example of a homogeneous group is given by the "free" nilpotent group of k generators and d steps. Its Lie algebra $\mathfrak{f}^{k,d}$ can be defined as follows.

First, let $\mathfrak{f}^k$ be the free Lie algebra on k generators. To define $\mathfrak{f}^k$, consider $\mathcal{P}^k$, the associative algebra of all noncommutative polynomials in indeterminates $X_1, \ldots, X_k$ with real coefficients. The algebra $\mathcal{P}^k$ has a Lie bracket, namely the commutator $[A,B] = AB - BA$, $A, B \in \mathcal{P}^k$, which gives $\mathcal{P}^k$ a Lie algebra structure. The free Lie algebra $\mathfrak{f}^k$ is then the smallest Lie subalgebra of $\mathcal{P}^k$ that contains $X_1, \ldots, X_k$; $\mathfrak{f}^k$ is, of course, infinite dimensional.

Now for each $d \geq 0$, let $\mathfrak{i}_d$ be the ideal in $\mathfrak{f}^k$ generated by all commutators of length greater than d, and set

$$\mathfrak{f}^{k,d} = \mathfrak{f}^k / \mathfrak{i}_d.$$

The $\mathfrak{f}^{k,d}$ are finite-dimensional Lie algebras and enjoy the following important properties.

(a) For each $\delta > 0$, the mapping $X_j \mapsto \delta X_j$, $j = 1, \ldots, k$, extends to an automorphism of $\mathfrak{f}^k$, and thus to an automorphism of $\mathfrak{f}^{k,d}$. As a result, the simply connected Lie groups corresponding to the $\mathfrak{f}^{k,d}$ are homogeneous; the exponents a_j that occur in the dilations are integers are integers, with $1 \leq a_j \leq d$.

(b) Every nilpotent Lie algebra $\mathfrak{l}$ occurs as the quotient $\mathfrak{f}^{k,d}/\mathfrak{i}$, for appropriate k, d, and a suitable ideal $\mathfrak{i}$.

The properties of the free Lie algebras $\mathfrak{f}^k$ are described in Jacobson [1962]. The significance of the nilpotent Lie algebras $\mathfrak{f}^{k,d}$ was stressed in Auslander [1972], [1973].

7.10 Suppose G is a semisimple Lie group, and let $G = KAN$ be its Iwasawa decomposition. The abelian group A acts, by conjugation, as automorphisms of the nilpotent group N. The underlying space of N is $\mathbf{R}^n$, for some n. If we choose a suitable one-parameter subgroup $\{\gamma(t) : t \in \mathbf{R}\}$ in A (i.e., such that $\gamma'(0)$ lies in the "positive Weyl chamber"), then with

$$\delta = e^{-t} \qquad \text{and} \qquad \delta \circ x = \gamma(t) \cdot x \cdot [\gamma(t)]^{-1}, \quad x \in N,$$

the mappings $x \mapsto \delta \circ x$ are automorphisms of N, making $N = \mathbf{R}^n$ a homogeneous group.

Analysis on this homogeneous group plays a significant role in the representation theory of G, via the study of intertwining operators, and also in the behavior of harmonic functions, in particular Poisson integrals, on the symmetric space G/K.

The basic facts concerning semisimple Lie groups and the Iwasawa decomposition may be found in Helgason [1962]. For applications of homogeneous groups to representation theory, see Knapp and Stein [1971], Knapp [1986]. Poisson integrals for symmetric spaces are discussed in Chapter 2 and in §7.11 below.

7.11 The ideas in the theory of maximal functions, as given in chapters 1 and 2, are well illustrated when considered in the context of homogeneous groups.

(a) Suppose $\mathbf{R}^n$ is given the structure of a homogeneous group, with norm function ρ, as in §5.1. Define

$$(Mf)(x) = \sup_{0<r<\infty} r^{-a} \int_{\rho(y)\leq r} |f(x \cdot y^{-1})|\, dy,$$

where a is the homogeneous dimension. Then M satisfies the L^p, $p > 1$, and weak-type estimates in Chapter 1, §3. Moreover, in analogy with §6.1 of Chapter 1 and §2.1 of Chapter 2, we can state that $(M_\Phi f)(x) \leq c(Mf)(x)$, where

$$(M_\Phi f)(x) = \sup_{0<\delta<\infty} |(f * \Phi_\delta)(x)|, \qquad \Phi_\delta(x) = \delta^{-a}\Phi(\delta^{-1} \circ x),$$

assuming $|\Phi(x)| \leq c[1+\rho(x)]^{-a-\varepsilon}$, for some $\varepsilon > 0$.

(b) An analogue of the proof in §4.2.1 for "singular" maximal functions is also valid in this context. Namely if $\Phi \in L^1$ is such that

$$\int |\Phi(y^{-1} \cdot x) - \Phi(x)|\, dx \leq \eta(\rho(y)),$$

and

$$\int_{\rho(x)\geq R} |\Phi(x)|\, dx \leq \eta(R^{-1}), \qquad R \geq 1,$$

for some Dini modulus η, then the mapping

$$f \mapsto \sup_j |(f * \Phi_{2^j})(x)|$$

is of weak-type (1,1) and is bounded on L^p, $1 < p \leq \infty$.

(c) Applications of results of this kind arise in the proof of the almost-everywhere convergence to boundary values for Poisson integrals on symmetric spaces. For this, see Stein [1983b], Sjögren [1986]; a survey of earlier results is given in Korányi [1972].

7.12 On every homogeneous group there is a norm function $\tilde{\rho}$, equivalent with the ρ defined in §5.1.3, that has the subadditivity property with constant 1; that is

$$\tilde{\rho}(x \cdot y) \leq \tilde{\rho}(x) + \tilde{\rho}(y), \qquad \text{and} \qquad \tilde{\rho}(x^{-1}) = \tilde{\rho}(x).$$

More particularly:

(a) On the Heisenberg group, $\tilde{\rho}(x) = (|z|^4 + t^2)^{1/4}$, for $x = [z, t]$, has this property.

(b) On any homogeneous group, we can take $\tilde{\rho}$ so that $\tilde{\rho}(x) = 1$ on a Euclidean sphere of (sufficiently) small radius in canonical coordinates (and extending by homogeneity). Note that this $\tilde{\rho}$ is C^∞ away from the origin.

(c) Observe that, in $\mathbf{R}^n$ with its usual additive structure, if

$$\rho(x) = \max_j |x_j|^{1/a_j},$$

then ρ has the subadditive property if $a_j \geq 1$ for all j.

Cygan [1981], Hebisch and Sikora [1990]; some of these results are implicit in Guivarch [1973].

C. Applications of homogeneous groups

7.13 The analysis of $\bar{\partial}_b$ and $\Box_b$ on the Heisenberg group in §2 and §3 can be used as a decisive tool in the corresponding study for strictly pseudo-convex domains. Suppose $\mathcal{D} \subset \mathbf{C}^{n+1}$ is such a domain with smooth boundary. Then, near any point on the boundary, we can choose a basis of the tangential Cauchy-Riemann vector fields $\overline{W}_1, \dots, \overline{W}_n$ so that (11) holds and in particular, after a suitable change of basis, so that

$$[\overline{W}_j, W_k] = 2i\delta_{jk}S \quad \text{modulo } \{W_j, \overline{W}_j\}_{j=1}^n.$$

We can now model these vector fields on the corresponding vector fields Z_j, $\bar{Z}_j$, T of the Heisenberg group. For each point $Q \in b\mathcal{D}$, we introduce a coordinate system centered at Q so that, if $P \in b\mathcal{D}$ is near Q, then P is given coordinates $(z,t) \in \mathbf{C}^n \times \mathbf{R}$ with

$$\exp\Big(\sum_{j=1}^n (z_j W_j + \bar{z}_j \overline{W}_j) + tS\Big)(Q) = P.$$

This defines a function Θ from a neighborhood of the diagonal in $b\mathcal{D} \times b\mathcal{D}$ to $\mathbf{H}^n$, namely

$$\Theta(P,Q) = [z,t] \in \mathbf{H}^n.$$

Note that $\Theta(P,Q) = -\Theta(Q,P)$, and that when $\mathcal{D} = \mathcal{U}^n$ (with $W_j = Z_j$, $S = T$) we have that $\Theta(P,Q) = Q^{-1} \cdot P$ is given by group multiplication on $\mathbf{H}^n$.

The key point is the approximation of W_j by Z_j, which can be formulated as follows. Suppose we fix Q and let W_j denote differentiation with respect to the P variable. Then, in the above coordinate system,

$$W_j = \frac{\partial}{\partial z_j} + i\bar{z}_j \frac{\partial}{\partial t} + \text{"higher order" terms}, \quad \text{as } P \to Q,$$

where "higher order" is in terms of the Heisenberg group homogeneity of z and t.

As a consequence, in the presence of a suitable Hermitian metric (a "Levi" metric), the Kohn Laplacian $\Box_b$ on q-forms has an approximate inverse, when $0 < q < n$, given by the operator

$$f \mapsto \int_{b\mathcal{D}} F_\alpha(\Theta(P,Q))\, f(Q)\, dQ \tag{$*$}$$

where $\alpha = n - 2q$ and $F_\alpha(z,t) = \gamma_\alpha^{-1}\phi_\alpha(z,t)$, as in §2.

For the development of this point of view and its applications, see Folland and Stein [1974]. Some further relevant results are:

(i) The technical requirement of the Levi metric is removed in Rothschild and Stein [1976].

(ii) Using the approach described in the appendix to this chapter, the results for $\Box_b$ above were applied to the $\bar{\partial}$-Neumann problem for strictly pseudoconvex domains; see Greiner and Stein [1977]. Generalizations may be found in R. Beals, Greiner, and Stanton [1987], D. Chang [1987].

(iii) Some alternative approaches to the analysis on strictly pseudoconvex boundaries via "approximation" by the Heisenberg group have been given by Dynin [1975], Taylor [1984], R. Beals and Greiner [1988]. The original technique, in terms of the Θ map described above, admits wide generalizations, as we will see in §7.14 below.

(iv) A theory of pseudo-differential operators, covering the solving operator $(*)$, as well as the examples arising in §7.15 of Chapter 12, has been described in Nagel and Stein [1979].

7.14 Let $X_1, \dots, X_k$ be a collection of smooth real vector fields, defined in an open subset of $\mathbf{R}^n$, so that their commutators (of length at most d) span the tangent space at every point. We describe two steps in the analysis of these vector fields, *lifting* and *approximation.*

(a) Lifting: Consider $\mathfrak{f}^{k,d}$, the free nilpotent Lie algebra of k generators and d steps, as described in §7.9. Suppose $\mathcal{N}$ is the simply connected Lie group whose Lie algebra is $\mathfrak{f}^{k,d}$, where the underlying space of $\mathcal{N}$ is identified with $\mathbf{R}^N$. Write

$$\mathbf{R}^N = \{(x,y) : x \in \mathbf{R}^n,\ y \in \mathbf{R}^{N-n}\}.$$

Then there exist vector fields $\tilde{X}_1, \dots, \tilde{X}_n$ in some open subset of $\mathbf{R}^N$ so that

(i) $\tilde{X}_j = X_j + a_j(x,y) \cdot \partial_y$, $1 \le j \le k$, and

(ii) The $\tilde{X}_j$ are free up to step d.

Property (ii) can be understood as follows: Suppose $Y_1, \dots, Y_k$ is a basis for the vector fields in $\mathfrak{f}^{k,d}$ having degree 1. For any ordered index set $I = (i_1, \dots, i_\ell)$, $1 \le i_j \le k$, write $|I| = \ell$ and

$$\tilde{X}_I = [X_{i_1}[X_{i_2}[\cdots[X_{i_{\ell-1}}, X_{i_\ell}]\cdots]]],$$

with a similar notation for Y_I. Then at any point

$$\dim \operatorname{span}\{\tilde{X}_I : |I| \le r\} = \dim \operatorname{span}\{Y_I : |I| \le r\}$$

for $1 \le r \le d$.

(b) Approximation: For any $P, Q \in \mathbf{R}^N$ that are sufficiently near each other, define $\Theta(P,Q) \in \mathcal{N}$ by

$$\Theta(P,Q) = \exp\Big(\sum_I u_I Y_I\Big)(0) \in \mathcal{N}, \quad \text{where } P = \exp\Big(\sum_I u_I \tilde{X}_I\Big)(Q).$$

Here the summation over I ranges over a collection of index sets so that $\{Y_I\}$ forms a basis of $\mathfrak{f}^{k,d}$.

The mapping Θ gives a coordinate system centered at Q, for P near Q; the coordinates are $\{u_I\}$. In these coordinates

$$\tilde{X}_j = Y_j + R_j, \qquad j = 1, \dots, k,$$

if the Y_j are expressed in the coordinates $\{u_I\}$, the error term R_j is of higher order (in terms of the homogeneity of $\mathcal{N}$).

As a result, if we consider the operator

$$\mathfrak{L} = \sum_{j=1}^{k} \tilde{X}_j^2,$$

it has an approximate fundamental solution, analogous to $(*)$ in §7.13, in terms of the (exact) homogeneous fundamental solution of $\sum_1^k Y_j^2$. By a simple procedure of descent (integration in the added variables), one then obtains an approximate fundamental solution of $\sum_1^k X_j^2$.

The approach outlined here is detailed in Rothschild and Stein [1976]. The technique of lifting has been given a more streamlined proof in Goodman [1976]. The operator $\mathfrak{L}$ was introduced in Hörmander [1967a], where its subellipticity was proved. The technique of lifting and approximation can also be used to study singular Radon transforms; see Chapter 11, §4.7.

7.15 The free nilpotent Lie algebras $\mathfrak{f}^{k,d}$ may be used, in a procedure analogous to the descent technique occurring in §2.4 of Chapter 11, to prove analogues of theorems on singular integrals and maximal functions for arbitrary nilpotent Lie groups (not just homogeneous groups). We state one particular result of this kind.

Let $\mathfrak{l}$ be an arbitrary nilpotent Lie algebra of dimension n, and suppose we realize its corresponding simply connected Lie group via canonical coordinates on $\mathbf{R}^n$, as in §7.7. Assume that the distribution K is a homogeneous Calderón-Zygmund distribution of the type characterized in Chapter 1, §8.19. Then, on this nilpotent group, the convolution operator $f \mapsto f * K$ extends to a bounded operator from L^p to itself, for $1 < p < \infty$. For further details, see Ricci and Stein [1988].

D. Miscellaneous topics

7.16 Let f be (say) a C^1 function defined on an open subset O of the boundary of the domain $\mathcal{U}^n$. Assume that f satisfies the tangential Cauchy-Riemann equations there (i.e., $\bar{\partial}_b f = 0$ on O). Then there exists an F, defined and continuous for all points in $\overline{\mathcal{U}}^n$ sufficiently close to O, holomorphic in the interior, and so that $F|_O = f$.

A general result of this kind, due to Lewy [1956], holds for any domain $\mathcal{D}$ whose Levi form has at least one positive eigenvalue; see also the account in Hörmander [1966]. In the above case (for the domain $\mathcal{U}^n$), it is possible to write a simple expression for F in terms of f.

We let $d\sigma$ denote the normalized invariant measure on the unit sphere of $\mathbf{C}^n$. When $z = (z', z_{n+1}) \in O$ and ε is sufficiently small, we have

$$F(z + \varepsilon\mathbf{i}) = F(z', z_{n+1} + \varepsilon i) = \int\limits_{|\zeta|=1} f(z' + \varepsilon^{1/2}\zeta, z_{n+1} + i\varepsilon + 2i\varepsilon^{1/2} z' \cdot \bar{\zeta})\, d\sigma(\zeta).$$

7.17 A variety of questions in harmonic analysis on $\mathbf{H}^n$ have been pursued, prompted in part by the analogy with problems in Fourier analysis on $\mathbf{R}^n$. We indicate briefly several of these directions.

(i) A multiplier theorem, dealing with operators given as functions of $\mathfrak{L}$ and T akin to those in §4.4 and §7.6 of Chapter 6; see de Michele and Mauceri [1979].

(ii) Further questions related to the spectral theory of $\mathfrak{L}$ and Radon transforms; in Strichartz [1991].

(iii) The analogue, as far as it is valid, of the (L^p, L^2) restriction theorem given in §2.1 of Chapter 9; see D. Müller [1990b].

7.18 Notions occurring in analysis on the Heisenberg group can be used in a problem of mathematical economics, namely the determination of the probablistic distributions of stock prices and the resulting formulas for option prices. The background for this is as follows.

If P is the price of a stock then, according to a standard model, it satisfies the stochastic differential equation

$$dP = \mu\, dt + \sigma P\, d\omega. \tag{$*$}$$

Here $d\omega$ is a Wiener process (describing Brownian motion), the quantities μ and σ are constants with σ representing the volatility of the stock. The probability distribution at time t determined by $(*)$ is given as the solution of a one-dimensional heat equation; as is well-known, it is a (log) normal distribution.

In certain settings, the assumption that the volatility σ is constant is not realistic. Instead, it is reasonable to assume that σ is itself a stochastic variable. The simplest hypothesis of this kind is that σ is governed by an Ornstein-Uhlenbeck-like process that satisfies the stochastic differential equation

$$d\sigma = \delta(\theta - \sigma)\, dt + k\, d\omega'. \tag{$**$}$$

Here $d\omega'$ is another Wiener process, independent of $d\omega$. The constant k determines the degree of stochastic change of σ, θ is the long-term mean of σ, and δ is a mean-reversion parameter.

The relevant probability distributions are then determined by the solution of an appropriate two-dimensional heat equation, which can be connected with the heat equation on the Heisenberg group given in §7.2. Exact formulas for the stock price distributions can be derived from this, or, more directly, by solving the differential equation

$$\left(\frac{1}{2}\frac{\partial^2}{\partial x^2} + (Ax + b)\frac{\partial}{\partial x} + Cx^2\right) = \frac{\partial U}{\partial t}(x, t) \tag{$***$}$$

with the initial condition $U(x, 0) \equiv 1$. It can be shown that

$$U(x, t) = \exp(Lx^2/2 + Mx + N),$$

where L, M, N are elementary (but complicated) functions of A, B, C, and t.

The stock-price distributions thus determined can be used to obtain a formula for option prices, giving a modification of the classical formula due to Black and Scholes [1973], originally derived in the context of $(*)$ alone (i.e., when σ is constant). Further details concerning the above are in E. Stein and J. Stein [1991]. It should also be noted the the solution of $(***)$ bears some relation to the "oscillator semigroup" as treated in, e.g., Howe [1988], D. Müller and Ricci [1990].

Notes

§*1.* The monograph of Folland and Kohn [1972] contains an account of the $\bar{\partial}$-complex, the $\bar{\partial}$-Neumann problem, and their boundary analogues, as well as further references to the literature. An alternative treatment of the problem $\bar{\partial}u = f$, using integral representation formulas, is in Range [1986].

§*2 and* §*3.* The results in these sections are taken from Folland and Stein [1974].

§*4.* The general nonsolvability of the Lewy equation was proved in Lewy [1957]. The necessary and sufficient conditions (Theorem 3) are in Greiner, Kohn, and Stein [1975]; some earlier considerations in this direction may be found in Sato, Kawai, and Kashiwara [1973].

§*5.* The role of the Heisenberg group (as well as more general homogeneous groups) in the $\bar{\partial}$-problem and other nonelliptic problems, was surmised in Stein [1970b]. The L^2 boundedness in §5.3 was proved in Knapp and Stein [1971]; the argument given here incorporates a suggestion of Ricci.

Bibliography

Aguilera, N., and C. Segovia

[1977] Weighted norm inequalities relating the g_λ^* and area functions, *Studia Math.* **61**: 293–303.

Andersen, K. F., and R. T. John

[1980] Weighted inequalities for vector-valued maximal functions and singular integrals, *Studia Math.* **69**: 19–31.

Arkhipov, G. I., and K. I. Oskolkov

[1987] On a special trigonometric series and its applications, *Mat. Sb.* **134**: 147–158.

Atiyah, M. F.

[1970] Resolution of singularities and division of distributions, *Comm. Pure Appl. Math.* **23**: 145–150.

Auslander, L.

[1972] *Personal communication.*

[1973] An exposition of the structure of solv-manifolds, *Bull. Amer. Math. Soc.* **79**: 227–261.

Balabane, M.

[1989] On the regularizing effect on Schrödinger-type groups, *Ann. Inst. Henri Poincaré* **6**: 1–14.

Beals, M.

[1982] *L^p Boundedness of Fourier Integral Operators.* Memoirs of the AMS #264, American Math. Society.

Beals, M., C. Fefferman, and R. Grossman

[1983] Strictly pseudoconvex domains, *Bull. Amer. Math. Soc.* **8**: 125–322.

Beals, R.

[1975] A general calculus of pseudo-differential operators, *Duke Math. J.* **42**: 1–42.

[1977] Characterization of pseudo-differential operators and applications, *Duke Math. J.* **44**: 45–57.

Beals, R., and C. Fefferman

[1973] On local solvability of linear partial differential equations, *Ann. of Math.* **97**: 482–498.

[1974] Spatially inhomogeneous pseudo-differential operators, *Comm. Pure Appl. Math.* **27**: 1–24.

Beals, R., and P. C. Greiner
[1988] *Calculus on Heisenberg Manifolds.* Annals of Math. Study #119, Princeton Univ. Press.

Beals, R., P. C. Greiner, and N. K. Stanton
[1987] L^p and Lipschitz estimates for the $\bar{\partial}$-equation and the $\bar{\partial}$-Neumann problem, *Math. Ann.* **277**: 185–196.

Beckner, W.
[1975] Inequalities in Fourier Analysis, *Ann. of Math.* **102**: 159–182.

Beckner, W., A. P. Calderón, R. Fefferman, and P. W. Jones (eds.)
[1983] *Conference on Harmonic Analysis in Honor of Antoni Zygmund.* Wadsworth.

Benedek, A., A. P. Calderón, and R. Panzone
[1962] Convolution operators on Banach-space valued functions, *Proc. Nat. Acad. Sci. U.S.A.* **48**: 356–365.

Bennett, C., R. A. DeVore, and R. Sharpley
[1981] Weak-L^∞ and BMO, *Ann. of Math.* **113**: 601–611.

Bennett, C., and R. Sharpley
[1988] *Interpolation of Operators.* Pure and Applied Math. #129, Academic Press.

Berenstein, A., and E. T. Sawyer
[1985] *Embedding and Multiplier Theorems for $H^p(\mathbf{R}^n)$.* Memoirs of the AMS #318, American Math. Society.

Bergh, J., and J. Löfström
[1976] *Interpolation Spaces.* Grund. der math. Wiss. #223, Springer-Verlag.

Bernstein, I. N., and S. I. Gelfand
[1969] The function P^λ is meromorphic, *Funct. Anal. Appl.* **3**: 68–69.

Besicovitch, A. S.
[1928] On Kakeya's problem and a similar one, *Mat. Zeit.* **27**: 312–320.
[1945] A general form of the covering principle and relative differentiation of additive functions, *Proc. Cambridge Philos. Soc.* **41**: 103–110.

Beurling, A., and L. Ahlfors
[1956] The boundary correspondence under quasi-conformal mappings, *Acta Math.* **96**: 125–142.

Björck, J. E.
[1973] On Fourier transforms of smooth measures carried by real-analytic submanifolds of $\mathbf{R}^n$, *preprint.*

Black, F., and M. Scholes
[1973] The pricing of options and corporate liabilities, *J. of Political Economy* **81**: 637–654.

Bony, J. M.
[1981] Calcul symbolique et propagation des singularités pour les equations aux dérivées partielles non linéares, *Ann. Sci. École Norm. Sup.* **14**: 209–256.

Boothby, W., and G. Weiss (eds.)
[1972] *Symmetric Spaces.* Marcel Dekker.

Bourdaud, G.
[1988] Une algèbre maximale d'opérateurs pseudo-differentiels, *Comm. P.D.E.* **13**: 1059–1083.

Bourgain, J.
[1983] Some remarks on Banach spaces in which martingale difference sequences are unconditional, *Ark. f. Mat.* **21**: 163–168.
[1986a] Averages in the plane over convex curves and maximal operators, *J. Analyse Math.* **47**: 69–85.
[1986b] Vector-valued singular integrals and H^1-BMO duality, *in* Chao and Woyczyinski [1986], 1–79.
[1986c] On the L^p bounds for maximal functions associated to convex bodies, *Israel J. Math.* **54**: 307–316.
[1986d] On high-dimensional maximal functions associated to convex bodies, *Amer. J. Math.* **108**: 1467–1476.
[1987] On dimension-free maximal inequalities for convex symmetric bodies in $\mathbf{R}^n$, *in* Lindenstrauss and Milman [1987], 168–176.
[1991a] Besicovitch-type maximal operators and applications to Fourier analysis, *Geom. and Funct. Anal.* **1**: 147–187.
[1991b] L^p estimates for oscillatory integrals in several variables, *Geom. and Funct. Anal.* **1**: 321–374.

Brenner, P.
[1977] $(L^p, L^{p'})$ estimates for Fourier integral operators related to hyperbolic equations, *Mat. Zeit.* **152**: 273–286.

Bruna, J., A. Nagel, and S. Wainger
[1988] Convex hypersurfaces and Fourier transforms, *Ann. of Math.* **127**: 333–365.

Burkholder, D. L.
[1964] Maximal inequalities as necessary conditions for almost-everywhere convergence, *Zeit f. War.* **3**: 75–88.
[1966] Martingale transforms, *Ann. of Math. Stat.* **37**: 1494–1505.
[1973] Distribution function inequalities for martingales, *Ann. Prob.* **1**: 19–42.
[1979] Martingale theory and harmonic analysis in Euclidean spaces, *in* Weiss and Wainger [1979], part 2, 283–301.
[1981] A geometrical characterization of Banach spaces in which martingale difference sequences are unconditional, *Ann. Prob.* **9**: 997–1011.
[1983] A geometric condition that implies the existence of certain singular integrals of Banach-space-valued functions, *in* Beckner *et al.* [1983], 270–286.

Burkholder, D. L., and R. F. Gundy
[1970] Extrapolation and interpolation of quasilinear operators on martingales, *Acta Math.* **124**: 249–304 .
[1972] Distribution function inequalities for the area integral, *Studia Math.* **44**: 527–544.
[1973] Boundary behavior of harmonic functions in a half-space and Brownian motion, *Ann. Inst. Fourier* **23**: 195–212.

Burkholder, D. L., R. F. Gundy, and M. L. Silverstein
[1971] A maximal function characterization of the class H^p, *Trans. Amer. Math. Soc.* **157**: 137–153.

Busemann, H., and W. Feller
[1934] Differentiation der L-integrale, *Fund. Math.* **22**: 226–256.

Caffarelli, L., E. Fabes, and C. E. Kenig
[1981] Completely singular elliptic-harmonic measures, *Indiana U. Math. J.* **30**: 917–924.

Caffarelli, L., E. Fabes, S. Mortola, and S. Salsa
[1981] Boundary behavior of non-negative solutions of elliptic operators in divergence form, *Indiana U. Math. J.* **30**: 621–640.

Calderón, A. P.
[1953] A general ergodic theorem, *Ann. of Math.* **58**: 182–191.
[1965] Commutators of singular integral operators, *Proc. Nat. Acad. Sci. U.S.A.* **53**: 1092–1099.
[1967] *Singular Integrals (ed.).* Proc. Symp. Pure Math. #10, American Math. Society.
[1977a] Cauchy integrals on Lipschitz curves and related operators, *Proc. Nat. Acad. Sci. U.S.A.* **74**: 1324–1327.
[1977b] An atomic decomposition of distributions in parabolic H^p spaces, *Adv. in Math.* **25**: 216–225.

Calderón, A. P., and A. Torchinsky
[1975] Parabolic maximal functions associated with a distribution, *Adv. in Math.* **16**: 1–64.
[1977] Parabolic maximal functions associated with a distribution II, *Adv. in Math.* **24**: 101–171.

Calderón, A. P., and R. Vaillancourt
[1972] A class of bounded pseudo-differential operators, *Proc. Nat. Acad. Sci. U.S.A.* **69**: 1185–1187.

Calderón, A. P., and A. Zygmund
[1950] On the theorem of Hausdorff-Young and its extensions, *in* Zygmund *et al.* [1950], 166–188.
[1952] On the existence of certain singular integrals, *Acta Math.* **88**: 85–139.
[1957] Singular integral operators and differential equations, *Amer. J. Math.* **79**: 801–821.
[1961] Local properties of solutions of elliptic partial differential equations, *Studia Math.* **20**: 171–225.

Carbery, A.
[1983] The boundedness of the maximal Bochner-Riesz operator on $L^4(\mathbf{R}^2)$, *Duke Math. J.* **50**: 409–416.
[1985] A weighted inequality for the maximal Bochner-Riesz operator on $\mathbf{R}^2$, *Trans. Amer. Math. Soc.* **287**: 673–680.
[1986a] Variants of the Calderón-Zygmund theory for L^p spaces, *Rev. Mat. Iberoamer.* **2**: 381–396.
[1986b] An almost-orthogonality principle with applications to maximal functions associated to convex bodies, *Bull. Amer. Math. Soc.* **14**: 269–273.

Carbery, A. (*continued*)
[1988] Differentiation in lacunary directions and an extension of the Marcinkiewicz multiplier theorem, *Ann. Inst. Fourier* **38**: 157–168.

Carbery, A., M. Christ, J. Vance, S. Wainger, and D. Watson
[1989] Operators associated to flat plane curves: L^p estimates via dilation methods, *Duke Math. J.* **59**: 675–700.

Carbery, A., and A. Seeger
[1993] H^p and L^p variants of multiparameter Calderón-Zygmund theory, *Trans. Amer. Math. Soc.* **334**: 719–747.

Carleson, L.
[1962] Interpolation by bounded analytic functions and the corona problem, *Ann. of Math.* **76**: 547–559.
[1966] On convergence and growth of partial sums of Fourier series, *Acta Math.* **116**: 135–157.
[1976] Two remarks on H^1 and BMO, *Adv. in Math.* **22**: 269–277.
[1980] An explicit unconditional basis in H^1, *Bull. des Sci. Math.* **104**: 405–416.

Carleson, L., and P. Sjölin
[1972] Oscillatory integrals and a multiplier problem for the disc, *Studia Math.* **44**: 287–299.

Carlsson, H., M. Christ, A. Córdoba, J. Douandikoetxea, J. L. Rubio de Francia, J. Vance, S. Wainger, and D. Weinberg
[1986] L^p estimates for maximal functions and Hilbert transforms, *Bull. Amer. Math. Soc.* **14**: 263–267.

Carlsson, H., P. Sjögren, and J.-O. Strömberg
[1985] Multiparameter maximal functions along dilation-invariant hypersurfaces, *Trans. Amer. Math. Soc.* **292**: 335–343.

Catlin, D.
[1983] Necessary conditions for subellipticity of the $\bar\partial$-Neumann problem, *Ann. of Math.* **117**: 147–171.
[1986] Subelliptic estimates for the $\bar\partial$-Neumann problem on pseudoconvex domains, *Ann. of Math.* **126**: 131–191.

Chang, D.-C.
[1987] *On L^p and Hölder Estimates for the $\bar\partial$-Neumann Problem on Strongly Pseudoconvex Domains.* Ph. D. Dissertation, Princeton University.

Chang, D.-C., S. G. Krantz, and E. M. Stein
[1992] Hardy spaces and elliptic boundary-value problems, *in* Nagel [1992], 119–131.

Chang, D.-C., A. Nagel, and E. M. Stein
[1988] Estimates for the $\bar\partial$-Neumann problem in pseudoconvex domains in $\mathbf{C}^2$ of finite type, *Proc. Nat. Acad. Sci. U.S.A.* **85**: 8771–8774.

Chang, S.-Y. A., and R. Fefferman
[1980] A continuous version of the duality of H^1 and BMO on the bidisc, *Ann. of Math.* **112**: 179–201.
[1982] The Calderón-Zygmund decomposition on product domains, *Amer. J. Math.* **104**: 455–468.

Chang, S.-Y. A., and R. Fefferman (*continued*)
[1985] Some recent developments in Fourier analysis and H^p theory on product domains, *Bull. Amer. Math. Soc.* **12**: 1–43.

Chang, S.-Y. A., J. M. Wilson, and T. Wolff
[1985] Some weighted norm inequalities concerning the Schrödinger operator, *Comm. Math. Helv.* **60**: 217–246.

Chanillo, S., and M. Christ
[1987] Weak $(1,1)$ bounds for oscillatory singular integrals, *Duke Math. J.* **55**: 141–155.

Chanillo, S., D. Kurtz, and G. Sampson
[1986] Weighted weak $(1,1)$ estimates for oscillatory kernels, *Trans. Amer. Math. Soc.* **295**: 127–145.

Chao, J. A., J. E. Gilbert, and P. A. Tomas
[1981] Molecular decompositions in H^p theory, *Rend. Circ. Mat. Palermo suppl.* **1**: 115–119.

Chao, J. A., and S. Janson
[1981] A note on H^1 q-martingales, *Pacific J. Math.* **97**: 307–317.

Chao, J. A., and W. A. Woyczyinski (eds.)
[1986] *Probability Theory and Harmonic Analysis.* Marcel Dekker.

Chen, L. K.
[1987] On a singular integral, *Studia Math.* **85**: 61–72.

Chern, S.-S., and S. Smale (eds.)
[1970] *Global Analysis (3 volumes).* Proc. Symp. Pure Math. #14–16, American Math. Society.

Christ, M.
[1985a] Hilbert transforms along curves: I. Nilpotent groups, *Ann. of Math.* **122**: 575–596.
[1985b] On almost-everywhere convergence of Bochner-Riesz means in higher dimensions, *Proc. Amer. Math. Soc.* **95**: 16–20.
[1988] Regularity properties of the $\bar{\partial}_b$ equation on weakly pseudoconvex CR manifolds of dimension 3, *J. Amer. Math. Soc.* **1**: 587–646.
[1990] A $T(b)$ theorem with remarks on analytic capacity and the Cauchy integral, *Colloq. Math.* **61**: 601–628.

Christ, M., J. Duoandikoetxea, and J. L. Rubio de Francia
[1986] Maximal operators related to the Radon transform and the Calderón-Zygmund method of rotations, *Duke Math. J.* **53**: 189–209.

Christ, M., A. Nagel, E. M. Stein, and S. Wainger
[1993] *Singular and Maximal Radon Transforms.* Annals of Math. Study, Princeton Univ. Press (*to appear*).

Christ, M., and J. L. Rubio de Francia
[1988] Weak-type $(1,1)$ bounds for rough operators II, *Invent. Math.* **93**: 225–237.

Christ, M., and C. D. Sogge
[1988] The weak-type L^1 convergence of eigenfunction expansions for pseudo-differential operators, *Invent. Math.* **94**: 421–453.

Christ, M., and E. M. Stein
[1987] A remark on singular Calderón-Zygmund theory, *Proc. Amer. Math. Soc.* **99**: 71–75.

Coifman, R. R.
[1972] Distribution function inequalities for singular integrals, *Proc. Nat. Acad. Sci. U.S.A.* **69**: 2838–2839.
[1974] A real-variable characterization of H^p, *Studia Math.* **51**: 269–274.

Coifman, R. R., and C. Fefferman
[1974] Weighted norm inequalities for maximal functions and singular integrals, *Studia Math.* **51**: 241–250.

Coifman, R. R., and L. Grafakos
[1992] Hardy space estimates for multilinear operators, *to appear.*

Coifman, R. R., P. W. Jones, and J. L. Rubio de Francia
[1983] Constructive decomposition of BMO functions and factorization of A_p weights, *Proc. Amer. Math. Soc.* **87**: 665–666.

Coifman, R. R., P. W. Jones, and S. Semmes
[1989] Two elementary proofs of the L^2 boundedness of Cauchy integrals on Lipschitz curves, *J. Amer. Math. Soc.* **2**: 553–564.

Coifman, R. R., P. L. Lions, Y. Meyer, and S. Semmes
[1989] Compacité par compensation et espaces de Hardy, *C. R. Acad. Sci. Paris* **309**: 945–949.

Coifman, R. R., A. McIntosh, and Y. Meyer
[1982] L'integrale de Cauchy sur les courbes lipschitziennes, *Ann. of Math.* **116**: 361–387.

Coifman, R. R., and Y. Meyer
[1978] *Au-delà des opérateurs pseudo-différentiels.* Astérisque #57, Société Math. de France.

Coifman, R. R., Y. Meyer, and E. M. Stein
[1985] Some new function spaces and their applications to harmonic analysis, *J. Funct. Anal.* **62**: 304–335.

Coifman, R. R., and R. Rochberg
[1980] Another characterization of BMO, *Proc. Amer. Math. Soc.* **79**: 249–254.

Coifman, R. R., R. Rochberg, and G. Weiss
[1976] Factorization theorems for Hardy spaces in several variables, *Ann. of Math.* **103**: 611–635.

Coifman, R. R., and G. Weiss
[1971] *Analyse harmonique non-commutative sur certains espaces homogenes.* Lecture Notes in Math. #242, Springer-Verlag.
[1977a] Extensions of Hardy spaces and their use in analysis, *Bull. Amer. Math. Soc.* **83**: 569–645.
[1977b] *Transference Methods in Analysis.* CBMS Regional Conference Ser. #31, American Math. Society.

Constantin, P., and J. Saut
[1988] Local smoothing properties of dispersive equations, *J. Amer. Math. Soc.* **1**: 431–439.

Copson, E. T.
[1965] *Asymptotic Expansions.* Cambridge Tracts in Math. and Math. Physics #55, Cambridge Univ. Press.

Cordes, H.
[1975] On compactness of commutators of multiplications and convolutions and boundedness of pseudo-differential operators, *J. Funct. Anal.* **18**: 115–131.

Córdoba, A.
[1977] The Kakeya maximal function and spherical summation multipliers, *Amer. J. Math.* **99**: 1–22.
[1979] A note on Bochner-Riesz operators, *Duke Math. J.* **46**: 505–511.
[1982] Geometric Fourier analysis, *Ann. Inst. Fourier* **32**: 215–226.

Córdoba, A., and C. Fefferman
[1976] A weighted norm inequality for singular integrals, *Studia Math.* **57**: 97–101.

Córdoba, A., and R. Fefferman
[1975] A geometric proof of the strong maximal theorem, *Ann. of Math.* **102**: 95–100.
[1977a] On differentiation of integrals, *Proc. Nat. Acad. Sci. U.S.A.* **74**: 2211–2213.
[1977b] On the equivalence between the boundedness of certain classes of maximal and multiplier operators in Fourier analysis, *Proc. Nat. Acad. Sci. U.S.A.* **74**: 423–425.

Córdoba, A., and B. Lopez-Melero
[1981] Spherical summation: A problem of E. M. Stein, *Ann. Inst. Fourier* **31**: 147–157.

Corput, J. G. van der
[1921] Zahlentheoretische Abschätzungen, *Math. Ann.* **84**: 53–79.

Cotlar, M.
[1955] A combinatorial inequality and its application to L^2 spaces, *Rev. Math. Cuyanā* **1**: 41–55.

Cotlar, M., and R. Cignoli
[1974] *An Introduction to Functional Analysis.* North-Holland.

Cotlar, M., and C. Sadosky
[1979] On the Helson-Szegö theorem and a related class of modified Toeplitz kernels, *in* Weiss and Wainger [1979], part 1, 383–407.
[1983] On some L^p versions of the Helson-Szegö theorem, *in* Beckner *et al.* [1983], 306–317.
[1990] Two-parameter lifting theorems and double Hilbert transforms in commutative and noncommutative settings, *J. Math. Anal. and Appl.* **150**: 439–480.

Cowling, M., S. Disney, G. Mauceri, and D. Müller
[1990] Damping oscillatory integrals, *Invent. Math.* **101**: 237–260.

Cowling, M., A. Dooley, A. Korányi, and F. Ricci
[1991] H-type groups and Iwasawa decompositions, *Adv. in Math.* **87**: 1–41.

Cowling, M., and G. Mauceri
[1987] Oscillatory integrals and Fourier transforms of surface-carried measures, *Trans. Amer. Math. Soc.* **304**: 53–68.

Cwickel, M.
[1987] *Personal communication.*

Cwickel, M., J. Peetre, Y. Sagher, and H. Wallin (eds.)
[1988] *Function Spaces and Applications (1986 Proceedings).* Lecture Notes in Math. #1302, Springer-Verlag.

Cygan, J.
[1981] Subadditivity of homogeneous norms on certain nilpotent Lie groups, *Proc. Amer. Math. Soc.* **83**: 69–70.

Dahlberg, B.
[1977] Estimates of harmonic measure, *Arch. for Rat. Mech. and Anal.* **65**: 275–288.
[1979] On the Poisson integral for Lipschitz and C^1 domains, *Studia Math.* **66**: 13–24.

Dahlberg, B., and C. E. Kenig
[1987] Hardy spaces and the L^p Neumann problem for Laplace's equation in a Lipschitz domain, *Ann. of Math.* **125**: 437–465.

Damek, E.
[1987] Geometry of a semi-direct extension of a Heisenberg-type nilpotent group, *Colloq. Math.* **53**: 255–268.

Daubechies, I.
[1988] Orthonormal bases of compactly supported wavelets, *Comm. Pure Appl. Math.* **61**: 909–996.

David, G.
[1984] Opérateurs intégraux singuliers sur certaines courbes du plan complexe, *Ann. Sci. École Norm. Sup.* **17**: 157–189.
[1988] Morceaux de graphes lipschitziens et intégrales singulières sur une surface, *Rev. Mat. Iberoamer.* **4**: 73–114.
[1991] *Wavelets and Singular Integrals on Curves.* Lecture Notes in Math. #1465, Springer-Verlag.

David, G., and J.-L. Journé
[1984] A boundedness criterion for generalized Calderón-Zygmund operators, *Ann. of Math.* **120**: 371–397.

David, G., J.-L. Journé, and S. Semmes
[1985] Opérateurs de Calderón-Zygmund, fonctions para-accrétives et interpolation, *Rev. Mat. Iberoamer.* **1**: 1–56.

Davies, E.
[1988] Lipschitz continuity of functions of operators in Schatten classes, *J. London Math. Soc.* **37**: 148–157.

Davies, R. O.
[1952] On accessibility of plane sets and differentiation of functions of two real variables, *Proc. Cambridge Philos. Soc.* **48**: 215–232.

Davis, B.
[1970] On the integrability of the martingale square function, *Israel J. Math.* **8**: 187–190.

Davis, K. M., and Y.-C. Chang
[1987] *Lectures on Bochner-Riesz Means.* London Math. Society Lecture Notes #114, Cambridge Univ. Press.

de Guzmán, M.
[1975] *Differentiation of Integrals in* $\mathbf{R}^n$. Lecture Notes in Math. #481, Springer-Verlag.
[1981] *Real-Variable Methods in Fourier Analysis.* Math. Study #46, North-Holland.

deLeeuw, K.
[1965] On L^p multipliers, *Ann. of Math.* **81**: 364–379.

de Michele, L., and G. Mauceri
[1979] L^p multipliers on the Heisenberg group, *Michigan Math. J.* **26**: 361–371.

Deng, D. G.
[1984] On a generalized Carleson inequality, *Studia Math.* **78**: 245–251.

Doob, J. L.
[1953] *Stochastic Processes.* Wiley.

Duistermaat, J. J.
[1973] *Fourier Integral Operators.* Courant Institute of Math. Sciences, New York Univ.

Durrett, R.
[1984] *Brownian Motion and Martingales in Analysis.* Wadsworth.

Duoandikoetxea, J.
[1986] Multiple singular integrals and maximal functions along hypersurfaces, *Ann. Inst. Fourier* **36**: 185–206.

Duoandikoetxea, J., and J. L. Rubio de Francia
[1985] Estimations indépendantes de la dimension pour les transformées de Riesz, *C. R. Acad. Sci. Paris* **300**: 193–196.
[1986] Maximal and singular integral operators via Fourier transform estimates, *Invent. Math.* **84**: 541–561.

Duren, P. L.
[1969] Extension of a theorem of Carleson, *Bull. Amer. Math. Soc.* **75**: 143–146.
[1970] *Theory of* H^p *Spaces.* Pure and Applied Math. #38, Academic Press.

Duren, P. L., B. Romberg, and A. Shields
[1969] Linear functionals on H^p spaces with $0 < p < 1$, *J. Reine Angew. Math.* **238**: 32–60.

Dynin, A. S.
[1975] Pseudo-differential operators on the Heisenberg group, *Dokl. Akad. Nauk SSSR* **225**: 1245–1248; transl. in *Soviet Math. Dokl.* **16**: 1608–1612.

Egorov, Y. P.
[1969] On canonical transformations of pseudo-differential operators (Russian), *Uspehi Mat. Nauk.* **24**: 235–236.

Erdélyi, A.
[1956] *Asymptotic Expansions.* Dover.

Eskin, G. I.
[1970] Degenerate elliptic pseudo-differential operators of principal type (transl.), *Math. USSR Sb.* **11**: 539–582.

Fabes, E., M. Jodeit, and N. Rivière
[1978] Potential techniques for boundary-value problems on C^1 domains, *Acta Math.* **141**: 165–186.

Fabes, E. B., and N. Rivière
[1966] Singular integrals with mixed homogeneity, *Studia Math.* **27**: 19–38.

Falconer, K. J.
[1985] *The Geometry of Fractal Sets.* Cambridge Tracts in Math. #85, Cambridge Univ. Press.

Fefferman, C.
[1970] Inequalities for strongly singular convolution operators, *Acta Math.* **124**: 9–36.
[1971a] Characterizations of bounded mean oscillation, *Bull. Amer. Math. Soc.* **77**: 587–588.
[1971b] The multiplier problem for the ball, *Ann. of Math.* **94**: 330–336.
[1973a] L^p bounds for pseudo-differential operators, *Israel J. Math.* **14**: 413–417.
[1973b] A note on spherical summation multipliers, *Israel J. Math.* **15**: 44–52.

Fefferman, C., and J. J. Kohn
[1988] Hölder estimates on domains of complex dimension two and on three dimensional CR manifolds, *Adv. in Math.* **69**: 233–303.

Fefferman, C., and B. Muckenhoupt
[1974] Two non-equivalent conditions for weight functions, *Proc. Amer. Math. Soc.* **45**: 99–104.

Fefferman, C., and D. H. Phong
[1978] On positivity of pseudo-differential operators, *Proc. Nat. Acad. Sci. U.S.A.* **75**: 4673–4674.
[1983] Subelliptic eigenvalue problems, *in* Beckner *et al.* [1983], 590–606.

Fefferman, C., N. Rivière, and Y. Sagher
[1974] Interpolation between H^p spaces: The real method, *Trans. Amer. Math. Soc.* **191**: 75–81.

Fefferman, C., and A. Sanchez-Calle
[1986] Fundamental solutions for second order subelliptic operators, *Ann. of Math.* **124**: 247–272.

Fefferman, C., and E. M. Stein
[1971] Some maximal inequalities, *Amer. J. Math.* **93**: 107–115.
[1972] H^p spaces of several variables, *Acta Math.* **129**: 137–193.

Fefferman, R.
[1978] A theory of entropy in Fourier analysis, *Adv. in Math.* **30**: 171–201.
[1979] A note on singular integrals, *Proc. Amer. Math. Soc.* **74**: 266–270.
[1983] On an operator arising in the Calderón-Zygmund method of rotations and the Bramble-Hilbert lemma, *Proc. Nat. Acad. Sci. U.S.A.* **80**: 3877–3878.
[1986] Calderón-Zygmund theory for product domain H^p spaces, *Proc. Nat. Acad. Sci. U.S.A.* **83**: 840–843.
[1987] Harmonic analysis on product spaces, *Ann. of Math.* **126**: 109–130.

Fefferman, R., R. F. Gundy, M. L. Silverstein, and E. M. Stein
[1982] Inequalities for ratios of functionals of harmonic functions, *Proc. Nat. Acad. Sci. U.S.A.* **79**: 7958–7960.

Fefferman, R., C. E. Kenig, and J. Pipher
[1991] The theory of weights and the Dirichlet problem for elliptic equations, *Ann. of Math.* **134**: 65–124.

Fefferman, R., and E. M. Stein
[1982] Singular integrals on product spaces, *Adv. in Math.* **45**: 117–143.

Feller, W.
[1966] *An Introduction to Probability Theory and its Applications vol. II.* Wiley.

Folland, G. B.
[1975] Subelliptic estimates and function spaces on nilpotent Lie groups, *Ark. f. Mat.* **13**: 161–207.
[1977] Applications of analysis on nilpotent groups to partial differential equations, *Bull. Amer. Math. Soc.* **83**: 912–930.
[1989] *Harmonic Analysis in Phase Space.* Annals of Math. Study #122, Princeton Univ. Press.

Folland, G. B., and J. J. Kohn
[1972] *The Neumann Problem for the Cauchy-Riemann Complex.* Annals of Math. Study #75, Princeton Univ. Press.

Folland, G. B., and E. M. Stein
[1974] Estimates for the $\bar{\partial}_b$-complex and analysis on the Heisenberg group, *Comm. Pure Appl. Math.* **27**: 429–522.
[1982] *Hardy Spaces on Homogeneous Groups.* Math. Notes #28, Princeton Univ. Press.

Frazier, M., and B. Jawerth
[1985] Decomposition of Besov spaces, *Indiana U. Math. J.* **34**: 777–799.
[1988] The φ-transform and applications to distribution spaces, *in* Cwickel *et al.* [1988], 223–246.

Frazier, M., B. Jawerth, and G. Weiss
[1991] *Littlewood-Paley Theory and the Study of Function Spaces.* CBMS Regional Conference Ser. #79, American Math. Society.

Frazier, M., R. Torres, and G. Weiss
[1988] The boundedness of Calderón-Zygmund operators on the spaces $F_p^{\alpha,q}$, *Rev. Mat. Iberoamer.* **4**: 41–72.

Gandulfo, A., J. García-Cuerva, and M. Taibleson
[1976] Conjugate system characterization of H^1: Counterexamples for the Euclidean plane and local fields, *Bull. Amer. Math. Soc.* **82**: 83–85.

García-Cuerva, J.
[1983] An extrapolation theorem in the theory of A_p weights, *Proc. Amer. Math. Soc.* **87**: 422–426.

García-Cuerva, J., and J. L. Rubio de Francia
[1985] *Weighted Norm Inequalities and Related Topics.* North-Holland.

Garnett, J. B.
[1981] *Bounded Analytic Functions.* Pure and Applied Math. #96, Academic Press.

Garnett, J. B., and P. W. Jones
[1978] The distance in BMO to L^∞, *Ann. of Math.* **108**: 373–393.

Garsia, A. M.
[1973] *Martingale Inequalities: Seminar Notes on Recent Progress.* Benjamin.

Gatto, A. E., and C. E. Guitérrez
[1983] On weighted norm inequalities for the maximal function, *Studia Math.* **76**: 59–62.

Gaveau, B.
[1977] Principe de moindre action, propogation de la chaleur, et estimées sous elliptiques sur certains groupes nilpotents, *Acta Math.* **139**: 95–153.

Gehring, F. W.
[1973] The L^p integrability of the partial derivatives of a quasiconformal mapping, *Acta Math.* **130**: 265–277.

Geller, D.
[1977] Fourier analysis on the Heisenberg group, *Proc. Nat. Acad. Sci. U.S.A.* **74**: 1328–1331.
[1980a] Fourier analysis on the Heisenberg group I: The Schwartz space, *J. Funct. Anal.* **36**: 205–254.
[1980b] Fourier analysis on the Heisenberg group II: Local solvability and homogeneous distributions, *Comm. P.D.E.* **5**: 475–560.
[1980c] Some results in H^p theory for the Heisenberg group, *Duke Math. J.* **47**: 365–390.
[1984] Spherical harmonics, the Weyl transform, and the Fourier transform on the Heisenberg group, *Canad. J. Math.* **36**: 615–684.
[1990] *Analytic Pseudo-Differential Operators for the Heisenberg Group and Local Solvability.* Math. Notes #37, Princeton Univ. Press.

Geller, D., and E. M. Stein
[1984] Estimates for singular convolution operators on the Heisenberg group, *Math. Ann.* **267**: 1–15.

Giaquinta, M.
[1983] *Multiple Integrals in the Calculus of Variations and Nonlinear Elliptic Systems.* Annals of Math. Study #105, Princeton Univ. Press.

Gilbert, J. E.
[1979] Nikishin-Stein theory and factorization with applications, *in* Weiss and Wainger [1979], part 2, 233–267.

Gindikin, S. G.
[1964] Analysis on homogeneous domains (Russian), *Uspehi Mat. Nauk* **19**: 3–92; transl. in *Russian Math. Surveys* (1964) **19**: 1–89.

Ginibre, J., and G. Velo
[1985] Scattering theory in the energy space for a class of nonlinear Schrödinger equations, *J. Math. Pures et Appl.* **64**: 363–401.

Gohberg, I., and M. Krein
[1970] *Theory and Applications of Volterra Operators on Hilbert Space.* Transl. of Math. Monographs #24, American Math. Society.

Goldberg, D.
[1979] A local version of real Hardy spaces, *Duke Math. J.* **46**: 27–42.

Goodman, R. W.
[1976] *Nilpotent Lie Groups.* Lecture Notes in Math. #562, Springer-Verlag.

Greenleaf, A.
[1981] Principal curvature in harmonic analysis, *Indiana U. Math. J.* **30**: 519–537.

Greenleaf, A., and G. Uhlmann
[1990] Estimates for singular Radon transforms and pseudo-differential operators with singular symbols, *J. Funct. Anal.* **89**: 202–232.

Greiner, P. C., J. J. Kohn, and E. M. Stein
[1975] Necessary and sufficient conditions for solvability of the Lewy equation, *Proc. Nat. Acad. Sci. U.S.A.* **72**: 3287–3289.

Greiner, P. C., and E. M. Stein
[1977] *Estimates for the $\bar{\partial}$-Neumann Problem.* Math. Notes #19, Princeton Univ. Press.

Gromov, M.
[1981] Groups of polynomial growth and expanding maps, *Publ. Math. I.H.E.S.* **53**: 53–73.

Gross, L.
[1975] Logarithmic Sobolev inequalities, *Amer. J. Math.* **97**: 1061–1083.

Grossman, A., G. Loupias, and E. M. Stein
[1969] An algebra of pseudo-differential operators and quantum mechanics in phase space, *Ann. Inst. Fourier* **18**: 343–368.

Guillemin, V., and S. Sternberg
[1977] *Geometric Asymptotics.* Math. Surveys and Monographs #14, American Math. Society.

Guitérrez, J. A.
[1982] *On the Boundedness of the Banach Space-Valued Hilbert Transform.* Ph.D. Dissertation, University of Texas, Austin.

Guivarch, Y.
[1973] Croissance polynomiale et périodes des fonctions harmoniques, *Bull. Soc. Math. de France* **101**: 333–379.

Gundy, R. F., and E. M. Stein
[1979] H^p theory for the polydisc, *Proc. Nat. Acad. Sci. U.S.A.* **76**: 1026–1029.

Gundy, R. F., and N. Varopoulos
[1979] Les transformations de Reisz et les integrales stochastiques, *C. R. Acad. Sci. Paris* **289**: 13–16.

Haimo, D. T. (ed.)
[1968] *Orthogonal Expansions and their Continuous Analogues (1967 Proceedings).* Southern Illinois Univ. Press.

Hardy, G. H.
[1913] Oscillatory Dirichlet's integrals, *Q. J. Pure and Applied Math.* **44**: 242–263.

Hardy, G. H., and J. E. Littlewood
[1928] Some properties of fractional integrals, *Mat. Zeit.* **27**: 565–606.
[1930] A maximal theorem with function-theoretic applications, *Acta Math.* **54**: 81–116.
[1932] Some properties of conjugate functions, *J. Reine Angew. Math.* **167**: 405–423.

Harvey, F., and J. Polking
[1984] The $\bar{\partial}$-Neumann solution to the inhomogeneous Cauchy-Riemann equations in the unit ball of $\mathbf{C}^n$, *Trans. Amer. Math. Soc.* **281**: 587–613.

Hausdorff, F.
[1937] *Set Theory.* Transl. from German 3rd edition, Chelsea (1962).

Hausner, M., and J. Schwartz
[1968] *Lie Groups and Lie Algebras.* Gordon and Breach.

Hebisch, W., and A. Sikora
[1990] A smooth subadditive homogeneous norm on a homogeneous group, *Studia Math.* **96**: 232–236.

Hedberg, L.
[1972] On certain convolution inequalities, *Proc. Amer. Math. Soc.* **36**: 505–510.

Helffer, B., and J. Nourrigat
[1985] *Hypoellipticité Maximale des Opérateurs Polynomes de Champs de Vecteurs.* Birkhäuser.

Helgason, S.
[1962] *Differential Geometry and Symmetric Spaces.* Pure and Applied Math. #12, Academic Press.
[1984] *Groups and Geometric Analysis: Integral Geometry, Invariant Differential Operators, and Spherical Functions.* Pure and Applied Math. #113, Academic Press.

Helson, H., and G. Szegö
[1960] A problem in prediction theory, *Ann. Mat. Pura Appl.* **51**: 107–138.

Herz, C. S.

[1954] On the mean inversion of Fourier and Hankel transforms, *Proc. Nat. Acad. Sci. U.S.A.* **40**: 996–999.

[1962a] Fourier transforms related to convex sets, *Ann. of Math.* **75**: 81–92.

[1962b] On the number of lattice points in a convex set, *Amer. J. Math.* **84**: 126–133.

[1974a] H^p spaces of martingales, $0 < p \leq 1$, *Zeit f. War.* **28**: 189–205.

[1974b] Bounded mean oscillation and regulated martingales, *Trans. Amer. Math. Soc.* **193**: 199–216.

Hirschman, I. I.

[1956] Multiplier transformations I, *Duke Math. J.* **26**: 222–242.

Hlawka, E.

[1950] Über Integrale auf konvexen Körper I, *Monatsh Math.* **54**: 1–36.

Hörmander, L.

[1958] On the division of distributions by polynomials, *Ark. f. Mat.* **3**: 555–568.

[1965] Pseudo-differential operators, *Comm. Pure Appl. Math.* **18**: 501–507.

[1966a] *Introduction to Complex Analysis in Several Variables.* Van Nostrand.

[1966b] Pseudodifferential operators and non-elliptic boundary problems, *Ann. of Math.* **83**: 129–209.

[1967a] Hypoelliptic second-order differential equations, *Acta Math.* **119**: 147–171.

[1967b] Pseudo-differential operators and hypoelliptic equations, *in* Calderón [1967], 138–183.

[1968] The spectral function of an elliptic operator, *Acta Math.* **121**: 193–218.

[1971] Fourier integral operators, *Acta Math.* **127**: 79–183.

[1973] Oscillatory integrals and multipliers on FL^p, *Ark. f. Mat.* **11**: 1–11.

[1979] The Weyl calculus of pseudodifferential operators, *Comm. Pure Appl. Math.* **32**: 355–443.

[1983] *The Analysis of Linear Partial Differential Operators I.* Grund. der math. Wiss. #256, Springer-Verlag.

[1985] *The Analysis of Linear Partial Differential Operators III-IV.* Grund. der math. Wiss. #274–275, Springer-Verlag.

Howe, R.

[1980a] On the role of the Heisenberg group in harmonic analysis, *Bull. Amer. Math. Soc.* **3**: 821–843.

[1980b] Quantum mechanics and partial differential equations, *J. Funct. Anal.* **38**: 188–254.

[1988] The oscillator semigroup, *in* Wells [1988], 61–132.

[1989] Transcending classical invariant theory, *J. Amer. Math. Soc.* **3**: 535–552.

Hruščev, S. V.

[1984] A description of weights satisfying the A_∞ condition of Muckenhoupt, *Proc. Amer. Math. Soc.* **90**: 253–257.

Hua, L. K.

[1963] *Harmonic Analysis of Functions of Several Complex Variables in the Classical Domains.* Transl. of Math. Monographs #6, American Math. Society.

Hulanicki, A.

[1976] The distribution of energy in the Brownian motion in the Gaussian field and analytic hypoellipticity of certain subelliptic operators on the Heisenberg group, *Studia Math.* **56**: 165–173.

Hunt, R. A.

[1968] On the convergence of Fourier series, *in* Haimo [1968], 235–255.

Hunt, R. A., D. S. Kurtz, and C. F. Neugebauer

[1983] A note on the equivalence of A_p and Sawyer's condition for equal weights, *in* Beckner *et al.* [1983], 156–158.

Hunt, R. A., B. Muckenhoupt, and R. L. Wheeden

[1973] Weighted norm inequalities for the conjugate function and Hilbert transform, *Trans. Amer. Math. Soc.* **176**: 227–251.

Igari, S.

[1963] An extension of the interpolation theorem of Marcinkiewicz II, *Tôhoku Math. J.* **15**: 343–358.

[1981] Decomposition theorem and lacunary convergence of Bochner-Riesz means of Fourier transforms in two variables, *Tôhoku Math. J.* **33**: 413–419.

Jacobson, N.

[1962] *Lie Algebras.* Interscience.

Janson, S.

[1981] Characterizations of H^1 by singular integral transforms on martingales and $\mathbf{R}^n$, *Math. Scand.* **46**: 140–152.

Janson, S., and P. W. Jones

[1982] Interpolation between H^p spaces: The complex method, *J. Funct. Anal.* **48**: 58–80 .

Jawerth, B.

[1986] Weighted norm inequalities: linearization, localization, and factorization, *Amer. J. Math.* **108**: 361–414.

Jerison, D., and C. E. Kenig

[1981] The Dirichlet problem in non-smooth domains, *Ann. of Math.* **113**: 367–382.

[1982] Boundary-value problems on Lipschitz domains, *in* Littman [1982], 2–68.

[1985] Unique continuation and absence of positive eigenvalues for Schrödinger operators, *Ann. of Math.* **121**: 463–494.

Jessen, B., J. Marcinkiewicz and A. Zygmund

[1935] Note on the differentiability of multiple integrals, *Fund. Math.* **25**: 217–234.

John, F.

[1964] Quasi-isometric mappings, in *Seminar 1962–1963*, 462–473. Instituto Nazionali di Alta Matematica.

John, F., and L. Nirenberg

[1961] On functions of bounded mean oscillation, *Comm. Pure Appl. Math.* **14**: 415–426.

Johnson, R.

[1982] Application of Carleson measures to partial differential equations and Fourier multiplier problems, *in* Mauceri, Ricci, and Weiss [1983], 16–72.

Jones, B. F.

[1964] A class of singular integrals, *Amer. J. Math.* **86**: 441–462.

Jones, P. W.

[1980a] Factorization of A_p weights, *Ann. of Math.* **111**: 511–530.

[1980b] Carleson measures and the Fefferman-Stein decomposition of $\mathrm{BMO}(\mathbf{R}^n)$, *Ann. of Math.* **111**: 197–208.

[1981] Interpolation between Hardy spaces, *in* Beckner *et al.* [1983], 437–451.

[1983] Homeomorphisms of the real line which preserve BMO, *Ark. f. Mat.* **21**: 229–231.

[1984] *Personal communication.*

Journé, J.-L.

[1983] *Calderón-Zygmund Operators, Pseudo-Differential Operators and the Cauchy Integral of Calderón.* Lecture Notes in Math. #994, Springer-Verlag.

[1985] Calderón-Zygmund operators on product spaces, *Rev. Mat. Iberoamer.* **1**: 55–91.

[1986] A covering lemma for product spaces, *Proc. Amer. Math. Soc.* **96**: 593–598.

[1989] *Personal communication.*

Jurkat, W. B., and G. Sampson

[1979] The L^p mapping problem for well behaved convolutions, *Studia Math.* **65**: 227–238.

Kaneko, M., and G. Sunouchi

[1985] On the Littlewood-Paley and Marcinkiewicz functions in higher dimensions, *Tôhoku Math. J.* **37**: 343–365.

Kapitanskii, L. V.

[1990] Some generalizations of the Strichartz-Brenner inequality, *Leningrad Math. J.* **1**: 693–726.

Kaplan, A.

[1980] Fundamental solutions for a class of hypoelliptic partial differential equations generated by composition of quadratic forms, *Trans. Amer. Math. Soc.* **258**: 147–153.

Kaplan, A., and F. Ricci

[1983] Harmonic analysis on groups of Heisenberg type, *in* Mauceri, Ricci, and Weiss [1983], 416–435.

Kashiwara, M., and M. Vergne

[1978] On the Segal-Shale-Weil representations and harmonic polynomials, *Invent. Math.* **44**: 1–47.

Katznelson, Y.

[1968] *An Introduction to Harmonic Analysis.* Wiley.

Kenig, C. E.

[1985] Recent progress on boundary-value problems on Lipschitz domains, *in* Treves [1985], 175–205.

[1986] Elliptic boundary value problems on Lipschitz domains, *in* Stein [1986a], 131–183.

Kenig, C. E., G. Ponce, and L. Vega

[1991] Oscillatory integrals and regularity of dispersive equations, *Indiana U. Math. J.* **40**: 33–69.

Kenig, C. E., A. Ruiz, and C. D. Sogge

[1987] Uniform Sobolev inequalities and unique continuation for second order constant coefficient differential equations, *Duke Math. J.* **55**: 329–349.

Knapp, A. W.

[1986] *Representation Theory of Semisimple Groups: An Overview Based on Examples.* Princeton Univ. Press.

Knapp, A. W., and E. M. Stein

[1971] Intertwining operators for semi-simple groups, *Ann. of Math.* **93**: 489–578.

Kohn, J. J., and L. Nirenberg

[1965] An algebra of pseudo-differential operators, *Comm. Pure Appl. Math.* **18**: 269–305.

Kolmogorov, A. N., and G. Seliverstov

[1925] Sur la convergence des séries de Fourier, *C. R. Acad. Sci. Paris* **178**: 303–305.

Komatsu, H. (ed.)

[1973] *Hyperfunctions and Pseudo-Differential Equations.* Lecture Notes in Math. #287, Springer-Verlag.

Korányi, A.

[1969] Harmonic functions on Hermitian hyperbolic space, *Trans. Amer. Math. Soc.* **135**: 507–516.

[1972] Harmonic functions on symmetric spaces, *in* Boothby and Weiss [1972], 379–412.

Korányi, A., and S. Vági

[1971] Singular integrals in homogeneous spaces and some problems of classical analysis, *Ann. Scuola Norm. Sup. Pisa* **25**: 575–648.

Krantz, S. G.

[1982a] Fractional integration on Hardy spaces, *Studia Math.* **63**: 87–94.

[1982b] *Function Theory of Several Complex Variables.* Wiley.

Kree, P.

[1965] Distributions quasi-homogènes, *C. R. Acad. Sci. Paris* **261**: 2560–2563.

[1966] Sur les multiplicateurs dans $\mathcal{F}L^p$, *Ann. Inst. Fourier* **16**: 31–89.

Kurtz, D. S.
[1980] Littlewood-Paley and multiplier theorems on weighted L^p spaces, *Trans. Amer. Math. Soc.* **259**: 235–254.

Landau, E.
[1927] *Vorlesungen über Zahlentheorie (vol. II).* Hirzel (Leipzig).

Latter, R. H.
[1978] A decomposition of $H^p(\mathbf{R}^n)$ in terms of atoms, *Studia Math.* **62**: 92–101.

Lax, P.
[1957] Asymptotic solutions of oscillatory initial value problems, *Duke Math. J.* **24**: 627–646.

Lemarié, P.
[1985] Continuité sur les espaces de Besov des opérateurs definis par des intégrales singulières, *Ann. Inst. Fourier* **35**: 175–187.

Lemarié, P., and Y. Meyer
[1986] Ondelettes et bases hilbertiennes, *Rev. Mat. Iberoamer.* **2**: 1–18.

Lewy, H.
[1956] On the local character of the solutions of an atypical linear differential equation in three variables and a related theorem for regular functions of two complex variables, *Ann. of Math.* **64**: 514–522.
[1957] An example of a smooth linear partial differential equation without solutions, *Ann. of Math.* **66**: 155–158.

Lindenstrauss, J., and V. D. Milman (eds.)
[1987] *Geometrical Aspects of Functional Analysis.* Lecture Notes in Math. #1267, Springer-Verlag.

Littlewood, J. E.
[1931] On functions subharmonic in a circle III, *Proc. London Math. Soc.* **32**: 222–234.

Littman, W.
[1963] Fourier transforms of surface-carried measures and differentiability of surface averages, *Bull. Amer. Math. Soc.* **69**: 766–770.
[1973] (L^p, L^q) estimates for singular integral operators, *in* Spencer [1973], 479–481.
[1982] *Studies in Partial Differential Equations (ed.).* MAA Study #23, Math. Assoc. of America.

Lojasiewicz, S.
[1959] Sur le problème de division, *Studia Math.* **18**: 87–136.

Macías, R. A., and C. Segovia
[1979a] Lipschitz functions on spaces of homogeneous type, *Adv. in Math.* **33**: 257–270.
[1979b] A decomposition into atoms of distributions on spaces of homogeneous type, *Adv. in Math.* **33**: 271–309.

Mackey, G.
[1965] Some remarks on symplectic automorphisms, *Proc. Amer. Math. Soc.* **16**: 393–397.

Malgrange, B.
[1974] Intégrales asymptotiques et monodromie, *Ann. École Norm. Sup.* **7**: 405–430.

Marcinkiewicz, J., and A. Zygmund
[1939a] Quelques inégalités pour les opérations linéaries, *Fund. Math.* **32**: 115–121.
[1939b] On the summability of double Fourier series, *Fund. Math.* **32**: 122–132.

Marshall, B.
[1981] Mixed norm estimates for the Klein-Gordon equation, *in* Beckner *et al.* [1983], 638–639.

Marshall, B., W. Strauss, and S. Wainger
[1980] $L^p - L^q$ estimates for the Klein-Gordon equation, *J. Math. Pures et Appl.* **59**: 417–440.

Marstrand, J. M.
[1979] Packing planes in $\mathbf{R}^3$, *Mathematika* **26**: 180–183.

Maslov, V. P.
[1965] *Theory of Perturbations and Asymptotic Methods (Russian).* Moscow State University.

Mauceri, G., M. A. Picardello, and F. Ricci
[1981] Twisted convolutions, Hardy spaces, and Hörmander multipliers, *Rend. Circ. Mat. Palermo suppl.* **1**: 191–202.

Mauceri, G., F. Ricci, and G. Weiss (eds.)
[1983] *Harmonic Analysis (1982 Proceedings).* Lecture Notes in Math. #992, Springer-Verlag.

Maurey, B.
[1974] *Théorèmes de factorisation pour les opérateurs linéaries à valeurs dans les espaces L^p.* Astérisque #11, Société Math. de France.
[1980] Isomorphismes entre espaces H^1, *Acta Math.* **145**: 79–120.

McConnell, T. R.
[1984] On Fourier multiplier transformations of Banach-space valued functions, *Trans. Amer. Math. Soc.* **285**: 739–757.

McNeal, J.
[1989] Boundary behavior of the Bergman kernel, *Duke Math. J.* **58**: 499–512.

Métivier, G.
[1980] Hypoellipticité analytique sur des groupes nilpotents, *Duke Math. J.* **47**: 195–221.

Merryfield, K.
[1985] On the area integral, Carleson measures and H^p in the polydisc, *Indiana U. Math. J.* **34**: 663–686.

Meyer, P. A.
[1976] Demonstration probabiliste des certains inégalités de Littlewood-Paley, in *Séminaire de Probabilités X* (P. A. Meyer, ed.), 125–183. Lecture Notes in Math. #511, Springer-Verlag.

Meyer, Y.
[1980] Régularité des solutions des équations aux dérivées partielles non linéaires, in *Seminaire Bourbaki 1979–80*, 293–302. Lecture Notes in Math. #842, Springer-Verlag.
[1990] *Ondelettes et Opérateurs (vol. I and II).* Hermann.

Meyer, Y., and R. R. Coifman
[1991] *Ondelettes et Opérateurs (vol. III).* Hermann.

Mihlin, S. G.
[1948] Singular integral equations, *Uspehi Mat. Nauk.* **3**: 29–112; transl. in *Amer. Math. Soc. Transl. Ser. 1 (1962)* **10**: 84–198.

Miller, W.
[1968] *Lie Theory and Special Functions.* Academic Press.

Milnor, J.
[1963] *Morse Theory.* Annals of Math. Study #51, Princeton Univ. Press.

Miyachi, A.
[1980a] On some Fourier multipliers for $H^p(\mathbf{R}^n)$, *J. Fac. Sci. U. Tokyo* **27**: 157–179.
[1980b] On some estimates for the wave equation in L^p and H^p, *J. Fac. Sci. U. Tokyo* **27**: 331–354.
[1990] H^p spaces over open subsets of $\mathbf{R}^n$ in terms of atoms, *Studia Math.* **44**: 205–228.

Mockenhaupt, G.
[1989] *Singuläre Integrale vom Bochner-Riesz Typ.* Thesis, Universität Siegen.

Mockenhaupt, G., A. Seeger, and C. D. Sogge
[1992a] Wave front sets, local smoothing and Bourgain's circular maximal theorem, *Ann. of Math.* **136**: 207–218.
[1992b] Local smoothing of Fourier integral operators and Carleson-Sjölin estimates, *to appear in J. Amer. Math. Soc.*

Morse, A. P.
[1947] Perfect blankets, *Trans. Amer. Math. Soc.* **6**: 418–442.

Muckenhoupt, B.
[1960] On certain singular integrals, *Pacific J. Math.* **10**: 239–261.
[1972] Weighted norm inequalities for the Hardy maximal function, *Trans. Amer. Math. Soc.* **165**: 207–226.
[1974] The equivalence of two conditions for weight functions, *Studia Math.* **49**: 101–106.

Muckenhoupt, B., and R. Wheeden
[1976] Weighted bounded mean oscillation and the Hilbert transform, *Studia Math.* **54**: 221–237.

Müller, D.
[1983] Calderón-Zygmund kernels carried by linear subspaces of homogeneous nilpotent Lie algebras, *Invent. Math.* **73**: 467–489.
[1985] Singular kernels supported on homogeneous submanifolds, *J. Reine Angew. Math.* **356**: 90–118.

Müller, D. (*continued*)
[1989] On Riesz means of eigenfunction expansions for the Kohn Laplacian, *J. Reine Angew. Math.* **401**: 113–121.
[1990a] A geometric bound for maximal functions associated to convex bodies, *Pacific J. Math.* **142**: 297–312.
[1990b] A restriction theorem for the Heisenberg group, *Ann. of Math.* **131**: 567–587.

Müller, D., and F. Ricci
[1990] Analysis of second-order differential operators on Heisenberg groups I, *Invent. Math.* **101**: 545–582.

Müller, S.
[1990] Higher integrability of determinants and weak convergence in L^1, *J. Reine Angew. Math.* **412**: 20–34.

Murai, T.
[1988] *A Real-Variable Method for the Cauchy Transform and Analytic Capacity.* Lecture Notes in Math. #1307, Springer-Verlag.

Murai, T., and A. Uchiyama
[1986] Good-λ inequalities for the area integral and the non-tangential maximal function, *Studia Math.* **83**: 251–262.

Nagel, A. (ed.)
[1992] *Proceedings of the Madison Symposium on Complex Analysis.* Contemporary Math. #137, American Math. Society.

Nagel, A., F. Ricci, and E. M. Stein
[1990] Harmonic analysis and fundamental solutions on nilpotent Lie groups, *in* Sadosky [1990], 249–275.

Nagel, A., N. Rivière, and S. Wainger
[1974] On Hilbert transforms along curves, *Bull. Amer. Math. Soc.* **80**: 106–108.
[1976] A maximal function associated to the curve (t, t^2), *Proc. Nat. Acad. Sci. U.S.A.* **73**: 1416–1417.

Nagel, A., J. P. Rosay, E. M. Stein, and S. Wainger
[1989] Estimates for the Bergman and Szegö kernels in $\mathbf{C}^2$, *Ann. of Math.* **129**: 113–149.

Nagel, A., W. Rudin, and J. Shapiro
[1982] Tangential boundary behavior of functions in Dirichlet-type spaces, *Ann. of Math.* **116**: 331–360.

Nagel, A., and E. M. Stein
[1979] *Lectures on Pseudo-Differential Operators: Regularity Theorems and Applications to Non-Elliptic Problems.* Math. Notes #24, Princeton Univ. Press.
[1984] On certain maximal functions and approach regions, *Adv. in Math.* **54**: 83–106.

Nagel, A., E. M. Stein, and S. Wainger

[1978] Differentiation in lacunary directions, *Proc. Nat. Acad. Sci. U.S.A.* **75**: 1060–1062.

[1979] Hilbert transforms and maximal functions related to variable curves, *in* Weiss and Wainger [1979], part 1, 95–98.

[1981] Boundary behavior of functions holomorphic in domains of finite type, *Proc. Nat. Acad. Sci. U.S.A.* **78**: 6596–6599.

[1985] Balls and metrics defined by vector fields I: Basic properties, *Acta Math.* **155**: 103–147.

Nagel, A., J. Vance, S. Wainger, and D. Weinberg

[1983] Hilbert transforms for convex curves, *Duke Math. J.* **50**: 735–744.

[1985] Maximal functions for convex curves, *Duke Math. J.* **52**: 715–722.

Nagel, A., and S. Wainger

[1977] L^2 boundedness of Hilbert transforms along surfaces and convolution operators homogeneous with respect to a multiple parameter group, *Amer. J. Math.* **99**: 761–785.

Namazi, J.

[1984] *A Singular Integral.* Ph. D. dissertation, Indiana University.

Nelson, E.

[1973] The free Markoff field, *J. Funct. Anal.* **12**: 221-227.

Nikishin, E. M.

[1970] Resonance theorems and super-linear operators, *Russian Math. Surveys* **25**: 125–187.

Nikodym, O.

[1927] Sur les ensembles accessibles, *Fund. Math.* **10**: 116–168.

Nirenberg, L.

[1968] Pseudo-differential operators, *in* Chern and Smale [1970], vol. 3, 149–167.

Nirenberg, L., and F. Treves

[1971] On local solvability of linear partial differential equations, *Comm. Pure Appl. Math.* **24**: 279–288.

Oberlin, D. M., and E. M. Stein

[1982] Mapping properties of the Radon transform, *Indiana U. Math. J.* **31**: 641–650.

Ogden, R. D., and S. Vági

[1972] Harmonic analysis and H^2-functions on Siegel domains of type II., *Proc. Nat. Acad. Sci. U.S.A.* **69**: 11–14.

Paley, R. E.

[1930] A proof of a theorem on averages, *Proc. London Math. Soc.* **31**: 289–300.

[1932] A Remarkable series of orthogonal functions I, *Proc. London Math. Soc.* **34**: 241–279.

Pan, Y.

[1991a] Uniform estimates for oscillatory integral operators, *J. Funct. Anal.* **100**: 207–220.

[1991b] Hardy spaces and oscillatory integral operators, *Rev. Mat. Iberoamer.* **7**: 55–64.

Pecher, H.

[1985] Nonlinear small data scattering for the wave and Klein-Gordon equations, *Mat. Zeit.* **185**: 261–270.

Peetre, J.

[1966] On convolution operators leaving $L^{p,\lambda}$ spaces invariant, *Ann. Mat. Pura Appl.* **72**: 295–304.

[1972] The Weyl transform and Laguerre polynomials, *Le Mathematiche (Catania)* **27**: 301–323.

[1976] *New Thoughts on Besov Spaces.* Duke Univ. Math. Ser., Duke University.

Peral, I., and J. L. Rubio de Francia

[1985] *Recent Progress in Fourier Analysis (1983 Proceedings).* Math. Study #111, North-Holland.

Peral, J.

[1980] L^p estimates for the wave equation, *J. Funct. Anal.* **36**: 114–145.

Petersen, K. E.

[1977] *Brownian Motion, Hardy spaces, and Bounded Mean Oscillation.* London Math. Society Lecture Notes #28, Cambridge Univ. Press.

Phong, D. H.

[1979] On integral representations for the Neumann operator, *Proc. Nat. Acad. Sci. U.S.A.* **76**: 1554–1558.

Phong, D. H., and E. M. Stein

[1982] Some further classes of pseudo-differential and singular integral operators arising in boundary-value problems I: Composition of operators, *Amer. J. Math.* **104**: 141–172.

[1983] Singular integrals related to the Radon transform and boundary value problems, *Proc. Nat. Acad. Sci. U.S.A.* **80**: 7697–7701.

[1986a] Hilbert integrals, singular integrals, and Radon transforms I, *Acta Math.* **157**: 99–157.

[1986b] Hilbert integrals, singular integrals, and Radon transforms II, *Invent. Math.* **86**: 75–113.

[1992] Oscillatory integrals with polynomial phases, *Invent. Math.* **110**: 39–62.

Piatetski-Shapiro, I. I.

[1961] *Geometry of Classical Domains and the Theory of Automorphic Functions.* Fizmatgiz (Moscow, in Russian; transl. Gordon and Breach, 1969).

Pipher, J.

[1986] Journé's covering lemma and its extension to higher dimensions, *Duke Math. J.* **53**: 683–690.

Prestini, E.
[1979] Restriction theorems for the Fourier transform to some manifolds in $\mathbf{R}^n$, *in* Weiss and Wainger [1979], part 1, 101–109.

Randol, B.
[1969a] On the Fourier transform of the indicator function of a planar set, *Trans. Amer. Math. Soc.* **139**: 271–278.
[1969b] On the asymptotic behaviour of the Fourier transform of the indicator function of a convex set, *Trans. Amer. Math. Soc.* **139**: 279–285.

Range, R. M.
[1986] *Holomorphic Functions and Integral Representations in Several Complex Variables.* Graduate Texts in Math. #108, Springer-Verlag.

Reimann, H. M.
[1974] Functions of bounded mean oscillation and quasi-conformal mappings, *Comm. Math. Helv.* **49**: 260–276.

Ricci, F.
[1982] Calderón-Zygmund kernels on nilpotent Lie groups, *in* Weiss and Ricci [1982], 217–227.
[1985] Commutative algebras of invariant functions on groups of Heisenberg type, *J. London Math. Soc.* **32**: 265–271.

Ricci, F., and E. M. Stein
[1987] Harmonic analysis on nilpotent groups and singular integrals I: Oscillatory integrals, *J. Funct. Anal.* **73**: 179–194.
[1988] Harmonic analysis on nilpotent groups and singular integrals II: Singular kernels supported on submanifolds, *J. Funct. Anal.* **78**: 56–84.
[1992] Multiparameter singular integrals and maximal functions, *Ann. Inst. Fourier* **42**: 637–670.

Riemann, B.
[1854] Über die Darstellbarkeit einer Funktion durch eine trigonometrische Reihe, in *Ges. Werke*, 227–271. Leipzig (1892); reprinted by Dover (1953).

Rockland, C.
[1978] Hypoellipticity for the Heisenberg group, *Trans. Amer. Math. Soc.* **240**: 1–52.

Rosenblum, M.
[1962] Summability of Fourier series in $L^p(d\mu)$, *Trans. Amer. Math. Soc.* **105**: 32–42.

Rothschild, L. P., and E. M. Stein
[1976] Hypoelliptic differential operators and nilpotent groups, *Acta Math.* **137**: 247–320.

Rothschild, L. P., and D. Tartakoff
[1982] Analytic relative fundamental solutions and projections for left invariant operators on the Heisenberg group, *Ann. École Norm. Sup.* **15**: 419–440.

Rubio de Francia, J. L.
[1980] Vector valued inequalities for operators on L^p spaces, *Bull. London Math. Soc.* **12**: 211–215.
[1981] Boundedness of maximal functions and singular integrals in weighted L^p spaces, *Proc. Amer. Math. Soc.* **83**: 673–679.
[1982] Factorization and extrapolation of weights, *Bull. Amer. Math. Soc.* **7**: 393–395.
[1984] Factorization theory and A_p weights, *Amer. J. Math.* **106**: 533–547.
[1985] A Littlewood-Paley inequality for arbitrary intervals, *Rev. Mat. Iberoamer.* **1**: 1–13.
[1986] Maximal functions and Fourier transforms, *Duke Math. J.* **53**: 395–404.

Rudin, W.
[1980] *Function Theory in the Unit Ball of* $\mathbf{C}^n$. Grund. der math. Wiss. #241, Springer-Verlag.

Ruiz, A.
[1985] On almost-everywhere convergence of solutions of hyperbolic equations to L^p initial data, *Trans. Amer. Math. Soc.* **287**: 167–188.

Runst, T.
[1985] Pseudo-differential operators of the exotic class $S^0_{1,1}$ in spaces of Besov and Triebel-Lizorkin types, *Ann. Global Anal. Geom.* **3**: 13–28.

Sadosky, C.
[1966] A note on parabolic fractional and singular integrals, *Studia Math.* **26**: 327–335.
[1990] *Analysis and Partial Differential Equations (ed.).* Lecture Notes in Pure and Applied Math. #122, Marcel Dekker.

Saks, S.
[1935] On the strong derivatives of functions of intervals, *Fund. Math.* **25**: 235–252.
[1937] *Theory of the Integral.* Hafner.

Sampson, G.
[1981] Oscillating kernels that map H^1 into L^1, *Ark. f. Mat.* **18**: 125–144.

Sanchez-Calle, A.
[1984] Fundamental solutions and geometry of the sum of squares of vector fields, *Invent. Math.* **78**: 143–160.

Sarason, D.
[1975] Functions of vanishing mean oscillation, *Trans. Amer. Math. Soc.* **207**: 391–405.

Sato, M., T. Kawai, and M. Kashiwara
[1973] Microfunctions and pseudo-differential equations, *in* Komatsu [1973], 265–529.

Sawyer, E. T.
[1982] A characterization of a two-weight norm inequality for maximal operators, *Studia Math.* **75**: 1–11.

Sawyer, S. A.
[1966] Maximal inequalities of weak type, *Ann. of Math.* **84**: 157–174.

Seeger, A., and C. D. Sogge
[1989] Bounds for eigenfunctions of differential operators, *Indiana U. Math. J.* **38**: 669–682.

Seeger, A., C. D. Sogge, and E. M. Stein
[1991] Regularity properties of Fourier integral operators, *Ann. of Math.* **134**: 231–251.

Seeley, R. T.
[1965] Refinement of the calculus at Calderón and Zygmund, *Indag. Math.* **27**: 521–531.
[1967] Elliptic singular integrals, *in* Calderón [1967], 308–315.

Segal, I. E.
[1959] Foundations of the theory of dynamical systems of infinitely many degrees of freedom I, *Mat.-Fys. Medd. Dansk. Vid. Selsk.* **31**: #12.
[1963] Transforms for operators and symplectic automorphisms over a locally compact abelian group, *Math. Scand.* **13**: 31–43.

Shale, D.
[1962] Linear symmetries of free Boson fields, *Trans. Amer. Math. Soc.* **103**: 149–167.

Siegel, C. L.
[1950] *Analytic Functions of Several Complex Variables.* Institute for Advanced Study.

Sjögren, P.
[1986] Admissible convergence of Poisson integrals in symmetric spaces, *Ann. of Math.* **124**: 313–335.

Sjölin, P.
[1971] Convergence almost everywhere of certain singular integrals and multiple Fourier series, *Ark. f. Mat.* **9**: 65–90.
[1974] Fourier multipliers and estimates for Fourier transforms of measures carried by smooth curves in $\mathbf{R}^2$, *Studia Math.* **51**: 169–182.
[1979] An H^p inequality for strongly singular integrals, *Mat. Zeit.* **165**: 231–238.
[1981] Convolution with oscillating kernels, *Indiana U. Math. J.* **30**: 47–56.
[1987] Regularity of solutions to the Schrödinger equation, *Duke Math. J.* **55**: 699–715.

Smith, K. T.
[1956] A generalization of an inequality of Hardy and Littlewood, *Canad. J. Math.* **8**: 157–170.

Sogge, C. D.
[1987a] A sharp restriction theorem for degenerate curves in $\mathbf{R}^2$, *Amer. J. Math.* **109**: 223–228.
[1987b] On the convergence of Riesz means on compact manifolds, *Ann. of Math.* **126**: 439–447.
[1987c] On the almost-everywhere convergence to L^p data for higher order hyperbolic operators, *Proc. Amer. Math. Soc.* **100**: 99–103.
[1988] Concerning the L^p norm of spectral clusters for second order elliptic operators on compact manifolds, *J. Funct. Anal.* **77**: 123–134.

Sogge, C. D. (*continued*)
[1989] Oscillatory integrals and unique continuation for second order elliptic differential equations, *J. Amer. Math. Soc.* **2**: 491–516.
[1991] Propagation of singularities and maximal functions in the plane, *Invent. Math.* **104**: 349–376.
[1993] *Fourier Integrals in Classical Analysis.* Cambridge Tracts in Math. #105, Cambridge Univ. Press.

Sogge, C. D., and E. M. Stein
[1985] Averages of functions over hypersurfaces in $\mathbf{R}^n$, *Invent. Math.* **82**: 543–556.
[1986] Averages over hypersurfaces II, *Invent. Math.* **86**: 233–242.
[1990] Averages over hypersurfaces: Smoothness of generalized Radon transforms, *J. Analyse Math.* **54**: 165–188.

Spanne, S.
[1966] Sur l'interpolation entre les espaces $\mathcal{L}_k^{p,\Phi}$, *Ann. Scuola Norm. Sup. Pisa* **20**: 625–648.

Spencer, D. C. (ed.)
[1973] *Partial Differential Equations.* Proc. Symp. Pure Math. #23, American Math. Society.

Stanton, N. K.
[1981] The solution of the $\bar{\partial}$-Neumann problem in a strictly pseudoconvex Siegel domain, *Invent. Math.* **65**: 137–174.

Stegenga, D. A.
[1976] Bounded Toeplitz operators on H^1 and applications of the duality between H^1 and the functions of bounded mean oscillation, *Amer. J. Math.* **98**: 573–589.

Stein, E. M.
[1958] On the functions of Littlewood-Paley, Lusin, and Marcinkiewicz, *Trans. Amer. Math. Soc.* **88**: 430–466.
[1961] On limits of sequences of operators, *Ann. of Math.* **74**: 140–170.
[1967] Singular integrals, harmonic functions, and differentiability properties of functions of several variables, *in* Calderón [1967], 316–335.
[1968] The analogues of Fatou's theorem and estimates for maximal functions, in *Geometry of Homogeneous Bounded Domains (C.I.M.E.* 3^{rd} *cycle)*, 291–307. Cremonese (Rome).
[1970a] *Topics in Harmonic Analysis Related to the Littlewood-Paley Theory.* Annals of Math. Study #63, Princeton Univ. Press.
[1970b] Some problems in harmonic analysis suggested by symmetric spaces and semisimple groups, in *Proc. Int'l. Congress of Math. Nice I*, 173–189. Gauthier-Villars.
[1970c] *Singular Integrals and Differentiability Properties of Functions.* Princeton Univ. Press.
[1971] L^p boundedness of certain convolution operators, *Bull. Amer. Math. Soc.* **77**: 404–405.
[1972] *Boundary Behavior of Holomorphic Functions of Several Complex Variables.* Math. Notes #11, Princeton Univ. Press.

Stein, E. M. (*continued*)

[1973a] Singular integrals and estimates for the Cauchy-Riemann equations, *Bull. Amer. Math. Soc.* **79**: 440–445.

[1973b] Pseudo-Differential Operators, *notes by D. H. Phong of a course given at Princeton University, 1972–1973.*

[1976a] Maximal functions: Spherical means, *Proc. Nat. Acad. Sci. U.S.A.* **73**: 2174–2175.

[1976b] Maximal functions: Homogeneous curves, *Proc. Nat. Acad. Sci. U.S.A.* **73**: 2176–2177.

[1976c] Maximal functions: Poisson integrals on symmetric spaces, *Proc. Nat. Acad. Sci. U.S.A.* **73**: 2547–2549.

[1979] Some problems in harmonic analysis, *in* Weiss and Wainger [1979], part 1, 3–20.

[1981] The differentiability of functions in $\mathbf{R}^n$, *Ann. of Math.* **113**: 383–385.

[1982] The development of square functions in the work of A. Zygmund, *Bull. Amer. Math. Soc.* **7**: 359–376.

[1983a] An H^1 function with non-summable Fourier expansion, *in* Mauceri, Ricci, and Weiss [1983], 193–200.

[1983b] Boundary behavior of harmonic functions on symmetric spaces: Maximal estimates for Poisson integrals, *Invent. Math.* **74**: 63–83.

[1985] Three variations on the theme of maximal functions, *in* I. Peral and J. L. Rubio de Francia [1985], 49–56.

[1986] Oscillatory integrals in Fourier analysis, *in* Stein [1986a], 307–355.

[1986a] *Beijing Lectures in Harmonic Analysis (ed.).* Annals of Math. Study #112, Princeton Univ. Press.

Stein, E. M., and J. Stein

[1991] Stock price distributions with stochastic volatility: An analytic approach, *Review of Financial Studies* **4**: 727–752.

Stein, E. M., and J.-O. Strömberg

[1983] Behavior of maximal functions in $\mathbf{R}^n$ for large n, *Ark. f. Mat.* **21**: 259–269.

Stein, E. M., M. H. Taibleson, and G. Weiss

[1981] Weak-type estimates for maximal operators on certain H^p spaces, *Rend. Circ. Mat. Palermo suppl.* **1**: 81–97.

Stein, E. M., and S. Wainger

[1970] The estimation of an integral arising in multiplier transformations, *Studia Math.* **35**: 101–104.

[1978] Problems in harmonic analysis related to curvature, *Bull. Amer. Math. Soc.* **84**: 1239–1295.

[1990] Discrete analogues of singular Radon transforms, *Bull. Amer. Math. Soc.* **23**: 537–544.

Stein, E. M., and G. Weiss

[1957] On the interpolation of analytic families of operators acting on H^p spaces, *Tôhoku Math. J.* **9**: 318–339.

[1960] On the theory of harmonic functions of several variables, I: The theory of H^p spaces, *Acta Math.* **103**: 25–62.

Stein, E. M., and G. Weiss (*continued*)

[1968] Generalization of the Cauchy-Riemann equations and representations of the rotation group, *Amer. J. Math.* **90**: 163–196.

[1971] *Introduction to Fourier Analysis on Euclidean Spaces.* Princeton Univ. Press.

Stein, E. M., and N. J. Weiss

[1969] On the convergence of Poisson integrals, *Trans. Amer. Math. Soc.* **140**: 34–54.

Stein, E. M., and A. Zygmund

[1967] Boundness of translation-invariant operators on Hölder spaces and L^p spaces, *Ann. of Math.* **85**: 337–349.

Strichartz, R. S.

[1970a] Convolutions with kernels having singularities on a sphere, *Trans. Amer. Math. Soc.* **148**: 461–471.

[1970b] *A priori* estimates for the wave equation and some applications, *J. Funct. Anal.* **5**: 218–235.

[1977] Restrictions of Fourier transforms to quadratic surfaces and decay of solutions of wave equations, *Duke Math. J.* **44**: 705–713.

[1982] Singular integrals supported on submanifolds, *Studia Math.* **74**: 137–151.

[1991] L^p harmonic analysis and Radon transforms on the Heisenberg group, *J. Funct. Anal.* **96**: 350–406.

Strömberg, J.-O.

[1976] Weak estimates for maximal functions with rectangles in certain directions, *Ark. f. Mat.* **15**: 229–240.

[1978] Maximal functions associated to rectangles with uniformly distributed directions, *Ann. of Math.* **107**: 399–402.

[1979a] Bounded mean oscillation with Orlicz norms and duality of Hardy spaces, *Indiana U. Math. J.* **28**: 511–544.

[1979b] Non-equivalence between two kinds of conditions on weight functions, *in* Weiss and Wainger [1979], part 1, 141–148.

[1981] A modified Franklin system and higher-order spline systems on $\mathbf{R}^n$ as unconditional bases for Hardy spaces, *in* Beckner *et al.* [1983], 475–493.

Strömberg, J.-O., and A. Torchinsky

[1989] *Weighted Hardy Spaces.* Lecture Notes in Math. #1381, Springer-Verlag.

Stroock, D. W., and S. R. S. Varadhan

[1974] A probabilistic approach to $H^p(R^d)$, *Trans. Amer. Math. Soc.* **192**: 245–260.

Sueiro, J.

[1986] On maximal functions and Poisson-Szegö integrals, *Trans. Amer. Math. Soc.* **298**: 653–669.

Svensson, I.

[1971] Estimates for the Fourier transform of the characteristic function of a convex set, *Ark. f. Mat.* **9**: 11–22.

Taberner, B. B.
[1985] On the restriction of the Fourier transform to a conical surface, *Trans. Amer. Math. Soc.* **292**: 321–332.

Taibleson, M. H.
[1964] On the theory of Lipschitz spaces of distributions on Euclidean n-space I, *J. Math. and Mech.* **13**: 407–480.
[1965] On the theory of Lipschitz spaces of distributions on Euclidean n-space II, *J. Math. and Mech.* **14**: 821–840.
[1966] On the theory of Lipschitz spaces of distributions on Euclidean n-space III, *J. Math. and Mech.* **15**: 973–981.
[1975] *Fourier Analysis on Local Fields.* Math. Notes #15, Princeton Univ. Press.

Taibleson, M. H., and G. Weiss
[1980] *The Molecular Characterization of Certain Hardy Spaces.* Astérisque #77 (67–149), Société Math. de France.

Taylor, M. E.
[1981] *Pseudodifferential Operators.* Princeton Univ. Press.
[1984] *Noncommutative Microlocal Analysis, part I.* Memoirs of the AMS #313, American Math. Society.

Thangavelu, S.
[1993] *Lectures on Hermite and Laguerre Expansions.* Math. Notes #42, Princeton Univ. Press.

Titchmarsh, E. C.
[1951] *The Theory of the Riemann Zeta Function.* Oxford at the Clarendon Press.

Tits, J.
[1981] Groupes á croissance polynomiale, in *Seminaire Bourbaki 1980–81*, Exposé #572, 176–188. Lecture Notes in Math. #901, Springer-Verlag.

Tomas, P. A.
[1975] A restriction theorem for the Fourier transform, *Bull. Amer. Math. Soc.* **81**: 477–478.
[1979] Restriction theorems for the Fourier transform, *in* Weiss and Wainger [1979], part 1, 111–114.

Torchinsky, A.
[1986] *Real-Variable Methods in Harmonic Analysis.* Pure and Applied Math. #123, Academic Press.

Torres, R.
[1992] *On the Boundedness of Certain Operators with Singular Kernels on Distribution Spaces.* Memoirs of the AMS #442, American Math. Society.

Treves, F.
[1982] *Introduction to Pseudo-Differential and Fourier Integral Operators (vols. 1 and 2).* Plenum Press.
[1985] *Pseudodifferential Operators and Applications (ed.).* Proc. Symp. Pure Math. #43, American Math. Society.

Triebel, H.
[1983] *Theory of Function Spaces.* Monographs in Math. #78, Birkhäuser.

Uchiyama, A.
[1980] A maximal function characterization of H^p on a space of homogeneous type, *Trans. Amer. Math. Soc.* **262**: 579–592.
[1982] A constructive proof of the Fefferman-Stein decomposition of BMO($\mathbf{R}^n$), *Acta Math.* **148**: 215–241.
[1986] On the radial maximal function of distributions, *Pacific J. Math.* **121**: 467–483.

Unterberger, A., and J. Bokobza
[1964] Les opérateurs de Calderón-Zygmund preécisés, *C. R. Acad. Sci. Paris* **259**: 1612–1614.
[1965] Sur une généralisation des opérateurs de Calderón-Zygmund et des espaces H^s, *C. R. Acad. Sci. Paris* **260**: 3265–3267.

van Hove, L.
[1952] Sur certains représentations unitaires d'un groupe fini de transformations, *Mem. Acad. Roy. de Belgique Classe des Sci.* **26**: #6.

Varadarajan, V. S.
[1974] *Lie Groups, Lie Algebras, and Their Representations.* Prentice-Hall.

Varčenko, A. N.
[1976] Newton polyhedra and estimation of oscillating integrals, *Funct. Anal. Appl.* **10**: 175–196.

Varopoulous, N.
[1977] BMO functions and the $\bar{\partial}$-equation, *Pacific J. Math.* **71**: 221–273.

Vega, L.
[1988] Schrödinger equations: Pointwise convergence to the initial data, *Proc. Amer. Math. Soc.* **102**: 874–878.

Vilenkin, N. J.
[1968] Laguerre polynomials, Whittaker functions and the representations of groups of bordered matrices, *Mat. Sb.* **75(117)**: 432–444; transl. in *Math. USSR Sb. (1968)* **4**: 399–410.

Voros, A.
[1978] An algebra of pseudo-differential operators and the asymptotics of quantum mechanics, *J. Funct. Anal.* **29**: 104–132.

Wainger, S.
[1965] *Special Trigonometric Series in k Dimensions.* Memoirs of the AMS #59, American Math. Society.
[1979] Applications of the Fourier transform to averages over lower dimensional sets, *in* Weiss and Wainger [1979], part 1, 85–94.

Watson, G. N.
[1922] *Theory of Bessel Functions.* Cambridge Univ. Press.

Weil, A.
[1963] Sur certaines groupes d'opérateurs unitaires, *Acta Math.* **111**: 143–211.
[1965] Sur la formule de Siegel dans la théorie des groupes classiques, *Acta Math.* **113**: 1–86.

Weiss, G., and F. Ricci (eds.)
[1982] *Harmonic Analysis (1981 Proceedings)*. Lecture Notes in Math. #908, Springer-Verlag.

Weiss, G., and S. Wainger (eds.)
[1979] *Harmonic Analysis in Euclidean Spaces (2 volumes)*. Proc. Symp. Pure Math. #35, American Math. Society.

Weiss, N. J.
[1968] An isometry of H^p spaces, *Proc. Amer. Math. Soc.* **19**: 1083–1086.
[1972] Fatou's theorem for symmetric spaces, *in* Boothby and Weiss [1972], 413–441.

Wells, R. O. (ed.)
[1988] *The Mathematical Heritage of Hermann Weyl.* Proc. Symp. Pure Math. #48, American Math. Society.

Weyl, H.
[1931] *The Theory of Groups and Quantum Mechanics.* Reprinted by Dover.

Wiener, N.
[1933] *The Fourier Integral and Certain of its Applications.* Cambridge Univ. Press.

Wilson, J. M.
[1985] On the atomic decomposition for Hardy spaces, *Pacific J. Math.* **116**: 201–207.

Wolff, T.
[1982] A note on interpolation spaces, *in* Weiss and Ricci [1982], 199–204.
[1992] Counter-examples with harmonic gradients, *to appear.*

Young, W. S.
[1982] Weighted norm inequalities for the Hardy-Littlewood maximal function, *Proc. Amer. Math. Soc.* **85**: 24–26.

Zó, F.
[1976] A note on the approximation of the identity, *Studia Math.* **55**: 111–122.

Zygmund, A.
[1959] *Trigonometric Series.* Cambridge Univ. Press.
[1969] On certain lemmas of Marcinkiewicz and Carleson, *J. Approx. Theory* **2**: 249–257.
[1974] On Fourier coefficients and transforms of functions of two variables, *Studia Math.* **50**: 189–202.

Zygmund, A., W. Transue, M. Morse, A. P. Calderón, and S. Bochner (eds.)
[1950] *Contributions to Fourier Analysis.* Annals of Math. Study #25, Princeton Univ. Press.

Author Index

Subject Index