Indian J. Pure Appl. Math., **48**(1): 1-58, March 2017

DOI: 10.1007/s13226-016-0198-z

SOME TOPICS IN THE HISTORY OF HARMONIC ANALYSIS IN THE TWENTIETH CENTURY

G. B. Folland

Department of Mathematics, University of Washington, Seattle, WA 98195-4350, *USA*
e-mail: folland@uw.edu

(*Received* 18 *June* 2016; *accepted* 5 *August* 2016)

In December 2015 I gave a series of six lectures at the Indian Institute of Science in which I sketched the thematic development of some of the main techniques and results of 20th-century harmonic analysis. The subjects of the lectures were, briefly, as follows:

1. Fourier series, 1900-1950.
2. Singular integrals (part I).
3. H^p, BMO, and singular integrals (part II).
4. Littlewood-Paley theory: the history of a technique.
5. Harmonic analysis on groups.
6. Wavelets.

I emphasized interconnections, both the way in which the material in the first lecture provided the roots out of which most of the developments in the other lectures grew, and the ways in which those developments interacted with each other. I included sketches of as many proofs as the time would permit: some very brief, but some fairly complete, especially those whose methodology is an important part of the subject. Much was omitted, of course, and there was a natural bias toward the areas where I have spent periods of my own mathematical life. Many developments, particularly those of the final quarter-century, received at most a brief mention.

This paper is a written account of these lectures with a few more details fleshed out, a few topics reorganized, and a few items added. I hope that others may find it an interesting narrative and a useful reference, and that it may lead some of them to share my enjoyment of exploring the original sources. I have tried to provide the references to those sources wherever possible, and for the more recent developments I also provide references to various expository works as the occasion arises. For the pre-1950 results discussed here and their proofs, however, there is one canonical reference, which I give here once and for all: Antoni Zygmund's treatise [96]. (The more fundamental ones can also be found in Folland [29].)

Key words : Fourier analysis; harmonic analysis; singular integral operators; Hardy spaces; Littlewood-Paley theory; wavelets

1. Some Notation and Terminology

We denote the circle group, considered either as $\mathbb{R}/2\pi\mathbb{Z}$ with coordinate θ or as $\{z \in \mathbb{C} : |z| = 1\}$ with coordinate $e^{i\theta}$, by $\mathbb{T}$; and we denote the unit disc in the plane, with polar coordinates written either in real form as (r, θ) or in complex form as $z = re^{i\theta}$, by $\mathbb{D}$. The Fourier series of an integrable function f on $\mathbb{T}$ is

$$f(\theta) \sim \sum_{-\infty}^{\infty} c_k e^{ik\theta}, \qquad c_k = \frac{1}{2\pi}\int_0^{2\pi} f(\theta)e^{-ik\theta}\,d\theta.$$

(We write $\sim$ rather than $=$ because the convergence of the series is a major issue to be discussed below.) The partial sums of this series are always taken to be the symmetric ones:

$$S_n^f(\theta) = \sum_{-n}^{n} c_k e^{ik\theta}. \tag{1}$$

Associated to every such Fourier series is the series

$$u(r, \theta) = \sum_{-\infty}^{\infty} c_k r^{|k|} e^{ik\theta} \qquad (0 \leq r < 1), \tag{2}$$

which converges uniformly on compact subsets of $\mathbb{D}$ to a harmonic function. For fixed $r < 1$, the series $u(r, \cdot)$ is called the rth *Abel mean* of the Fourier series of f.

The function u can also be expressed as the *Poisson integral* of f:

$$u(r, \theta) = \int_0^{2\pi} P(r, \theta - \phi) f(\phi)\,d\phi, \tag{3}$$

where P is the *Poisson kernel*:

$$P(r, \theta) = \sum_{-\infty}^{\infty} r^{|k|} e^{ik\theta} = \frac{1 - r^2}{1 + r^2 - 2r\cos\theta}. \tag{4}$$

We shall also need the analogue of this for functions on $\mathbb{R}^n$. If $f \in L^p(\mathbb{R}^n)$ $(1 \leq p \leq \infty)$, its *Poisson integral* is the harmonic function u on

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

defined by

$$u(t,x) = (P_t * f)(x),$$

where P_t is the *Poisson kernel*

$$P_t(x) = \frac{\Gamma((n+1)/2)}{\pi^{(n+1)/2}} \frac{t}{(|x|^2+t^2)^{(n+1)/2}} \tag{5}$$

(see, for example, [28] or [87]) and $*$ denotes *convolution*:

$$g * h(x) = \int g(x-y)h(y)\,dy = \int g(y)h(x-y)\,dy.$$

If $E \subset \mathbb{T}$ or $E \subset \mathbb{R}^n$, χ_E denotes the characteristic function of E: $\chi_E(x) = 1$ if $x \in E$, $\chi_E(x) = 0$ otherwise.

Although we follow the classical practice in taking Fourier series to be 2π-periodic, for Fourier transforms we shall prefer to put the factors of 2π in the exponents. Thus, we define the Fourier transform of a function $f \in L^1(\mathbb{R}^n)$ by

$$\widehat{f}(\xi) = \mathcal{F}f(\xi) = \int_{\mathbb{R}^n} e^{-2\pi i\xi\cdot x} f(x)\,dx.$$

The Fourier inversion formula (valid literally if $\widehat{f} \in L^1(\mathbb{R}^n)$ and suitably interpreted in the general case) is then

$$f(x) = \mathcal{F}^{-1}\widehat{f}(x) = \int_{\mathbb{R}^n} e^{2\pi i\xi\cdot x} \widehat{f}(\xi)\,d\xi.$$

We record here one specific Fourier transform that will be needed in several places: the Fourier transform of the Poisson kernel (5) is

$$\widehat{P}_t(\xi) = e^{-2\pi t|\xi|}. \tag{6}$$

2. Fourier Series in 1900: Pointwise Convergence

The year 1900 — or, better, 1902 — is a good starting point for the history of modern harmonic analysis, because the latter is the year when the Lebesgue integral [51] was born. Before this fundamental tool became available, one did not have the conceptual machinery to state, much less prove, most of the results that now form the basis of the subject. Of course, the use of trigonometric series to solve problems coming from physics and other sciences goes back to Euler and the Bernoullis in the 18th century, but before the Lebesgue revolution the development of a rigorous theory (apart from results of an elementary nature) was largely restricted to questions about pointwise convergence.

The first great advance in this direction was made by Dirichlet [16] in 1829. He proved that the Fourier series of a periodic function f that is piecewise continuous and piecewise monotone converges to $\frac{1}{2}[f(\theta+) + f(\theta-)]$ at every θ, and in particular to $f(\theta)$ at every θ at which f is continuous. Once the Riemann integral and the notion of "bounded variation" became available, it required only a slight elaboration of Dirichlet's argument to show that the hypothesis "piecewise continuous and piecewise monotone" could be replaced by "of bounded variation on $[0, 2\pi]$". This sufficient condition for pointwise convergence remains unmatched for its combination of utility and generality.

The next major paper on the subject was Riemann's *Habilitationsschrift* [69], which dates from 1854 although it was not formally published until after Riemann's death in 1866. Riemann main concern was the question of convergence of trigonometric series $\sum c_k e^{ik\theta}$ without *a priori* assumptions on the nature of the coefficients c_k. The results that he considered the main point of the paper are no longer of broad interest (though some of them are discussed in Chapter IX of Zygmund [96]). However, the paper is significant as the birthplace of the Riemann integral as a precisely defined concept, and two of its theorems are still widely used: the fact that the Fourier coefficients c_n of a Riemann integrable function tend to zero as $n \to \infty$ (which became the "Riemann-Lebesgue lemma" when generalized to Lebesgue integrable functions), and the fact that the convergence or divergence of the Fourier series of an integrable function f at a point x_0 depends only on the behavior of f in an arbitrarily small neighborhood of x_0. It also contains some notable counterexamples and a nice account of the earlier work on trigonometric series and their applications.

The other landmark of the pre-1900 theory is the discovery by du Bois-Reymond in 1873 [18] of a continuous periodic function whose Fourier series diverges at one point. (Simpler examples are now known, and a fairly easy Baire-category argument shows that the Fourier series of "most" functions in $C(\mathbb{T})$ are not everywhere convergent.)

After 1900 the emphasis shifted from pointwise convergence to other types of convergence and methods of summation, so we shall close the discussion of pointwise convergence of Fourier series by citing three later major results. First, in 1914 Sergei Bernstein [1] showed that the Fourier series of a function that is Hölder continuous of exponent $> \frac{1}{2}$ converges absolutely and hence uniformly on $\mathbb{T}$. (Such functions need not be of bounded variation on any interval.) Second, in 1923 A. N. Kolmogorov [46] constructed a function in $L^1(\mathbb{T})$ whose Fourier series diverges almost everywhere, and three years later [47] he outdid himself by producing a function in $L^1(\mathbb{T})$ whose Fourier series diverges *everywhere*. After that, the outstanding open question was whether the Fourier series of a function in $L^p(\mathbb{T})$ ($p > 1$), or even in $C(\mathbb{T})$, necessarily converges almost everywhere; an affirmative answer was given in 1966 by Lennart Carleson [6] for $p = 2$ and extended shortly afterwards to all $p > 1$ by Richard Hunt [40]; see also Fefferman [21].

3. The First Decade: 1900-1910

The first major advance in the theory of Fourier series after 1900 did not actually make use of the Lebesgue integral: it is L. Fejér's theorem [24] on summability of Fourier series. Instead of evaluating a Fourier series of a function f as the limit of its partial sums S_n^f defined by (1), we evaluate it as the limit of the *average* of the first n partial sums (the nth *Cesàro mean*),

$$\sigma_n^f(\theta) = \frac{1}{n}[S_1^f(\theta) + \cdots + S_n^f(\theta)],$$

or as the limit of rth Abel mean $u(r,\theta)$ defined by (2) as $r \to 1$.

***Theorem* 1** *(Fejer,* 1903*) — If f is Riemann integrable on $\mathbb{T}$, then $\sigma_n^f(\theta) \to f(\theta)$ as $n \to \infty$ and $u(r,\theta) \to f(\theta)$ as $r \to 1$ at every θ where f is continuous. If f is continuous everywhere, then $\sigma_n^f \to f$ and $u(r,\cdot) \to f$ uniformly.*

Fejér proved the results about σ_n^f much as we do today, by writing σ_n^f as the convolution of f with the so-called Fejér kernel and examining the properties of the latter, then deduced the results about u by invoking the known connections between Cesàro and Abel means.

Not long afterward, Fatou [19] investigated the boundary behavior of u in the light of the new Lebesgue integral and established a cluster of interesting results. By combining one of them with Lebesgue's version of the fundamental theorem of calculus, one obtains the following remarkable improvement on Fejér's pointwise convergence theorem, involving the important notion of nontangential convergence. Namely, if $g(r,\theta)$ is a function on the unit disc [resp., $g(t,x)$ is a function on $\mathbb{R}_+^{n+1}$], and $h(\theta)$ is a function on $\mathbb{T}$ [resp., $h(x)$ is a function on $\mathbb{R}^n$], we say that $g(r,\theta) \to h(\theta_0)$ *nontangentially* [resp., $g(t,x) \to h(x_0)$ *nontangentially*] if, for every $c > 0$, $g(r,\theta) \to h(\theta_0)$ as $(r,\theta) \to (1,\theta_0)$ in the region where $|\theta - \theta_0| < c(1-r)$ [resp., $g(t,x) \to h(x_0)$ as $(t,x) \to (0,x_0)$ in the region where $|x - x_0| < ct$].

***Theorem* 2** *(Fatou,* 1906*) — If $f \in L^1(\mathbb{T})$ and u is defined by* (2)*, then $u(r,\theta) \to f(\theta_0)$ nontangentially at almost every θ_0 — more precisely, at every θ_0 in the Lebesgue set of f.*

One of the fundamental properties of the Fourier basis $\{e^{ik\theta}\}$ is that it is orthonormal (with respect to the measure $d\theta/2\pi$), as was recognized in the 19th century in connection with Sturm-Liouville theory. This aspect of the theory was put in the spotlight in a paper of F. Riesz[1] [70] and almost simultaneously in a paper of E. Fischer [25] concerning square-integrable functions (the collection of which had not yet been named "L^2," though we shall employ that notation). In modern language,

[1] The "F." could stand for Frigyes, Friedrich, or Frédéric, all linguistic variants of the same name, depending on whether Riesz was writing in Hungarian, German, or French.

Reisz's and Fischer's results both amount to the completeness of L^2; but here they are in their original form.

***Theorem* 3** *(F. Riesz, Fischer;* 1907*)* —

(a) (Fischer) If $\{f_n\}$ is a sequence in $L^2([a,b])$ such that $\int_a^b |f_n(x) - f_m(x)|^2\,dx \to 0$ as $m, n \to \infty$, there is an $f \in L^2([a,b])$ such that $\int_a^b |f_n(x) - f(x)|^2\,dx \to 0$.

(b) (Riesz) If $\{\phi_k\}$ is an orthonormal sequence in $L^2([a,b])$ and $\{c_k\}$ is a sequence of numbers such that $\sum |c_k|^2 < \infty$, there is an function $f \in L^2([a,b])$ such that $\int_a^b f(x)\overline{\phi_k(x)}\,dx = c_k$ for all k.

Riesz's theorem is a corollary of Fischer's, as Fischer pointed out; the function in question is, of course, $f = \sum c_k\phi_k$, and the point is that the partial sums of this series satisfy Fischer's hypothesis. However, Riesz's (slightly earlier) argument did not invoke this result explicitly, and it is easy to deduce Fischer's theorem from it.

The year 1910 witnessed several developments that are relevant to our story. First, Riesz [71] invented L^p spaces ($1 \le p < \infty$) and developed their basic theory. It is then an easy corollary of Fejér's results on uniform convergence in Theorem 1, together with the density of $C(\mathbb{T})$ in $L^p(\mathbb{T})$, that the analogous results hold for convergence in L^p; that is, if $f \in L^p(\mathbb{T})$ ($p < \infty$) then $\|\sigma_n^f - f\|_p \to 0$ as $n \to \infty$ and $\|u(r, \cdot) - f\|_p \to 0$ as $r \to 1$.

Second, Michel Plancherel published the memoir on integral representations of functions [67] that led to the attachment of his name to the fact that the Fourier transform is a unitary operator on $L^2(\mathbb{R})$. In fact, this result is not explicitly in Plancherel's paper. His main results, in our context, are a rigorous definition of the Fourier cosine transform

$$\mathcal{F}_c f(\xi) = \int_0^\infty f(x)\cos(\xi x)\,dx$$

for $f \in L^2_{\mathbb{R}}(0, \infty)$ (all of Plancherel's functions are real-valued) and a proof of the inversion formula for it, namely, $\mathcal{F}_c^2 = (\pi/2)I$. The unitarity of $\mathcal{F}_c$ (up to a scalar factor) is, however, an immediate consequence: since $\int (\mathcal{F}_c f)g = \iint f(x)g(\xi)\cos(\xi x)\,dx\,d\xi = \int f(\mathcal{F}_c g)$, we have $\int (\mathcal{F}_c f)^2 = \int f(\mathcal{F}_c^2 f) = (\pi/2)\int f^2$.

The third development is Alfred Haar's investigation [33] of expansion of functions in terms of orthonormal bases (part of his doctoral thesis). The first part of this paper concerns the orthonormal bases for $L^2([a,b])$ consisting of eigenfunctions of Sturm-Liouville problems; Haar shows that they share with the Fourier basis the property that the series associated to a continuous function f need not converge pointwise but is always Cesàro summable to f. (This is not unexpected, since the high-frequency eigenfunctions of such problems tend to look a lot like sine waves.) He then asks whether there is an orthonormal basis for $L^2([0,1])$ with the property that the expansion of any continuous function is uniformly convergent, and he produces the affirmative answer now known as the *Haar basis*. It is defined as follows; for future reference we consider its elements as functions on $\mathbb{R}$, although for the present we are interested in them only on $[0,1]$. We set

$$\psi_0 = \chi_{[0,1]}, \qquad \psi_1 = \chi_{[0,1/2)} - \chi_{(1/2,1]}, \tag{7}$$

and for $n = 2^j + k$ for $j \geq 1$ and $0 \leq k < 2^j$,

$$\psi_n(x) = 2^{j/2}\psi_1(2^j x - k.) \tag{8}$$

(That is, for $n = 2^j + k$, ψ_n is $2^{j/2}$ on the left half of $[2^{-j}k, 2^{-j}(k+1)]$, $-2^{j/2}$ on the right half, and 0 elsewhere.) It is an elementary exercise to check that (i) $\{\psi_n\}_0^\infty$ is orthonormal in $L^2([0,1])$ and (ii) the linear span of $\psi_0, \ldots, \psi_{2^j-1}$ is the set of all functions on $[0,1)$ that are constant on each subinterval $(2^{-j}k, 2^{-j}(k+1))$ and satisfy $f(x) = \frac{1}{2}[f(x-) + f(x+)]$ at the break points $x = 2^{-j}k$; it follows easily that $\{\psi_n\}_0^\infty$ is an orthonormal basis for $L^2([0,1))$. Moreover, since the ψ_n's are bounded, the coefficients $\langle f, \psi_n\rangle = \int_0^1 f(x)\psi_n(x)\,dx$ make sense for all $f \in L^1([0,1])$.

***Theorem* 4** *(Haar,* 1910*) — If $f \in L^1([0,1])$, the series $\sum_0^\infty \langle f, \psi_n\rangle\psi_n(x)$ converges to $f(x)$ at every x in the Lebesgue set of f (in particular, almost everywhere). If f is continuous on $[0,1]$, the convergence is uniform.*

4. Fourier Series and Complex Analysis, 1915-1930

Many of the developments in Fourier analysis between the two world wars had to do, directly or indirectly, with the connection between Fourier series on $\mathbb{T}$ and holomorphic or harmonic functions on $\mathbb{D}$. We have already encountered the harmonic extension $u(r,\theta)$ of a function $f(\theta)$, defined by (2). The harmonic conjugate of u that vanishes at the origin is

$$v(r,\theta) = \frac{1}{i}\sum_{-\infty}^{\infty} c_k(\operatorname{sgn} k) r^{|k|} e^{ik\theta}, \qquad \operatorname{sgn} k = \begin{cases} k/|k| & \text{if } k \neq 0, \\ 0 & \text{if } k = 0. \end{cases} \tag{9}$$

(By "harmonic conjugate" we mean a function v that satisfies the Cauchy-Riemann equations $v_x = -u_y$, $v_y = u_x$. This makes sense even if u is complex-valued, although the original researchers generally had in mind that f and u are real-valued.) It is known (see Zygmund [96, §VII.1]) that

$\lim_{r\to 1} v(r,\theta)$ exists almost everywhere; it is formally given by the series $(1/i)\sum c_k(\operatorname{sgn} k)e^{ik\theta}$. We denote this function by $\widetilde{f}(\theta)$ and call it the (*Fourier*) *conjugate* function of f:

$$\widetilde{f}(\theta) = \lim_{r\to 1} v(r,\theta) \sim \frac{1}{i}\sum_{-\infty}^{\infty} c_k(\operatorname{sgn} k)e^{ik\theta}. \tag{10}$$

One of the main lines of development concerns what we now call H^p spaces or Hardy spaces. There are now several types of Hardy spaces, but the original ones are defined as

$$H^p(\mathbb{D}) = \left\{F \text{ holomorphic on } \mathbb{D} : \|F\|_{H_p} \equiv \sup_{r<1}[M_F^p(r)]^{1/p} < \infty\right\} \qquad (0 < p < \infty), \tag{11}$$

where

$$M_F^p(r) = \frac{1}{2\pi}\int_0^{2\pi} |F(re^{i\theta})|^p\, d\theta. \tag{12}$$

($\|\cdot\|_{H_p}$ is a norm only for $p \geq 1$.) These spaces are named in honor of G. H. Hardy, but Hardy did not invent them. Rather, in a 1915 paper [36] he proved that if $F(z)$ is holomorphic in the disc $|z| < R$, then $M_F^p(r)$ is an increasing function of r for $r < R$, and moreover log $M_F^p(r)$ is a convex function of $\log r$. The spaces $H^p(\mathbb{D})$ were first formally defined, and named H^p in recognition of Hardy's theorem, in a 1923 paper of F. Riesz [72].

The first serious work on these spaces, however, came several years earlier, in the Riesz brothers' only joint paper [73]. In the first part of this paper they showed that the boundary values of a holomorphic function on $\mathbb{D}$, under suitable hypotheses including the H^p condition for $p > 1$, cannot vanish on a set of positive linear measure in $\mathbb{T}$. They then turned to the case $p = 1$ and proved the following fundamental result:

***Theorem* 5** *(F. and M. Riesz,* 1916*) — A holomorphic function $F(z) = \sum_0^\infty a_k z^k$ on $\mathbb{D}$ belongs to $H^1(\mathbb{D})$ if and only if $\sum_0^\infty a_k e^{ik\theta}$ is the Fourier series of a function $f(e^{i\theta})$ in $L^1(\mathbb{T})$. In this case, F and f determine each other by the relations*

$$F(z) = \frac{1}{2\pi i}\int_{|w|=1} \frac{f(w)}{w-z}\, dw, \qquad f(e^{i\theta}) = \lim_{r\to 1} F(re^{i\theta}) \textit{ for a.e. } \theta.$$

The "if" implication of Theorem 5 is quite easy, for F can be derived from f not only by the Cauchy integral as above but by the Poisson integral (3). The facts that $P(r,\theta) \geq 0$ and $\int_0^{2\pi} P(r,\theta)\, d\theta = 1$ for all $r < 1$ (see (4)) then easily imply that $F \in H^1(\mathbb{D})$. The converse is the hard part; the essential ingredient in its proof is the following result, which is of interest in its own right:

***Theorem* 6** *(F. and M. Riesz,* 1916*) — Let μ be a complex Borel measure on $\mathbb{T}$. If the Fourier coefficients $c_k = (1/2\pi)\int_0^{2\pi} e^{-ik\theta} d\mu(\theta)$ vanish for $k < 0$, then μ is absolutely continuous with respect to Lebesgue measure.*

The reader of [73], however, may have trouble locating this result therein. What is actually stated and proved there is as follows: Suppose $F(z)$ is bounded on $\mathbb{D}$ and its boundary values $F(e^{i\theta})$ on $\mathbb{T}$ exist and constitute a function of bounded variation. Then this function is continuous (this was already known), so that the image of $\mathbb{T}$ under it is a rectifiable curve $F(\mathbb{T})$; and the image of any set of measure zero in $\mathbb{T}$ has measure zero in $F(\mathbb{T})$. (To obtain Theorem 6 from this, one takes $F(z)$ to be the integrated series $\sum_0^\infty c_k z^{k+1}/(k+1)$, so that $d\mu(\theta)$ is the Lebesgue-Stieltjes measure $dF(e^{i\theta})$.)

Theorem 5 can be restated in the following way, which we present as a corollary. (The f in Theorem 5 corresponds to $f + i\widetilde{f}$ here.)

Corollary 7 — Suppose $f \in L^1(\mathbb{T})$, and let u, v, and $\widetilde{f}$ be defined by (2), (9) and (10). Then $u + iv \in H^1(\mathbb{D})$ if and only if $\widetilde{f} \in L^1(\mathbb{T})$.

Corollary 7 remains true if L^1 and H^1 are replaced by L^p and H^p with $p > 1$, but it is easier in the latter case. In fact, there it is a corollary of the analogous result for harmonic functions; namely, a harmonic function $u(r, \theta)$ on $\mathbb{D}$ satisfies $\sup_{r<1} \|u(r, \cdot)\|_{L^p(\mathbb{T})} < \infty$ $(1 < p < \infty)$ if and only if u is the Poisson integral of an $f \in L^p(\mathbb{T})$. The "if" implication follows from elementary properties of the Poisson kernel as above, and for the converse, one can use the weak compactness of the closed unit ball of L^p (proved by F. Riesz in [71]) to show that as $r \to 1$, $u(r, \cdot)$ converges weakly in L^p to an $f \in L^p(\mathbb{T})$ of which u is the Poisson integral. For $p = 1$, however, this breaks down and one can conclude only that u is the Poisson integral of a measure on $\mathbb{T}$.

There is more to the story, of course. Simple examples show that the conjugate of an L^1 function need not be in L^1, but what about L^p for $p > 1$? The case $p = 2$ is obvious: since the Fourier basis is orthogonal, from (10) we have $\|\widetilde{f}\|_2^2 = 2\pi \sum_{k\neq 0} |c_k|^2 \leq \|f\|_2^2$. But for $p \neq 2$, the answer lies deeper. It was announced by Marcel Riesz in 1924 [74], though he did not get around the publishing the details of his proof until 1927 [75]:

***Theorem* 8** *(M. Riesz,* 1924*) — The map $f \mapsto \widetilde{f}$ is bounded on $L^p(\mathbb{T})$ for $1 < p < \infty$.*

This result is now seen as the prototype example of the theory of singular integrals, which we shall discuss in §7, but the tools needed to build that theory did not yet exist in 1924. (The weak-type estimate for $p = 1$ was proved by Kolmogorov [48] in 1925, but the interpolation theorem from which the L^p estimate then follows came considerably later; see §5.) Rather, Riesz devised a clever argument that exploited the connection with complex analysis. It has the disadvantage that it does not generalize at all, but it is still worth a little attention for its esthetic value. It shows that if u and v are the Poisson integrals of f and $\widetilde{f}$, then

$$\int_0^{2\pi} |v(r,\theta)|^p \, d\theta \le C_p \int_0^{2\pi} |u(r,\theta)|^p \, d\theta \qquad (0 < r < 1) \tag{13}$$

with C_p independent of r and f, and the result follows by letting $r \to 1$.

We explain the idea by working out the case $p = 4$. We may assume that f is real-valued (so that u and v are too) and that the constant term c_0 in the Fourier series of f vanishes. Let $F = u + iv$. Then $F(0) = c_0 = 0$, so by the Cauchy integral formula,

$$0 = \int_{|z|=r} \frac{F(z)^4}{iz} \, dz = \int_0^{2\pi} (u(r,\theta) + iv(r,\theta))^4 \, d\theta.$$

Taking real parts yields

$$\int v^4 = 6 \int u^2 v^2 - \int u^4 \le 6 \left[\int u^4 \int v^4 \right]^{1/2}$$

(with obvious abbreviations), and this immediately gives (13) with $C_4 = 36$.

The same idea, with some elaboration, works when p is any even integer, and Riesz managed to push it further so that it works when p is anything except an odd integer. (The emergence of an inconvenient factor of $\cos(\pi p/2)$ at a certain point spoils the argument in these exceptional cases.) But if $p > 1$ is an odd integer, its conjugate exponent $p/(p-1)$ lies in the interval $(1, 2)$, where the result is valid, so a simple duality argument completes the proof.

As Riesz observed, his theorem has the following important consequence that has no apparent connection to complex analysis:

Corollary 9 — If $f \in L^p(\mathbb{T})$ with $1 < p < \infty$, the partial sums S_n^f of its Fourier series given by (1) converge to f in the L^p norm as $n \to \infty$.

The proof is easy. Since the trigonometric polynomials are dense in L^p (a consequence of Fejér's theorem), by the usual $\epsilon/3$-argument it is enough to show that $\|S_n^f\|_p \le C_p \|f\|_p$ with C_p independent of n. But $S_n^f = (E_{-n} P E_n - E_{n+1} P E_{-(n+1)}) f$ where $P(\sum_{-\infty}^{\infty} c_k e^{ik\theta}) = \sum_0^{\infty} c_k e^{ik\theta}$ and $E_n f(\theta) = e^{in\theta} f(\theta)$. Clearly E_n is an isometry on L^p for every n, and $Pf = \frac{1}{2}(f + i\widetilde{f})$ (assuming, as above, that $c_0 = 0$); the result therefore follows from Theorem 8.

The L^2 convergence of a Fourier series is, of course, unaffected by the order in which the terms are added up, as any rearrangement of an orthonormal basis is again an orthonormal basis; but the validity of Corollary 9 is strongly dependent on the fact that the partial sums S_n^f are taken to be the standard ones. It is well known that in any Banach space $\mathfrak{X}$, a series $\sum_1^\infty x_n$ converges unconditionally (that is, $\sum_1^\infty x_{\sigma(n)}$ converges to the same sum for any permutation σ of $\mathbb{Z}^+$) if and only the series $\sum_1^\infty \epsilon_n x_n$ converges for all choices of $\epsilon_n \in \{-1, 1\}$. (See, for example, [42].) In 1930 Littlewood [52] showed that this is not the case when $\mathfrak{X} = L^p(\mathbb{T})$ ($p \neq 2$) and $x_n = S_n^f - S_{n-1}^f$, and shortly afterward Paley and Zygmund [63] obtained the following remarkable generalization of Littlewood's result.

Paley and Zygmund made use of the convenient encoding of sequences of ± 1's by means of the *Rademacher functions*, first studied in [68]. They are most easily defined analytically as

$$r_n(t) = \operatorname{sgn}[\sin(2^{n+1}\pi t)] \qquad (0 \leq t \leq 1.) \tag{14}$$

In other words, as long as t is not a dyadic rational, $r_n(t) = (-1)^{d_{n+1}(t)}$ where $d_n(t) \in \{0, 1\}$ is the nth digit in the base-2 decimal expansion of t (but $r_n(j/2^k) = 0$ for $n \geq k$). Thus any sequence $\{\epsilon_n\}$ of ± 1's except those that contain only finitely many 0s or finitely many 1s can be written as $\{r_n(t)\}$ for a unique $t \in (0, 1)$. Moreover, it is easily seen that the map $t \mapsto \{r_n(t)\}$ is measure-preserving from Lebesgue measure on $(0, 1)$ (with the dyadic rationals omitted) to the natural probability measure on $\prod_0^\infty \{-1, 1\}$ determined by a sequence of tosses of a fair coin.

***Theorem* 10** *(Paley-Zygmund,* 1930*) — Given a sequence $\{c_n\}_{-\infty}^\infty$ of complex numbers, let $A_n(\theta) = c_n e^{in\theta} + c_{-n} e^{-in\theta}$.*

(a) If $\sum |c_n|^2 < \infty$, for almost every $t \in (0, 1)$ the series $\sum_0^\infty r_n(t) A_n(\theta)$ converges a.e. on $\mathbb{T}$ to a function in $\bigcap_{p<\infty} L^p(\mathbb{T})$.

(b) If $\sum |c_n|^2 = \infty$, for almost every $t \in (0, 1)$ the series $\sum_0^\infty r_n(t) A_n(\theta)$ diverges (and indeed is not Cesàro or Abel summable) a.e. on $\mathbb{T}$.

Thus, by taking the c_n's to be the Fourier coefficients of a function in $L^2 \setminus L^p$ (if $p > 2$) or in $L^p \setminus L^2$ (if $p < 2$), one obtains examples of functions in L^p such that the L^p convergence of their Fourier series is destroyed by suitable insertions of factors of ± 1.

The proof of Theorem 10 is a rather easy consequence of Fubini's theorem together with the following properties of the Rademacher functions, which we shall meet again in §10; see Zygmund [96, §V.8] for details.

Lemma 11 — If $\sum_0^\infty |a_n|^2 < \infty$, the series $\sum a_n r_n(t)$ converges a.e. on $[0, 1]$ to a function $f \in \bigcap_{p<\infty} L^p([0, 1])$; moreover, there are constants A_p and B_p depending only on p such that $A_p \sum |a_n|^2 \leq \|f\|_p^2 \leq B_p \sum |a_n|^2$. On the other hand, if $\sum_0^\infty |a_n|^2 = \infty$, the series $\sum a_n r_n(t)$ diverges (and indeed is not Cesàro or Abel summable) for a.e. $t \in [0, 1]$.

The situation with L^p convergence of Fourier series is quite different if one considers not the whole sequence of partial sums S_n^f but only a subsequence $S_{n(m)}^f$ where $n(m)$, roughly speaking, grows exponentially with m. To be specific, let $n(m) = 2^m$; thus $S_{n(m)}^f$ is the mth partial sum of the series $c_0 + \sum_{l=0}^\infty \Delta_l^f$ where $\Delta_l^f = \sum_{2^{l-1}<|k|\le 2^l} c_k e^{ik\theta}$. It turns out that the series $c_0 + \sum \Delta_l^f$ does converge unconditionally in L^p when $f \in L^p(\mathbb{T})$, $1 < p < \infty$, and the same is true for other similar choices of $n(m)$. But this is a small part of a much longer story, the "Littlewood-Paley theory," which we shall recount in §10.

The final development from this period on our agenda is the great 1930 paper of Hardy and Littlewood [37]. In it they begin by introducing the concept of nonincreasing rearrangements and prove some fundamental results about them; then they introduce the Hardy-Littlewood maximal function for functions in $L^1(\mathbb{T})$,

$$Mf(\theta) = \sup_{0<|t|<\pi} \frac{1}{t} \int_0^t |f(\theta+\phi)|\, d\phi,$$

and the nontangential maximal functions for functions on $\mathbb{D}$,

$$u_c^*(\theta) = \sup_{(r,\phi)\in S_c(\theta)} |u(r,\phi)| \qquad (0 \le c < 1), \tag{15}$$

where the nontangential approach region $S_c(\theta)$ may be taken to be the convex hull of the disc $r \le c$ and the point $(1,\theta)$. (The precise shape of $S_c(\theta)$ outside a neighborhood of $(1,\theta)$ is unimportant, as is the value of c, except that when $c = 0$ u_c^* becomes the radial maximal function $u_0^*(\theta) = \sup_{0<r<1} |u(r,\theta)|$.) The main results are as follows:

***Theorem* 12** *(Hardy-Littlewood,* 1930*)* —

(a) For $1 < p \le \infty$ there is a constant A_p such that $\|Mf\|_p \le A_p \|f\|_p$ for all $f \in L^p(\mathbb{T})$.

(b) For $0 \le c < 1$ there is a constant B_c such that $u_c^*(\theta) \le B_c Mf(\theta)$ whenever $f \in L^1(\mathbb{T})$ and u is its harmonic extension to $\mathbb{D}$; consequently, $\|u_c^*\|_p \le A_p B_c \|f\|_p$ for $p > 1$.

(c) For $0 < p < \infty$ and $0 \le c < 1$ there is a constant $C_{p,c}$ such that $\|F_c^*\|_p \le C_{p,c} \|F\|_{H^p}$ for all $F \in H^p(\mathbb{D})$. (Here, of course, $F_c^* = u_c^*$ where $u(r,\theta) = F(re^{i\theta})$).

Part (c) is a corollary of part (b) when $p > 1$, for then F is the harmonic extension of an $f \in L^p(\mathbb{T})$ and $\|F\|_{H^p} = \|f\|_p$, but for $p \le 1$ it expresses a special property of holomorphic functions that is not shared by all harmonic functions.

5. Interpolation Theorems

Among the powerful tools in the more modern development of harmonic analysis are some theorems to the effect that if one has L^p estimates on an operator for two different values of p, one also obtains such estimates for the intermediate values of p. The first such theorem was proved by Marcel Riesz [76]. Let us first state it in the general form in which it is now familiar:

***Theorem* 13** *(M. Riesz,* 1927*; Thorin,* 1939*) — Let* (X,μ) *and* (Y,ν) *be measure spaces and* $p_0, p_1, q_0, q_1 \in [1,\infty]$*. Suppose* T *is a linear map from* $L^{p_0}(\mu) + L^{p_1}(\mu)$ *to* $L^{q_0}(\nu) + L^{q_1}(\nu)$ *that is bounded from* $L^{p_j}(\mu)$ *to* $L^{q_j}(\nu)$ *for* $j = 0,1$*. Then* T *is bounded from* $L^{p_t}(\mu)$ *to* $L^{q_t}(\nu)$ *for* $0 < t < 1$*, where* $p_t^{-1} = (1-t)p_0^{-1} + tp_1^{-1}$ *and* $q_t^{-1} = (1-t)q_0^{-1} + tq_1^{-1}$*. More precisely, if*

$$M(p,q) = \sup\bigl\{\|Tf\|_q : \|f\|_p = 1\bigr\},$$

then $\log M(p_t, q_t)$ *is a convex function of* t*.*

As Riesz observed, this general theorem is a corollary of the following, apparently much more special, result in finite-dimensional linear algebra:

***Theorem* 14** *— Let* (A_{jk}) *be a complex* $m \times n$ *matrix, and for* $\alpha, \beta \in [0,1]$ *let*

$$\mathcal{M}(\alpha,\beta) = \sup\left\{\Bigl|\sum_{jk} A_{jk} y_j x_k\Bigr| : \sum_1^n |x_j|^{1/\alpha} = \sum_1^m |y_k|^{1/\beta} = 1\right\}.$$

(If $\alpha = 0$ *or* $\beta = 0$*, the condition on* x *or* y *is interpreted as* $\max|x_j| = 1$ *or* $\max|y_k| = 1$*.) Then* $\log \mathcal{M}(\alpha,\beta)$ *is a convex function on the square* $[0,1]\times[0,1]$*.*

To make the connection, observe that the $M(p,q)$ in Theorem 13 can be rewritten as

$$M(p,q) = \sup\Bigl\{\Bigl|\int (Tf)g\,d\nu\Bigr| : \|f\|_p = \|g\|_{q'} = 1\Bigr\},$$

where $q' = q/(q-1)$ is the conjugate exponent to q, and moreover it is enough to take f and g to be simple functions, i.e., $f = \sum_1^n a_j \chi_{E_j}$ and $g = \sum_1^m b_k \chi_{F_k}$ where the E_j's (resp. the F_k's) are disjoint subsets of X (resp. of Y) of finite measure. But for such f and g (for a fixed set of E_j's and F_k's) the quantity $M(p,q)$ reduces to a quantity of the form $\mathcal{M}(\alpha,\beta)$ with $\alpha = 1/p$ and $\beta = 1/q'$, and the result follows.

Riesz proved Theorem 14 only for (α,β) in the triangle $\{(\alpha,\beta) \in [0,1]^2 : \alpha+\beta \geq 1\}$, which yields Theorem 13 only under the condition that $q_j \geq p_j$ for $j = 0,1$. The lovely proof via the "three lines theorem" of complex analysis, which renders this restriction superfluous, was discovered by Riesz's student G. Thorin [92].

Thorin's argument points the way toward a powerful and wide-ranging generalization of Theorem 13, due to Elias M. Stein [79] (see also [87]), in which the single operator T is replaced by an analytic family of operators:

***Theorem* 15** *(Stein,* 1956*) — Let (X,μ), (Y,ν), p_t, and q_t be as in Theorem 13, and let $\Sigma = \{x+iy : x\in[0,1],\ y\in\mathbb{R}\}$. Suppose that for each $z\in\Sigma$ we have a linear map T_z from simple functions on X to locally integrable*[2] *functions on Y such that $z\mapsto \int (T_z f)g$ is continuous on Σ and holomorphic on its interior for all simple functions f on X and g on Y. Assume that*

$$\|T_{iy}f\|_{q_0} \le M_0(y)\|f\|_{p_0} \quad \textit{and} \quad \|T_{1+iy}f\|_{q_1} \le M_1(y)\|f\|_{p_1},$$

and there are constants $C_0, C_1, C_{fg} > 0$ and $a<\pi$ such that

$$\left|\int (T_{x+iy}f)g\right| \le \exp[C_{f,g}e^{a|y|}], \quad M_0(y) \le \exp[C_0 e^{a|y|}], \quad M_1(y)\le \exp[C_1 e^{a|y|}].$$

Then for $0<t<1$ there is a constant M_t (which can be estimated in terms of the functions $M_0(y)$ and $M_1(y)$) such that $\|T_t f\|_{q_t} \le M_t\|f\|_{p_t}$.

The conditions on the growth of $\int(T_{x+iy}f)g$, $M_0(y)$, and $M_1(y)$ in y are extremely weak; they are needed to guarantee the hypotheses of the Phragmén-Lindelöf-type theorem on which the proof rests. If these quantities are bounded uniformly in y, one can use the three lines theorem instead to conclude that $\log M_t$ is a convex function of t.

As a simple illustration of the use of Theorem 15, we consider operators on L^2 Sobolev spaces. If k is a positive integer, the *Sobolev space* $L^{2,k} = L^{2,k}(\mathbb{R}^n)$ is defined to be the space of all $f\in L^2(\mathbb{R}^n)$ whose distribution derivatives $\partial^\alpha f$ are also in $L^2(\mathbb{R}^n)$ for $|\alpha|\le k$. (We are employing the usual multi-index notation as in [29]: α is an n-tuple of nonnegative integers and $|\alpha| = \alpha_1+\cdots+\alpha_n$.) Since $(\partial^\alpha f)\widehat{\ }(\xi) = (2\pi i\xi)^\alpha \widehat{f}(\xi)$, by the Plancherel theorem we have $f\in L^2_k$ if and only if the functions $\xi\mapsto \xi^\alpha\widehat{f}(\xi)$ are in L^2 for $|\alpha|\le k$. But it is easily verified that $\sum_{0\le|\alpha|\le k}|\xi^\alpha|^2$ is bounded above and below by constant multiples of $(1+|\xi|^2)^k$, so $L^{2,k} = L^2_k$ where, for any $s\in\mathbb{R}$, we define

$$L^2_s = \{f : \Lambda^s f\in L^2\}, \quad \text{where} \quad (\Lambda^s f)\widehat{\ }(\xi) = (1+|\xi|^2)^{s/2}\widehat{f}(\xi.) \tag{16}$$

(When $s\ge 0$, the elements of Λ^2_s are L^2 functions, but when $s<0$, they are tempered distributions whose Fourier transforms are locally L^2 functions. The space L^2_s is often denoted by H_s.)

Proposition 16 — Suppose $s_0<s_1$ and $r_0<r_1$, and suppose T is a bounded linear map from $L^2_{s_0}$ to $L^2_{r_0}$ whose restriction to $L^2_{s_1}$ is bounded from $L^2_{s_1}$ to $L^2_{r_1}$. Then the restriction of T to $L^2_{s_t}$ is bounded from $L^2_{s_t}$ to $L^2_{r_t}$ for $0<t<1$, where $s_t = (1-t)s_0 + ts_1$ and $r_t = (1-t)r_0+tr_1$.

[2] i.e., integrable on sets of finite measure

Indeed, since Λ^s is an isomorphism from $L^2 = L^2_0$ to L^2_s for every $s \in \mathbb{R}$, T is bounded from L^2_s to L^2_r if and only if $\Lambda^r T \Lambda^{-s}$ is bounded on L^2, so Theorem 16 follows by applying Theorem 15 to $T_z = \Lambda^{r(z)} T \Lambda^{-s(z)}$ where $s(z) = (1-z)s_0 + zs_1$ and $r(z) = (1-z)r_0 + zr_1$. The necessary estimates as $|\operatorname{Im} z| \to \infty$ are trivial since $\Lambda^{x+iy} = \Lambda^x \Lambda^{iy}$ and Λ^{iy} is unitary on L^2.

For any $p \in [1,\infty]$ we can define L^p Sobolev spaces $L^{p,k}$ and L^p_s just as above, simply replacing L^2 by L^p (but keeping Λ^s unchanged), and the analogue of Proposition 16 generalizes provided that we have appropriate estimates on Λ^{iy}. However, for $p \neq 2$ these estimates are no longer obvious, nor is it obvious that $L^{p,k} = L^p_k$ — and indeed these claims are no longer valid for $p = 1$ or $p = \infty$. They are, however, valid for $1 < p < \infty$, and we shall indicate the proofs in §7 and §10.

We now turn to the other major interpolation theorem, whose setting is as follows: (X,μ) and (Y,ν) are again measure spaces, and T is now a map from some space $\mathcal{D}$ of measurable functions on X to the space of measurable functions on Y that is *quasi-linear*: that is, there is a $K > 0$ such that $|T(f+g)| \leq K(|Tf| + |Tg|)$ for all $f, g \in \mathcal{D}$. For $1 \leq p, q \leq \infty$ we say that T is *strong type* (p,q) if $L^p(\mu) \subset \mathcal{D}$ and there is a constant C such that $\|Tf\|_q \leq C\|f\|_p$, and *weak type* (p,q) ($q < \infty$) if $L^p(\mu) \subset \mathcal{D}$ and there is a constant C such that

$$\nu\big(\{y : |Tf(y)| > \alpha\}\big) \leq C\left(\frac{\|f\|_p}{\alpha}\right)^q; \tag{17}$$

we also agree that "weak type (p,∞)" means "strong type (p,∞)." We observe that the quantity on the left of (17) is at most $(\|Tf\|_q/\alpha)^q$ (Chebyshev's inequality), so strong type (p,q) implies weak type (p,q).

***Theorem* 17** *(Marcinkiewicz,* 1939*) — Suppose T is weak types (p_0,q_0) and (p_1,q_1) where $p_0 \leq q_0$, $p_1 \leq q_1$, and $q_0 \neq q_1$. Then T is strong type (p_t,q_t) for $0 < t < 1$, where $p_t^{-1} = (1-t)p_0^{-1} + tp_1^{-1}$ and $q_t^{-1} = (1-t)q_0^{-1} + tq_1^{-1}$.*

This result was announced by J. Marcinkiewicz in a short note [59] in 1939. Not long afterwards, his career was cut short by World War II, and he left only a brief sketch of a proof in a letter to Zygmund. Marcinkiewicz's theorem languished in obscurity for 17 years — in their fundamental paper [4] Calderón and Zygmund derived the special case they needed without quoting [59] — until Zygmund wrote up a full account in [95]. Since then it has become part of the analysts' standard toolkit.

6. The Transition to $\mathbb{R}^n$

Much of the development of harmonic analysis in the second half of the 20th century had roots in the material we discussed in §2, but with some shifts of focus. The earlier work concerned Fourier series as a technique of studying functions on the circle $\mathbb{T}$ and harmonic and analytic functions on the disc $\mathbb{D}$. In many respects it is quite easy to shift attention to Fourier transforms as a technique for studying functions on $\mathbb{R}$ and harmonic and analytic functions on the upper half-plane. We shall not go into detail about all the analogous results that can be obtained in this setting, singling out only one for particular attention. Namely, the analogue on $\mathbb{R}$ of the map $f \to \widetilde{f}$ taking a Fourier series to its conjugate is the *Hilbert transform* H, the unitary operator on $L^2(\mathbb{R})$ defined by

$$\widehat{Hf}(\xi) = \frac{1}{i}(\operatorname{sgn}\xi)\widehat{f}(\xi),$$

which can also be defined without reference to the Fourier transform as

$$Hf(x) = \frac{1}{\pi}\lim_{\epsilon\to 0}\int_{|y|>\epsilon}\frac{f(x-y)}{y}\,dy,$$

where the limit exists in the L^2 norm. (It also exists pointwise when f is also Hölder continuous.) As Marcel Riesz [75] showed, his theorem on conjugate series remains valid in this setting, with much the same proof:

***Theorem* 18** *(M. Riesz,* 1924*) — H is bounded on L^p for $1 < p < \infty$.*

Of greater significance than the shift from $\mathbb{T}$ to $\mathbb{R}$ is the shift from $\mathbb{R}$ to $\mathbb{R}^n$, which necessitates weaning oneself away from a reliance on complex function theory and developing a new set of tools to replace it, but which also offers a wealth of new applications. This will be the main theme of the next few sections.

In the early days, one obstacle to unleashing the power of the Fourier transform was that its use seemed to be restricted to functions that are rather tame at infinity, so that all the integrals in question converge. Various mathematicians in the 1930s made attempts to resolve such issues as well as related ones concerning notions of "generalized derivatives." In part they were goaded by the physicists who refused to be limited by scruples about rigor and employed formulas such as

$$\int_{\mathbb{R}} e^{2\pi i x\xi}\,d\xi = \delta(x) \quad (\delta = \text{Dirac "delta-function"}) \tag{18}$$

with great success. But it was Laurent Schwartz who, in the 1940s, developed a conceptual framework that offered a simple and painless resolution of most of these difficulties: his theory of distributions. The basic ideas were presented in [77] and [78], and the theory gained immediate popularity when more extensive expository accounts became available soon afterwards.

The basics of this theory are sufficiently accessible (see, for example, [29]) that we shall not go into any detail here. However, the Schwartz space $\mathcal{S} = \mathcal{S}(\mathbb{R}^n)$ of C^∞ functions which, together with all their derivatives, decay faster than polynomially at infinity, and its dual space $\mathcal{S}'$, the space of tempered distributions, will occasionally enter the discussion (they already did so in the previous section). We recall that among the operations on $\mathcal{S}$ that extend continuously to operations on $\mathcal{S}'$ are the Fourier transform, differentiation, translation, composition by invertible linear transformations of $\mathbb{R}^n$, and multiplication by C^∞ functions which, together with all their derivatives, grow at most polynomially at infinity. In particular, the constant function 1 and the point mass at the origin are Fourier transforms of each other (the rigorous interpretation of (18)). Moreover, composition with dilations is well defined on $\mathcal{S}'$, so it makes sense to say that a distribution is homogeneous of degree α (namely, $F \circ \delta_r = r^\alpha F$ for $r > 0$, where $\delta_r(x) = rx$), and it is easily verified that its Fourier transform is then homogeneous of degree $-n - \alpha$.

We shall denote the action of a distribution $F \in \mathcal{S}'$ on a test function $\phi \in \mathcal{S}$ by $\langle F, \phi\rangle$. This pairing is bilinear. In §11 and §12 we also use the notation $\langle\cdot,\cdot\rangle$ for the sesquilinear inner product on L^2, but the meaning will be clear from the context.

7. Singular Integrals

One of the most significant milestones in the development of harmonic analysis in $\mathbb{R}^n$ was the Calderón-Zygmund theory of singular integral operators, a far-reaching generalization of the theory of the Hilbert transform, particularly Theorem 18. Various types of singular integrals — that is, operators such as the Hilbert transform involving integrals that are not absolutely convergent, usually defined in some principal-value sense — had previously been studied by several people, notably Tricomi, Giraud, and Mikhlin; but Alberto Calderón and Antoni Zygmund were the first to make a systematic study of L^p boundedness, and the techniques they developed have proved to be of wider importance.

Their initial paper [4] of 1952 deals with convolution operators of the form $T_K f = f * K$ on functions on $\mathbb{R}^n$, where K is a distribution that is homogeneous of degree $-n$ and agrees away from the origin with a function possessing some mild smoothness properties. In more detail, suppose K is a function that is of class C^1 on $\mathbb{R}^n \setminus \{0\}$ and homogeneous of degree $-n$ (i.e., $K(rx) = r^{-n}K(x)$ for $r > 0$), and that possesses the *mean-zero property*:

$$\int_{|x|=1} K(x)d\sigma(x) = 0 \qquad (\sigma = \text{surface measure on the unit sphere}). \tag{19}$$

The smoothness condition can be relaxed, as we shall discuss later. We also observe that (19) is equivalent to

$$\int_{a<|x|<b} K(x)\,dx = 0 \qquad (0 < a < b < \infty.) \tag{20}$$

Such a K fails to be integrable near the origin and near infinity, as integration in polar coordinates shows that $\int_{a<|x|<b}|K(x)|dx$ is proportional to $\int_a^b dt/t = \log(b/a)$. However, K defines a tempered distribution (still denoted by K) by the formula

$$\begin{aligned}\langle K, \phi\rangle &= \lim_{\epsilon\to 0}\int_{|x|>\epsilon} K(x)\phi(x)\,dx \\ &= \int_{|x|\le 1} K(x)[\phi(x)-\phi(0)]\,dx + \int_{|x|>1} K(x)\phi(x)\,dx.\end{aligned} \tag{21}$$

The two formulas agree because of (20), and the integrals in the second one are absolutely convergent for any $\phi \in \mathcal{S}$ because ϕ decays rapidly at infinity and $|\phi(x)-\phi(0)| = O(|x|)$.

It is easily checked that K is homogeneous of degree $-n$ *as a distribution*, and hence its Fourier transform $\widehat{K}$ is homogeneous of degree 0. Moreover, $\widehat{K}$ is not merely a distribution but a function that is continuous except at the origin. To see this, pick a C^∞ function ϕ with $\phi(x) = 1$ for $|x| \le 1$ and $\phi(x) = 0$ for $|x| \ge 2$, and write $K = \phi K + (1-\phi)K$. ϕK has compact support, so its Fourier transform is C^∞. Moreover, ∇K is homogeneous of degree $-n-1$, so $|\nabla[(1-\phi)K](x)| \le c|x|^{-n-1}$ for $|x|$ large; hence $\nabla[(1-\phi)K] \in L^1$, so its Fourier transform $2\pi i\xi[(1-\phi)K]\widehat{\ }(\xi)$ is a continuous function; but then $[(1-\phi)K]\widehat{\ }$ is itself continuous except at $\xi = 0$.

Being homogeneous of degree 0 and continuous away from the origin, $\widehat{K}$ is bounded, and it follows that *the operator T_K initially defined on $\mathcal{S}$ by*

$$(T_K f)\widehat{\ } = \widehat{K}\widehat{f},$$

or, equivalently,

$$T_K f(x) = f * K(x) = \lim_{\epsilon\to 0}\int_{|y|>\epsilon} f(x-y)K(y)\,dy, \tag{22}$$

extends to a bounded operator on L^2. (Incidentally, if K is continuous and homogeneous of degree $-n$ on $\mathbb{R}^n\setminus\{0\}$ but does not satisfy the mean-zero property, the second formula in (21) still defines a tempered distribution, but it is not homogeneous and its Fourier transform is not a bounded function.)

The fundamental theorem of Calderón and Zygmund generalizes this to L^p:

***Theorem* 19** *(Calderón-Zygmund,* 1952*) — Suppose K is of class $C^{(1)}$ and homogeneous of degree $-n$ on $\mathbb{R}^n\setminus\{0\}$ and that K satisfies (19). Then the operator T_K defined by (22) is weak type $(1,1)$ and bounded on L^p for $1<p<\infty$.*

The proof of Theorem 19 is just as important as the result, so we sketch the ideas. The main point is the weak type $(1,1)$ estimate. Since T_K is bounded on L^2, as we have just observed, the

Marcinkiewicz interpolation theorem then implies that T is bounded on L^p for $1 < p < 2$; and since $\int (T_K f)g = \int f(T_{\widetilde{K}} g)$ where $\widetilde{K}(x) = K(-x)$ (which satisfies the same hypotheses as K), a simple duality argument yields the boundedness on L^p for $2 < p < \infty$.

The main tool for obtaining the weak type $(1,1)$ estimate is the following lemma, which is also useful in other situations. Some notation: If $E \subset \mathbb{R}^n$ is a measurable set, we recall that χ_E is its characteristic function, and we denote its Lebesgue measure by $|E|$. By a *cube* we shall mean a translate of a set of the form $Q = [0, r)^n \subset \mathbb{R}^n$, and we call r the *side length* of Q. If Q is a cube and f is a locally integrable function, $m_Q f$ will denote the mean value of f on Q:

$$m_Q f = \frac{1}{|Q|} \int_Q f(x)\,dx.$$

Also, $2Q$ will denote the cube with the same center as Q and side length twice as big. Finally, for $m \in \mathbb{Z}$, $\mathcal{Q}_m$ will denote the collection of cubes with side length 2^{-m} and vertices in $(2^{-m}\mathbb{Z})^n$. (Thus the cubes in $\mathcal{Q}_{m+1}$ are obtained from the cubes in $\mathcal{Q}_m$ by bisecting their sides.)

Lemma 20 (Calderón-Zygmund, 1952) — Given a nonnegative $h \in L^1$ and $\alpha > 0$, there is a sequence $\{Q_j\}$ of disjoint cubes such that (i) $\alpha < m_{Q_j} h \leq 2^n \alpha$ for all j, (ii) $\sum |Q_j| \leq \|h\|_1/\alpha$, and (iii) $h(x) \leq \alpha$ for a.e. $x \in \mathbb{R}^n \setminus \bigcup Q_j$.

To prove the lemma, one picks $M \leq 0$ so that $m_Q h \leq 2^n \alpha$ for all $Q \in \mathcal{Q}_M$ (possible since $m_Q h \leq 2^{Mn} \|h\|_1$ for $Q \in \mathcal{Q}_M$), and one puts those $Q \in \mathcal{Q}_M$ into the sequence that satisfy $m_Q h > \alpha$ (there are only finitely many). One then proceeds inductively for $m \geq M$: having put some cubes from $\mathcal{Q}_m$ into the sequence, one puts those cubes $Q \in \mathcal{Q}_{m+1}$ into the sequence that are not contained in previously accepted cubes and satisfy $m_Q h > \alpha$. Then (i) is true by construction, (ii) is true since $\|h\|_1 \geq \sum \int_{Q_j} h > \alpha \sum |Q_j|$, and (iii) follows from the Lebesgue differentiation theorem since $m_Q f \leq \alpha$ for all $Q \in \bigcup \mathcal{Q}_m$ that contain x.

Returning to Theorem 19, given $f \in L^1$ and $\alpha > 0$, we wish to show that
$|\{x : |T_K f(x)| > \alpha\}| \leq C\|f\|_1/\alpha$. Let $\{Q_j\}$ be as in Lemma 20 with $h = |f|$, and set

$$\Omega = \bigcup Q_j, \qquad b_j(x) = [f(x) - m_{Q_j} f]\chi_{Q_j}(x), \qquad b = \sum b_j,$$
$$g(x) = f(x) - b(x) = \begin{cases} m_{Q_j} f & \text{for } x \in Q_j, \\ f(x) & \text{for } x \notin \Omega. \end{cases}$$

It is enough to show that $|\{x : |T_K g(x)| > \alpha\}|$ and $|\{x : |T_K b(x)| > \alpha\}|$ are dominated by $\|f\|_1/\alpha$. But by a simple calculation, $g \in L^2$ and $\|g\|_2^2 \leq (2^{2n}+1)\alpha\|f\|_1$, and T_K is bounded on L^2, so the estimate for $|\{x : |T_K g(x)| > \alpha\}|$ follows from Chebyshev's inequality. On the other hand, let

$$B_\alpha = \{x : |T_K b(x)| > \alpha\}, \qquad \widetilde{\Omega} = \bigcup 2Q_j,$$

and let y_j and r_j be the center and side length of Q_j. Then $|B_\alpha| \le |\widetilde{\Omega}| + |B_\alpha \setminus \widetilde{\Omega}|$, and $|\widetilde{\Omega}| \le 2^n|\Omega| = 2^n \sum |Q_j| \le 2^n\|f\|_1/\alpha$, so it is enough to estimate $|B_\alpha \setminus \widetilde{\Omega}|$. But for $x \notin \widetilde{\Omega}$ and $y \in Q_j$ we have $|x - y_j| \ge 2r_j \ge 2|y - y_j|$ and (by the mean value theorem, since ∇K is homogeneous of degree $-n-1$)

$$|K(x-y) - K(x-y_j)| \le C|y-y_j||x-y_j|^{-n-1}. \tag{23}$$

Thus, since $\int_{Q_j} b_j = 0$ for all j, for $x \notin \widetilde{\Omega}$ we have

$$\begin{aligned} |Tb(x)| = \Big|\sum Tb_j(x)\Big| &\le \left|\sum \int_{Q_j} [K(x-y) - K(x-y_j)]b_j(y)\,dy\right| \\ &\le C \sum \int_{Q_j} |y-y_j||x-y_j|^{-n-1}|b_j(y)|\,dy, \end{aligned}$$

and hence

$$|B_\alpha \setminus \widetilde{\Omega}| \le \frac{1}{\alpha}\int_{\mathbb{R}^n\setminus\widetilde{\Omega}} |Tb(x)|\,dx \le \frac{C}{\alpha}\sum \int_{\mathbb{R}^n\setminus 2Q_j}\int_{Q_j} |y-y_j||x-y_j|^{-n-1}|b_j(y)|\,dy\,dx.$$

But

$$\int_{\mathbb{R}^n\setminus 2Q_j} |y-y_j||x-y_j|^{-n-1}\,dx \le r_j \int_{|z|>2r_j} |z|^{-n-1}\,dz \le cr_j \cdot r_j^{-1}, \tag{24}$$

so

$$|B_\alpha \setminus \widetilde{\Omega}| \le \frac{C'}{\alpha}\sum \int_{Q_j} |b_j(y)|\,dy \le \frac{C''}{\alpha}\|f\|_1,$$

and we are done.

The great feature of this argument is that it is extremely robust and can easily be generalized, provided one has the L^2 boundedness of T_K as a starting point.

In the first place, one can replace the integral kernels $K(x-y)$ by more general kernels $K(x,y)$. The only property of K needed here is an analogue of (23) or (24):

$$|K(x,y) - K(x,y_0)| \le C|y-y_0||x-y_0|^{-n-1} \quad \text{for} \quad |x-y_0| \ge 2|y-y_0|, \tag{25}$$

$$\int_{|x-y_0|\ge 2|y-y_0|} |K(x,y) - K(x,y_0)|\,dx \le c, \tag{26}$$

together with a similar estimate with the roles of the two arguments of K switched (in order to apply duality to obtain the L^p boundedness for $p > 2$). One could replace the quantity on the right of (25) by $C|y-y_0|^\alpha|x-y_0|^{-n-\alpha}$ for some $\alpha > 0$, which still yields (26).

Moreover, one can replace $\mathbb{R}^n$ equipped with the Euclidean distance $d(x,y) = |x-y|$ and Lebesgue measure by a metric space X equipped with its distance function $d(x,y)$ and a measure μ that has the *doubling property* with respect to d: that is, there is a constant A independent of x and r such that $\mu(B_{2r}(x)) \leq A\mu(B_r(x))$, where $B_r(x) = \{y : d(x,y) < r\}$. (One can even weaken the triangle inequality and assume only that there is a constant B such that $d(x,z) \leq B[d(x,y)+d(y,z)]$.) It might appear that Lemma 20 is special to Euclidean space, but the use of cubes is a convenience rather than a necessity. One can replace the sequence of cubes Q_j by a sequence of balls B_j that have only a controlled amount of overlap so that $\mu(\bigcup B_j)$ is comparable to $\sum \mu(B_j)$; the functions b_j are then supported in sets obtained by "disjointifying" the B_j's. The argument in this general setting was worked out by Coifman and Weiss [10] and Korányi and Vagi [49]; see also Stein [84].

Finally, one can consider operators on vector-valued functions. That is, let $\mathfrak{X}_1$ and $\mathfrak{X}_2$ be Banach spaces. One considers a kernel K taking values in the space of bounded operators from $\mathfrak{X}_1$ to $\mathfrak{X}_2$; then the operator T_K maps $\mathfrak{X}_1$-valued functions to $\mathfrak{X}_2$-valued functions. To obtain the L^2 boundedness of T_K in a straightforward way one may need $\mathfrak{X}_1$ and $\mathfrak{X}_2$ to be Hilbert spaces, but once this is done, the proof of Theorem 19 (with absolute values replaced by norms in appropriate places) yields the boundedness of T_K from $L^p(\mathbb{R}^n, \mathfrak{X}_1)$ to $L^p(\mathbb{R}^n, \mathfrak{X}_2)$ — or a generalization with $\mathbb{R}^n$ replaced by a metric space as above.

The most classical enlargement of the Calderón-Zygmund theory, which came not long after Theorem 19 in [5] and other papers, was to the "variable-coefficient singular integrals"

$$Tf(x) = \lim_{\epsilon\to 0} \int_{|x-y|>\epsilon} K(x, x-y) f(y)\, dy \tag{27}$$

where, for each x, $K(x,\cdot)$ is smooth and homogeneous of degree $-n$ on $\mathbb{R}^n \setminus \{0\}$ and has the mean-zero property, and $K(x,\cdot)$ depends smoothly and boundedly on x in a suitable sense. (We are being deliberately imprecise about the meaning of "smooth." For some purposes minimal smoothness conditions are important; for others it is best to assume C^∞.) The operator (27) can be re-expressed in terms of the Fourier transform as

$$Tf(x) = \int e^{2\pi i\xi\cdot x} \sigma(x,\xi) \widehat{f}(\xi)\, d\xi, \tag{28}$$

where, for each x, $\sigma(x,\cdot)$ is the Fourier transform of $K(x,\cdot)$. As this is homogeneous of degree 0, the boundedness of T on L^2 is easy to establish, assuming only some boundedness of K and hence σ in the first variable.

The representation (28) of singular integral operators leads directly to the theory of pseudo-differential operators, which has become an essential tool in the study of partial differential equations.

To be precise, a (classical) *symbol of order* m ($m \in \mathbb{R}$) is a C^∞ function σ on $\mathbb{R}^{2n}$ such that

$$|\partial_\xi^\alpha \partial_x^\beta \sigma(x,\xi)| \le C_{\alpha\beta}(1+|\xi|)^{m-|\alpha|} \tag{29}$$

for all multi-indices α and β — that is, σ behaves qualitatively like a homogeneous function of ξ of degree m for large ξ and is smooth for small ξ. Examples include smooth functions that are homogeneous of degree m for large ξ (i.e., $\sigma(x,r\xi) = r^m\sigma(x,\xi)$ for $r \ge 1$ and $|\xi| \ge c > 0$) and the symbol $\sigma(x,\xi) = (1+|\xi|^2)^{m/2}$ of the operator Λ^m in (16). A (classical) *pseudo-differential operator of order* m is an operator of the form (28) where σ is a symbol of order m; the common notation for the operator T associated to the symbol σ is $\sigma(x,D)$, where D stands for $(2\pi i)^{-1}(\partial/\partial x)$. Such operators are defined initially on the Schwartz space $\mathcal{S}$, and under our hypothesis (29) on σ they map $\mathcal{S}$ into itself. (One can generalize by requiring the estimates (29) to hold only for x in compact subsets of an open $\Omega \subset \mathbb{R}^n$, with $C_{\alpha\beta}$ depending on the set; then the operator maps $\mathcal{S}$ into $C^\infty(\Omega)$, and one obtains only local L^p estimates.)

The class of pseudo-differential operators includes all partial differential operators with coefficients that are C^∞ and bounded along with all their derivatives (the boundedness being inessential if one only wants local estimates); they are the ones whose symbols $\sigma(x,\xi)$ are polynomials in ξ, thus justifying the notation $\sigma(x,D)$ and the name "pseudo-differential." Pseudo-differential operators of order 0 are a mild generalization of singular integral operators of the form (27) or (28), modified so that σ is smooth near $\xi = 0$. It is easy to see that they are bounded on L^2 and that the associated kernels $K(x,x-y)$ ($K(x,\cdot)$ being the inverse Fourier transform of $\sigma(x,\cdot)$) satisfy estimates of the form (23), so the Calderón-Zygmund theory yields L^p boundedness for $1 < p < \infty$. Moreover, one can convert a pseudo-differential operator of arbitrary order m to a pseudo-differential operator of order 0 by composing with the operator Λ^{-m} defined by (16), and this yields boundedness of pseudo-differential operators as maps between L^p Sobolev spaces.

What makes pseudo-differential operators really useful is the fact that they form a $*$-algebra in which the product and adjoint can easily be calculated in terms of the symbols, up to error terms of lower order. More precisely, if σ_1 and σ_2 are symbols of orders m_1 and m_2, the product $\sigma_1(x,D)\sigma_2(x,D)$ is of the form $\tau(x,D)$ where τ is a symbol of order m_1+m_2; moreover, there is an asymptotic expansion of τ as a series of symbols of order m_1+m_2-j ($j = 0,1,2,\ldots$), constructed from products of derivatives of σ_1 and σ_2, whose leading term is $\sigma_1\sigma_2$. A similar result holds for the adjoint $\sigma(x,D)^*$.

The symbolic calculus of pseudo-differential operators was developed by Joseph J. Kohn and Louis Nirenberg [45], building on earlier results by Calderón and Zygmund and their predecessors, and shortly afterward Lars Hörmander [39] invented the somewhat more general and more convenient form that we have just sketched. See Folland [28] for a quick introduction to pseudo-differential operators and their applications and Taylor [91] for a more comprehensive account.

At this point we can resolve an issue left open in §5, namely, the fact that when k is a positive integer, the Sobolev space L^p_k, defined as $\Lambda^{-k}(L^p)$ where Λ^k is given by (16), coincides with the space $L^{p,k}$ of functions whose distribution derivatives of order $\leq k$ are in L^p; more generally, $f \in L^p_{s+k}$ if and only if $\partial^\alpha f \in L^p_s$ for $0 \leq |\alpha| \leq k$.

By induction, it suffices to consider $k = 1$. The essential point is that the functions $\sigma_j(x,\xi) = \sigma_j(\xi) = 2\pi i \xi_j (1+|\xi|^2)^{-1/2}$ are symbols of order 0, so the associated operators $\partial_j \Lambda^{-1} = \Lambda^{-1}\partial_j$ are bounded on L^p; also, Λ^{-1} itself is of order -1, which is even better, so it is also bounded on L^p. Thus, if $f \in L^p_{s+1}$, then $\Lambda^{s+1} f \in L^p$, hence $\Lambda^s \partial_j f = (\partial_j \Lambda^{-1})\Lambda^{s+1} f \in L^p$ and $\Lambda^s f = \Lambda^{-1}\Lambda^{s+1} f \in L^p$, hence f and $\partial_j f$ are in L^p_s. Conversely, if f and $\partial_j f$ are in L^p_s, then $\Lambda^s f$ and $\Lambda^s \partial_j f$ are in L^p, and hence so is

$$\Lambda^{-1}\Lambda^s f - \frac{1}{4\pi^2}\sum (\Lambda^{-1}\partial_j)(\Lambda^s \partial_j f) = \Lambda^{s-1}\left[I - \frac{1}{4\pi^2}\sum \partial_j^2\right] f = \Lambda^{s-1}\Lambda^2 f = \Lambda^{s+1} f,$$

so $f \in L^p_{s+1}$.

8. H^p Spaces: the Real-Variable Theory

The theory of H^p spaces on the disc has an obvious analogue on the upper half-plane $\mathbb{U} = \{(x,y) \in \mathbb{R}^2 : y > 0\}$. To wit, one has the Hardy spaces

$$H^p(\mathbb{U}) = \left\{ F \text{ holomorphic on } \mathbb{U} : \sup_{y>0} \int_{\mathbb{R}} |F(x+iy)|^p \, dx < \infty \right\} \qquad (0 < p < \infty),$$

and many results on $H^p(\mathbb{D})$ carry over to $H^p(\mathbb{U})$ with obvious modifications. For example, nontangential convergence works much the same way, a nontangential approach region of a point $(x_0, 0)$ on the real axis now being simply a cone $\{(x,y) : |x - x_0| < cy\}$. Moreover, if $f \in L^p(\mathbb{R})$ $(1 \leq p \leq \infty)$, f has a harmonic extension u to $\mathbb{U}$ such that $u(\cdot, y) \in L^p$ for each $y > 0$, given by the Poisson integral.

This last fact generalizes immediately to functions on $\mathbb{R}^n$: the Poisson integral of an $f \in L^p(\mathbb{R}^n)$ is a harmonic function on $\mathbb{R}^{n+1}_+$. However, as in §1, we denote the extra coordinate on $\mathbb{R}^{n+1}_+$ by t rather than y and write it first; that is, $\mathbb{R}^{n+1}_+$ is taken to consist of points (t,x) with $t > 0$ and $x \in \mathbb{R}^n$. For future reference we note that the Poisson kernel P_t defined by (5) satisfies $\int P_t = \|P_t\|_1 = 1$ for all t, so that $\|f * P_t\|_p \leq \|f\|_p$.

One possibility for generalizing H^p spaces to higher dimensions is to study functions of several complex variables on appropriate domains in $\mathbb{C}^n$; see, for example, Stein-Weiss [87, Chapter III]. Our concern here is with a different one that focuses instead on harmonic functions that satisfy appropriate analogues of the Cauchy-Riemann equations. To achieve the right orientation, recall that $F = u + iv$ is holomorphic on a domain in $\mathbb{C}$ if and only if (u, v) satisfies $u_x = v_y$ and $v_x + u_y = 0$. The first of these equations says that (v, u) is the gradient of a function U, and the second one then says that U is harmonic.

With this in mind, in 1960 Stein and Guido Weiss [86] considered $(n + 1)$-tuples $(u_0, \dots, u_n)$ of harmonic functions on $\mathbb{R}^{n+1}_+$ that satisfy the generalized Cauchy-Riemann equations

$$\frac{\partial u_j}{\partial x_k} = \frac{\partial u_k}{\partial x_j} \text{ for } 0 \le j, k \le n, \qquad \frac{\partial u_0}{\partial x_0} + \cdots + \frac{\partial u_n}{\partial x_n} = 0, \tag{30}$$

where we have written the coordinates on $\mathbb{R}^{n+1}_+$ as $(x_0, \dots, x_n)$ instead of $(t, x) = (t, x_1, \dots, x_n)$. Again, (30) is equivalent to the condition $(u_0, \dots, u_n) = \nabla U$ where U is harmonic. Stein and Weiss defined a Hardy space that we shall denote by $H^p_{\text{harm}}(\mathbb{R}^{n+1}_+)$ for $(n - 1)/n < p < \infty$ to be the set of all $(u_0, \dots, u_n)$ that satisfy (30) and

$$\sup_{t>0} \int_{\mathbb{R}^n} \left(|u_0(t, x)|^2 + \cdots + |u_n(t, x)|^2\right)^{p/2} dx < \infty. \tag{31}$$

The restriction $p > (n - 1)/n$ is necessary to obtain a satisfactory theory; in a nutshell, the reason is that this is the range of p for which the integrand in (31) is subharmonic. The theory can be extended to all $p > 0$ by considering systems of harmonic functions satisfying more complicated generalizations of the Cauchy-Riemann equations. We shall refer briefly to the resulting H^p spaces in the sequel without going into any detail; see [87, Chapter VI] for more information.

If $(u_0, \dots, u_n) \in H^p_{\text{harm}}(\mathbb{R}^{n+1}_+)$, as $t \to 0$ $u_j(t, \cdot)$ converges in the topology of tempered distributions to an $f_j \in \mathcal{S}'(\mathbb{R}^n)$ whose behavior at infinity is sufficiently tame that it can be convolved wtih the Poisson kernel, and u_j can be recovered from f_j by the Poisson integral: $u_j(t, \cdot) = P_t * f_j$. Moreover, as in the case $n = 1$, $u_1, \cdots, u_n$ (or $f_1, \cdots, f_n$) are completely determined by u_0 (or f_0) via the most straightforward generalization of the Hilbert transform, the *Riesz transforms*:[3]

$$[R_j f]\widehat{\ }(\xi) = \frac{\xi_j}{i|\xi|} \widehat{f}(\xi.) \tag{32}$$

This is easily seen formally: if $(u_0, \dots, u_n) = \nabla U$ and $U(t, \cdot) = P_t * f$, then by (6), $\widehat{U}(t, \xi) = e^{-2\pi t|\xi|} \widehat{f}(\xi)$, so $\widehat{u}_0(t, \xi) = -2\pi|\xi| e^{-2\pi t|\xi|} \widehat{f}(\xi)$ and $\widehat{u}_j(t, \xi) = 2\pi i \xi_j e^{-2\pi t|\xi|} \widehat{f}(\xi)$ for $j > 0$.

[3] The name "Riesz transform" was introduced by Stein and Weiss.

Hence, for $j > 0$, $u_j(t, \cdot) = R_j u_0(t, \cdot)$, and in the limit as $t \to 0$, $f_j = R_j f_0$. (Making this into a rigorous argument is straightforward.) In short, for $p > (n-1)/n$, $H^p_{\text{harm}}(\mathbb{R}^{n+1}_+)$ can be identified via the correspondence $(u_0, \ldots, u_n) \leftrightarrow f_0$ with a space of distributions on $\mathbb{R}^n$ that we denote by $H^p(\mathbb{R}^n)$:

$$H^p(\mathbb{R}^n) = \big\{ f_0 \in \mathcal{S}'(\mathbb{R}^n) : (u_0, \ldots, u_n) \in H^p_{\text{harm}}(\mathbb{R}^{n+1}_+), \text{ where } u_0(t, \cdot) = P_t * f_0 \text{ and } u_j(t, \cdot) = R_j u_0(t, \cdot) \text{ for } j > 0 \big\}. \tag{33}$$

For $0 < p \leq (n-1)/n$, $H^p(\mathbb{R}^n)$ may be defined as a space of distributions in a similar but more complicated way.

When $p \geq 1$ we can say more. The Riesz transforms are Calderón-Zygmund singular integral operators, so they are bounded on L^p for $p > 1$; hence, in this case, $H^p(\mathbb{R}^n)$ is nothing but $L^p(\mathbb{R}^n)$. For $p = 1$ there is an analogue of the F. and M. Riesz theorem: if $(u_0, \ldots, u_n) \in H^1_{\text{harm}}(\mathbb{R}^{n+1}_+)$, then $u_0, \ldots u_n$ are the Poisson integrals of functions $f_1, \ldots, f_n \in L^1(\mathbb{R}^n)$. Hence,

$$H^1(\mathbb{R}^n) = \big\{ f \in L^1(\mathbb{R}^n) : R_j f \in L^1(\mathbb{R}^n) \text{ for } j = 1, \ldots, n \big\}. \tag{34}$$

In both these cases the convergence of $u_j(t, \cdot)$ to its boundary function f_j takes place in the L^p norm and pointwise a.e. (For a single harmonic function u on $\mathbb{R}^{n+1}_+$ that satisfies $\sup_{t>0} \int |u(t,x)|^p \, dx < \infty$, u is the Poisson integral of an L^p function if $p > 1$, but in general it is only the Poisson integral of a measure if $p = 1$.)

The next ingredient in H^p theory comes from a different direction, a 1961 paper of Fritz John and Louis Nirenberg [41] that put the spotlight on functions of "bounded mean oscillation." With notation and terminology concerning cubes and mean values as in §7, we define

$$\text{BMO}(\mathbb{R}^n) = \left\{ f \in L^1_{\text{loc}}(\mathbb{R}^n) : \|f\|_{\text{BMO}} \equiv \sup_Q m_Q\big(|f - m_Q f|\big) < \infty \right\}, \tag{35}$$

the supremum being taken over all cubes in $\mathbb{R}^n$. (Note that $L^\infty(\mathbb{R}^n) \subset \text{BMO}(\mathbb{R}^n)$). If instead one fixes a cube Q_0, takes $f \in L^1(Q_0)$, and takes the supremum only over cubes $Q \subset Q_0$, one obtains the space $\text{BMO}(Q_0)$. John and Nirenberg's main result is that there are constants $B, b > 0$ depending only on n such that if $f \in \text{BMO}(Q_0)$, then

$$\big|\{x : |f(x) - m_{Q_0} f| > \alpha\}\big| \leq B \exp(-b\alpha / \|f\|_{\text{BMO}(Q_0)}),$$

which implies in particular that $f \in L^p(Q_0)$ for all $p < \infty$. This result was immediately applied by John and by Jürgen Moser to problems in partial differential equations.

For our purposes, however, the important point is that $\mathrm{BMO}(\mathbb{R}^n)$ becomes a Banach space with norm $\|\cdot\|_{\mathrm{BMO}}$ after one identifies functions that are equal a.e. and functions that differ by a constant, and we have Charles Fefferman's duality theorem (announced in [20], proved in detail in [22]):

***Theorem* 21** *(C. Fefferman,* 1971*) —* $\mathrm{BMO}(\mathbb{R}^n)$ *is the dual space of* $H^1(\mathbb{R}^n)$*.*

The meaning of this is as follows. First note that if $f \in H^1(\mathbb{R}^n)$ then $\int f = 0$, for by (35), $(R_j f)\widehat{\ }(\xi)$ must be continuous, but by (32), this cannot be true at $\xi = 0$ unless $\widehat{f}(0) = \int f = 0$. If f is C^∞ with compact support and $\int f = 0$, then $f \in H^1(\mathbb{R}^n)$, and the set of all such f is dense in $H^1(\mathbb{R}^n)$. If $\phi \in \mathrm{BMO}(\mathbb{R}^n)$, moreover, for such f the integral $\int f\phi$ is well defined and unaffected if a constant is added to ϕ. Fefferman's theorem says that (i) the map $f \mapsto \int f\phi$ extends to a bounded linear functional on $H^1(\mathbb{R}^n)$, (ii) the norm of this functional (taking the norm on $H^1(\mathbb{R}^n)$ to be $\|f\|_1 + \sum \|R_j f\|_1$) is equivalent to $\|\phi\|_{\mathrm{BMO}}$, and (iii) every such functional arises in this way.

The final liberation of H^p spaces from analytic function theory came from their characterization in terms of maximal functions. The first step was Hardy and Littlewood's result (Theorem 12(c) in this paper) that the nontangential maximal function of an $F \in H^p(\mathbb{D})$ is in L^p for all $p > 0$; the analogous result for $H^p(\mathbb{U})$ is also valid. Forty years elapsed before Donald Burkholder, Richard Gundy, and Martin Silverstein [2] completed this picture by proving the converse, using ideas from probability theory:

***Theorem* 22** *(Burkholder-Gundy-Silverstein,* 1971*) — If* u *is a real-valued harmonic function on* $\mathbb{D}$ *or* $\mathbb{U}$ *whose nontangential maximal function* u^* *is in* L^p *(*$0 < p < \infty$*), then* u *is the real part of a holomorphic function in* $H^p(\mathbb{D})$ *or* $H^p(\mathbb{U})$*.*

Here u^* is defined by (15) with $c = 1$ for functions on $\mathbb{D}$; for functions on $\mathbb{R}^{n+1}_+$ (recalling that $\mathbb{R}^2_+$ is $\mathbb{U}$ with the variables switched),

$$u^*(x) = \sup_{|y-x|<t<\infty} |u(t,y)|. \tag{36}$$

Theorem 22 – and the following Theorem 23 — are still valid if u^* is replaced by u^*_c, defined by (36) with the condition $|y - x| < t$ replaced by $|y - x| < ct$. We take $c = 1$ for simplicity.

With Theorem 22 in hand, it did not take Fefferman and Stein [22] long to establish the analogous result in n variables, and to go even further by showing that the Poisson integral could be replaced by general smooth approximate identities. To explain this, we need some notation.

First, if $\phi \in L^1(\mathbb{R}^n)$, we set

$$\phi_t(x) = t^{-n}\phi(t^{-1}x) \qquad (t > 0.)$$

(Thus the mass of ϕ_t concentrates at the origin as $t \to 0$, and the factor t^{-n} makes $\int \phi_t$ independent of t. Observe that this notation is consistent with the definition (4) of the Poisson kernel P_t if we take $P = P_1$.) We recall that if $\int \phi = 1$, then as $t \to 0$, $\phi_t * f \to f$ in the L^p norm and pointwise a.e. for any $f \in L^p$; moreover, if $\phi \in \mathcal{S}$, then $\phi_t * f \to f$ in the topology of $\mathcal{S}$ or $\mathcal{S}'$ for any f in $\mathcal{S}$ or $\mathcal{S}'$. For $\phi \in \mathcal{S}$ and $f \in \mathcal{S}'$, we introduce the following radial and nontangential maximal functions:

$$M_\phi^+ f(x) = \sup_{t>0} |\phi_t * f(x)|, \qquad M_\phi^* f(x) = \sup_{|y-x|<t<\infty} |\phi_t * f(y)|.$$

(Again, the condition $|y-x| < t$ could be replaced by $|y-x| < ct$, $c > 0$.)

Second, we introduce an enlargement of $\mathcal{S}$ that includes the Poisson kernel,

$$\mathcal{P} = \big\{\phi \in C^\infty(\mathbb{R}^n) : |\partial^\alpha \phi(x)| \le C_\alpha (1+|x|)^{-n-1-|\alpha|} \text{ for all multi-indices } \alpha\big\},$$

and a certain bounded subset of $\mathcal{S}$,

$$\mathcal{A} = \big\{\phi \in \mathcal{S} : |\phi(x)| \le (1+|x|)^{-n-1} \text{ and } |\nabla\phi(x)| \le (1+|x|)^{-n-1}\big\}$$

(note that there are no unspecified constants in these inequalities).

***Theorem* 23** *(Fefferman-Stein,* 1972*) — Suppose $(n-1)/n < p < \infty$. For $f \in \mathcal{S}'$, the following are equivalent:*

(a) $M_\phi^+ f \in L^p$ for some $\phi \in \mathcal{S}$ with $\int \phi = 1$.

(b) $M_\phi^* f \in L^p$ for some $\phi \in \mathcal{S}$ with $\int \phi = 1$.

(c) $M_\phi^* \in L^p$ for all $\phi \in \mathcal{S}$, and in fact $\sup_{\phi\in\mathcal{A}} M_\phi^* f \in L^p$.

(d) f extends to a bounded functional on $\mathcal{P}$, so that its Poisson integral $u(t,\cdot) = P_t * f$ is well defined, and $u^* \in L^p$.

(e) $f \in H^p(\mathbb{R}^n)$ as defined in (33).

These equivalences remain valid for all $p > 0$, except that in (c) $\mathcal{A}$ must be redefined, depending on the value of p, to involve stronger bounds on ϕ and its derivatives up to a certain order, and in (e) $H^p(\mathbb{R}^n)$ must be redefined as indicated after (33).

With this theorem in hand, the real-variable H^p space is now understood to be the space of all tempered distributions on $\mathbb{R}^n$ that satisfy the equivalent conditions (a)-(c), which no longer have anything to do with harmonic or holomorphic functions. Instead, it is a significant extension of the class of L^p spaces from the range $1 < p < \infty$ the full range $0 < p < \infty$, including a modification of

L^1 that turns out to have better properties than L^1 itself in many respects. For one thing, as Fefferman and Stein showed in [22], singular integral operators of the sort discussed in Theorem 19 are bounded on H^1, and hence they also extend to bounded operators on BMO; this is a useful substitute for L^p boundedness in these two exceptional cases.

There is another important real-variable characterization of $H^p(\mathbb{R}^n)$ for $p \leq 1$, the "atomic decomposition". A p-*atom* is a bounded function a that is supported in a cube Q and satisfies (i) $\|a\|_\infty \leq |Q|^{-1/p}$ and (ii) $\int x^\alpha a(x)\,dx = 0$ for all multi-indices α with $|\alpha| \leq n(p^{-1}-1)$. It is easy to verify that p-atoms belong to H^p; more generally, if $f = \sum c_j a_j$ (convergence in $\mathcal{S}'$) where the a_j's are p-atoms and the c_j's are positive numbers with $\sum c_j^p < \infty$, then $f \in H^p$. Conversely, every $f \in H^p$ can be represented this way. The latter result, due to R. Coifman [9] when $n = 1$ and to R. H. Latter [50] in general, takes some work to prove, using a variant of the decomposition $f = g + \sum b_j$ in the proof of Theorem 19.

See [84] for more about real-variable H^p spaces on $\mathbb{R}^n$. Like the theory of singular integrals, this theory has been extended from the original setting of $\mathbb{R}^n$ with its standard translations and dilations to a number of other situations. See Coifman-Weiss [11], Folland-Stein [32], and the references given in these works.

9. Singular Integrals Revisited

As we have pointed out, the Calderón-Zygmund machine yields L^p boundedness of a large collection of singular integral operators in a wide variety of contexts, provided that L^2 boundedness is known to begin with. In the original setting and its immediate generalizations, the Fourier transform is the essential tool for reaching that starting point. In other situations, however, the Fourier transform may be unavailable or ineffective, and one needs to find other methods.

The single most powerful tool for this purpose is a functional-analytic proposition that was proved independently by Stein and Mischa Cotlar in the late 1960s (see [43]). It concerns estimates for a sum $\sum T_j$ of bounded operators on a Hilbert space $\mathcal{H}$, given that $\|T_j\| \leq M < \infty$ for all j. Of course, all one can say in general is that $\|\sum_1^n T_j\| \leq nM$. But suppose there are two sequences of mutually orthogonal subspaces of $\mathcal{H}$, $\{\mathfrak{X}_j\}$ and $\{\mathcal{Y}_j\}$, such that T_j maps $\mathfrak{X}_j$ into $\mathcal{Y}_j$ and $T_j = 0$ on $\mathfrak{X}_j^\perp$. Then, denoting the orthogonal projection of $x \in \mathcal{H}$ onto $\mathfrak{X}_j$ by x_j and applying the Pythagorean theorem first on $\bigoplus \mathcal{Y}_j$ and then on $\bigoplus \mathfrak{X}_j$ we see that

$$\Big\|\sum T_j x\Big\|^2 = \Big\|\sum T_j x_j\Big\|^2 = \sum \|T_j x_j\|^2 \leq M^2 \sum \|x_j\|^2 \leq M^2 \|x\|^2,$$

so $\|\sum T_j\| \leq M$. The above conditions on the T_j's are equivalent to the conditions $T_i^* T_j = 0$

$= T_i T_j^*$ for all i, j (verifying this is an amusing exercise), and the idea of the Cotlar-Stein lemma is that the conclusion remains true (with a small modification) if these equations are only approximately valid.

Lemma 24 (Cotlar, Stein; 1969) — Suppose that $\{T_j\}_1^\infty$ is a sequence of bounded operators on a Hilbert space $\mathcal{H}$ and there is a function $\phi : \mathbb{Z} \to (0, \infty)$ such that (i) $\|T_i^* T_j\| \leq \phi(i-j)^2$ and $\|T_i T_j^*\| \leq \phi(i-j)^2$ for all i, j (in particular, $\|T_j\| = \|T_j^* T_j\|^{1/2} \leq \phi(0)$ for all j) and (ii) $\sum_{-\infty}^{\infty} \phi(k) = A < \infty$. Then $\|\sum_1^n T_j\| \leq A$ for all n.

Stein's proof of this result, which originally appeared in Knapp-Stein [44] and may also be found in [27] and [84] among other places, is a delight; the reader who is not familiar with it will be well rewarded by spending a few minutes learning it.

The Cotlar-Stein lemma is typically applied to integral operators by breaking the integral kernel up into a sum of small pieces. For example, one can prove the L^2 boundedness of the operators T_K in Theorem 19 by writing $K = \sum_{-\infty}^{\infty} K_j$, where $K_j(x) = K(x)$ if $2^j \leq |x| < 2^{j+1}$ and $K_j(x) = 0$ otherwise. In fact, the first application of the lemma was to an analogue of these operators in a noncommutative setting, as we shall discuss in §11.

Another early application was to the L^2 theory of generalized pseudo-differential operators. We recall that classical pseudo-differential operators are operators of the form (28) where the symbol σ satisfies (29). More generally, for $m \in \mathbb{R}$ and $0 \leq \delta \leq \rho \leq 1$ one defines the symbol class $S^m_{\rho,\delta}$ to be the set of C^∞ functions σ on $\mathbb{R}^n \times \mathbb{R}^n$ that satisfy

$$|\partial_\xi^\alpha \partial_x^\beta \sigma(x, \xi)| \leq C_{\alpha,\beta}(1 + |\xi|)^{m - \rho|\alpha| + \delta|\beta|}. \tag{37}$$

When $\rho > \delta$, the L^2 theory works much the same way as in the classical case $\rho = 1$, $\delta = 0$, but there are interesting operators with symbols in the borderline classes with $\rho = \delta$. Calderón and Rémi Vaillancourt [3] used the Cotlar-Stein lemma (or rather an easy generalization of it where the sum $\sum T_j$ is replaced by an integral $\int T_\lambda d\mu(\lambda)$) to prove:

***Theorem* 25** *(Calderón-Vaillancourt,* 1972*) — If T is given by (28) with symbol $\sigma \in S^0_{\rho,\rho}$ ($0 \leq \rho < 1$), then T is bounded on L^2.*

Somewhat different proofs, which apply the lemma in its original form by using a partition of unity to break the symbol $\sigma(x, \xi)$ into a sum of symbols supported in the regions where $2^{j-1} < |\xi| < 2^{j+2}$, can be found in [27] and [84].

Operators with symbols in $S^0_{\rho,\delta}$ have integral kernels of Calderón-Zygmund type — that is, satisfying estimates of the form (25) — only when $\rho = 1$; for $\rho < 1$ they are generally not bounded on

L^p for $p \neq 2$. On the other hand, operators with symbols in $S^0_{1,\delta}$ do have Calderón-Zygmund kernels, even for the extreme case $\delta = 1$; but in that extreme case they are generally not bounded on L^2 (and hence the L^p theory also breaks down). See [84, Chapter VII].

In 1984 Guy David and Jean-Lin Journé [15] achieved what one might call the apotheosis of Calderón-Zygmund theory with their "$T(1)$ theorem." To state it we need some terminology. Suppose T is a continuous linear map from $\mathcal{S}$ to $\mathcal{S}'$; thus $\langle Tu, v\rangle$ is well defined for $u, v \in \mathcal{S}$, where $\langle\cdot,\cdot\rangle$ denotes the pairing of $\mathcal{S}'$ with $\mathcal{S}$. We say that T is *weakly bounded* if there is a constant $C > 0$ and an integer $K \geq 0$ such that, for all $r > 0$, $|\langle Tu, v\rangle| \leq Cr^n$ whenever u and v are supported in a cube of side length r and $|\partial^\alpha u|, |\partial^\alpha v| \leq r^{-|\alpha|}$ for $|\alpha| \leq K$. (This condition on u and v is invariant under dilations.)

Three important observations: First, if T is bounded on L^2 then T is weakly bounded, with $C = \|T\|_{L^2\to L^2}$ and $K = 0$; but weak boundedness is a much weaker and more easily verified condition. Second, if T is weakly bounded, T extends continuously to a map from C^K_c (the space of C^K functions of compact support, with K as in the preceding definition) to its dual. Third, if T is weakly bounded, then so is its adjoint T^* defined by $\langle T^*u, v\rangle = \langle Tv, u\rangle$.

Next, a *standard kernel* is a continuous function K on $\{(x, y) \in \mathbb{R}^n \times \mathbb{R}^n : x \neq y\}$ that satisfies, for some $C, \delta > 0$,

$$\begin{gathered} |K(x,y)| \leq C|x-y|^{-n}, \\ |K(x,y) - K(x',y)| + |K(y,x) - K(y,x')| \leq C\frac{|x-x'|^\delta}{|x-y|^{n+\delta}} \text{ when } |x-x'| < \tfrac{1}{2}|x-y|. \end{gathered} \tag{38}$$

The map $T : \mathcal{S} \to \mathcal{S}'$ is said to be *associated to* the kernel K if, for any $u \in \mathcal{S}$, the distribution Tu agrees with the function $x \mapsto \int K(x, y)u(y)\,dy$ on the complement of the support of u. Note that in this case, the adjoint T^* is associated to the kernel $K^*(x, y) = K(y, x)$, which is also standard.

We shall call an operator $T : \mathcal{S} \to \mathcal{S}'$ that is weakly bounded and associated to a standard kernel a *weak Calderón-Zygmund operator*. If, in addition, T is bounded on L^2, and hence bounded on L^p for $1 < p < \infty$, we shall call T a *Calderón-Zygmund operator*.

If T is a weak Calderón-Zygmund operator, there is a natural way to define Tf for any bounded C^∞ function f as a linear functional on the space $\mathfrak{X} = \{u \in C^\infty_c : \int u = 0\}$. Indeed, given such a u supported in $\{x : |x| \leq R\}$, pick $\phi \in C^\infty_c$ with $\phi(x) = 1$ for $|x| \leq 2R$ and write $f_1 = \phi f$, $f_2 = (1 - \phi)f$. Then $f_1 \in \mathcal{S}$, so $\langle Tf_1, u\rangle$ is well defined. On the other hand, since $\int u = 0$ we have

$\int K(x,y)u(x)\,dx = \int [K(x,y) - K(0,y)]u(x)\,dx$ for $|y| > 2R$, so we may define

$$\langle Tf_2, u\rangle = \iint [K(x,y) - K(0,y)]f_2(y)u(x)\,dy\,dx,$$

the latter integral being absolutely convergent in view of (38). We then set $Tf = Tf_1 + Tf_2$; it is easy to check that this is independent of the choice of ϕ. Finally, noting that $\mathfrak{X}$ is dense in $H^1(\mathbb{R}^n)$, we say that $Tf \in \mathrm{BMO}$ if the functional Tf extends to a bounded functional on $H^1(\mathbb{R}^n)$.

***Theorem* 26** *(David-Journé,* 1984*) — Suppose $T : \mathcal{S} \to \mathcal{S}'$ is a weak Calderón-Zygmund operator. Then T is a Caldeón-Zygmund operator if and only if* $T(1) \in \mathrm{BMO}$ *and* $T^*(1) \in \mathrm{BMO}$.

The proof in [15] (see also [84]) consists of using some auxiliary operators to reduce to the case where $T(1) = T^*(1) = 0$ and then using the Cotlar-Stein lemma. There is now an alternative proof of both of these steps using wavelets, which we shall sketch in §12.

As a simple application, consider a pseudo-differential operator T whose symbol σ belongs to the class $S^0_{1,1}$ as in (37). As we noted earlier, T is associated to a standard kernel, and it is easily seen to be weakly bounded. Moreover, one can read off $T(1)$ directly from (28):

$T(1)(x) = \int e^{2\pi i\xi\cdot x}\sigma(x,\xi)\delta(\xi)\,d\xi = \sigma(x,0)$, which is a bounded function. Hence:

Corollary 27 — Suppose T is a pseudo-differential operator with symbol in $S^0_{1,1}$.

(a) T is bounded on L^2 if and only if $T^*(1) \in \mathrm{BMO}$.

(b) If T^* is also a pseudo-differential operator with symbol in $S^0_{1,1}$, then T is bounded on L^2.

The theory of singular integrals has been further extended in a number of significant directions, motivated by problems in partial differential equations and other areas of analysis; here we can only sketch a couple of the main ideas.

First, one can consider operators that resemble Calderón-Zygmund singular integrals but involve integration over lower-dimensional manifolds. The first such operators to receive intensive study — by Stein, Stephen Wainger, and others in the 1970s; see the survey paper [85] — were the *Hilbert transforms on curves*:

$$Tf(x) = \int_{-1}^{1} f(x - \gamma(t))\frac{dt}{t},$$

where f is a function on $\mathbb{R}^n$ and $\gamma : [-1,1] \to \mathbb{R}^n$ is a smooth curve with $\gamma(0) = 0$. The operators involving integration over higher-dimensional manifolds are known as *singular Radon transforms*: they have the form

$$Tf(x) = \int_{M_x} K(x,y)f(y)d\sigma(y),$$

where M_x is a smooth k-dimensional submanifold of $\mathbb{R}^n$ containing x and varying smoothly in x, $K(x,\cdot)$ is a kernel of Calderón-Zygmund type on M_x with singularity at x, and σ is surface measure on M_x. The validity (or not) of L^p estimates for these operators turns out to depend strongly on curvature properties of the curve γ or the manifolds M_x. This theory was developed for the hypersurface case ($k = n-1$) by D. H. Phong and Stein [65, 66], and in general by Michael Christ, Alexander Nagel, Stein, and Wainger [8]. We note that the proofs of the L^2 estimates in [65] and [66] use the Calderón-Vaillancourt theorem and the $T(1)$ theorem, respectively; the proof in [8] uses the Cotlar-Stein lemma directly.

Second, one can consider "multi-parameter" singular integrals, where the underlying geometry involves not the usual dilations $x \mapsto rx$ on $\mathbb{R}^n$ but a multi-parameter family of dilations. The most basic examples are convolution operators $Tf = f * K$ on $\mathbb{R}^n \times \mathbb{R}^m$ where $K(x,y)$ is homogeneous (or approximately homogeneous) of degree $-n$ in x and of degree $-m$ in y and has an appropriate "mean-zero" property. Note that this entails K being singular not just at the origin but on the whole subspaces $x = 0$ and $y = 0$. When $K(x,y) = K_1(x)K_2(y)$ where K_1 and K_2 are Calderón-Zygmund kernels on $\mathbb{R}^n$ and $\mathbb{R}^m$, L^p estimates follow easily from the classical theory, but the general case requires other techniques. In this setting the theory is due to Robert Fefferman and Stein [23], and a number of generalizations and variants have been developed since, including "multi-parameter singular Radon transforms"; see Street [89] for a comprehensive account as well as references to the literature.

10. Littlewood-Paley Theory

We now return to the theorems of Littlewood and Paley that we alluded to in §4. They were announced in 1931 in [53], but the full proofs did not appear until 1936 in [54]. (Paley died in a skiing accident in 1933, at the age of 26.) The main goal is the following theorem, which Littlewood and Paley gave in a slightly more general form than we shall do. Given a function f on $\mathbb{T}$ with Fourier series $\sum_{-\infty}^{\infty} c_k e^{ik\theta}$, let

$$\Delta_l(\theta) = \sum_{2^{l-1} < |k| \le 2^l} c_k e^{ik\theta}, \tag{39}$$

so that $\sum_{-\infty}^{\infty} c_k e^{ik\theta} = c_0 + \sum_0^{\infty} \Delta_l$. For simplicity in stating the results below, we shall assume that $c_0 \ (= \int f) \ = 0$.

Theorem 28 *(Littlewood-Paley,* 1931*) — For $1 < p < \infty$ there are constants $A_p, B_p > 0$ such that for all $f \in L^p(\mathbb{T})$ with $\int f = 0$,*

$$A_p \|f\|_p \le \left\| \left(\sum |\Delta_l|^2 \right)^{1/2} \right\|_p \le B_p \|f\|_p.$$

Corollary 29 — If $1 < p < \infty$, $f \in L^p(\mathbb{T})$, and $\int f = 0$, then for any sequence $\{\epsilon_l\}$ with $\epsilon_l \in \{-1, 1\}$ the series $\sum \epsilon_l \Delta_l$ converges in $L^p(\mathbb{T})$, and its L^p norm is comparable to $\|f\|_p$.

These results are of considerable interest in their own right, but what has developed into a widely applicable tool is an auxiliary function that Littlewood and Paley used in proving them, the "g-function." This, in its original form, pertains to Fourier series on $\mathbb{T}$ of power-series type — that is, $f(\theta) = \sum_1^\infty c_k e^{ik\theta}$ — and their holomorphic extensions $F(z) = \sum_1^\infty c_k z^k$. For such an f, we define

$$g_f(\theta) = \left(\int_0^1 (1-r)|F'(re^{i\theta})|^2 \, dr \right)^{1/2}, \tag{40}$$

and we have

***Theorem* 30** *(Littlewood-Paley,* 1931*) — For $1 < p < \infty$ there are constants $A_p, B_p > 0$ such that $A_p\|f\|_p \le \|g_f\|_p \le B_p\|f\|_p$ for all $f \in L^p(\mathbb{T})$ whose Fourier coefficients c_k vanish for $k \le 0$.*

(The assumption $c_0 = 0$ is clearly appropriate here since c_0 disappears in passing from F to F'.)

As explained in [53], there is a simple heuristic argument that leads from the function $(\sum |\Delta_l|^2)^{1/2}$ in Theorem 28 to the function g_f. The basic intuition is that the partial sum $\sum_1^n a_k$ of a series behaves like the Abel mean $\sum_1^\infty a_k r^k$ with $r = 1 - (1/n)$. Granting this, since Δ_l is the difference between the 2^lth and the 2^{l-1}th partial sums of the Fourier series of f, one has

$$\Delta_l(\theta) \approx F((1-2^{-l})e^{i\theta}) - F((1-2^{1-l})e^{i\theta}) = \int_{1-2^{1-l}}^{1-2^{-l}} F'(re^{i\theta}) \, dr,$$

and hence (since $(1-2^{-l}) - (1-2^{1-l}) = 2^{-l}$)

$$\begin{aligned} \sum |\Delta_l|^2 &\approx \sum \left| \int_{1-2^{1-l}}^{1-2^{-l}} F'(re^{i\theta}) \, dr \right|^2 \le \sum 2^{-l} \int_{1-2^{1-l}}^{1-2^{-l}} |F'(re^{i\theta})|^2 \, dr \\ &\approx \sum \int_{1-2^{1-l}}^{1-2^{-l}} (1-r)|F'(re^{i\theta})|^2 \, dr = \int_0^1 (1-r)|F'(re^{i\theta})|^2 \, dr = g_f(\theta)^2. \end{aligned}$$

The actual deduction in [54] of Theorem 28 from Theorem 30, however, is nowhere near as simple as this heuristic argument would suggest. It is quite convoluted and involves a partial analogue of Theorem 30 for a more complicated version (called g_f^*) of the g-function.

As with the theory of H^p spaces, it was natural to seek analogues of Theorem 30 for functions on $\mathbb{R}^n$ rather than $\mathbb{T}$, using harmonic functions on $\mathbb{R}^{n+1}_+$. This theory was first developed in a 1958 paper of Stein [80], with some extensions and refinements afterward. The most straightforward analogue of the g-function in this setting is the following: given $f \in L^p(\mathbb{R}^n)$ $(1 \le p \le \infty)$, let $u(t, \cdot) = P_t * f$

be its Poisson integral as in (5), and set

$$g_f(x) = \left(\int_0^\infty |\nabla_{t,x} u(x,t)|^2 t\, dt\right)^{1/2}. \tag{41}$$

It will also be of interest to consider variants of this expression involving only t-derivatives of u:

$$g_f^k(x) = \left(\int_0^\infty \left|\frac{\partial^k u}{\partial t^k}(x,t)\right|^2 t^{2k-1}\, dt\right)^{1/2} \qquad (k = 1, 2, 3, \ldots) \tag{42}$$

Observe that g_f^1 differs from g_f only in the substitution of $\partial u/\partial t$ for $\nabla_{t,x} u$, and that $(g_f)^2 = (g_f^1)^2 + (\widetilde{g}_f)^2$ where $\widetilde{g}_f$ is defined by (41) with $\nabla_{t,x}$ replaced by ∇_x. Here is the analogue of Theorem 30:

***Theorem* 31** — *Let G_f denote any of the functions g_f or g_f^k ($k \geq 1$). For $1 < p < \infty$ there are constants $A_p, B_p > 0$ such that $A_p \|f\|_p \leq \|G_f\|_p \leq B_p \|f\|_p$ for all $f \in L^p(\mathbb{R}^n)$. What is more, each G_f is an isometry on L^2 up to a constant factor.*

The Calderón-Zygmund technology yields a proof that is easy enough to be sketched here; see [81] for details. First, the L^2 result is an immediate corollary of the Plancherel theorem. Indeed, since $u(t,\cdot)\widehat{\ }(\xi) = e^{-2\pi t|\xi|}\widehat{f}(\xi)$ by (6), we have

$$\int \left|\frac{\partial u}{\partial t}(t,x)\right|^2 dx = \int |\nabla_x u(t,x)|^2\, dx = \int 4\pi^2 |\xi|^2 e^{-4\pi t|\xi|} |\widehat{f}(\xi)|^2\, d\xi,$$

so since $\int_0^\infty t e^{-4\pi t|\xi|}\, dt = 1/16\pi^2|\xi|^2$, Fubini's theorem yields $\|g_f^1\|_2^2 = \|\widetilde{g}_f\|_2^2 = \frac{1}{4}\|f\|_2^2$ and hence $\|g_f\|_2^2 = \frac{1}{2}\|f\|_2^2$. A similar calculation shows that $\|g_f^k\|_2^2 = [(2k-1)!/4^k]\|f\|_2^2$.

The L^p boundedness of the g-functions can now be obtained by a clever application of the generalization of Theorem 19 to vector-valued functions, as discussed in §7. We show how this works for g_f; the argument for g_f^k is essentially the same. Let

$$\mathcal{H} = \left\{\phi : (0,\infty) \to \mathbb{C}^{n+1} : \|\phi\|_{\mathcal{H}}^2 = \int_0^\infty |\phi(t)|^2 t\, dt < \infty\right\}.$$

For $\epsilon > 0$ and $x \in \mathbb{R}^n$, define $K_\epsilon(x) \in \mathcal{H}$ by $K_\epsilon(x)(t) = \nabla_{t,x} P_{t+\epsilon}$. Routine estimates on the Poisson kernel and its derivatives show that $K_\epsilon(x)$ is indeed in $\mathcal{H}$, and that $\|K_\epsilon(x)\|_{\mathcal{H}} \leq C|x|^{-n}$ and $\|\nabla_x K_\epsilon(x)\|_{\mathcal{H}} \leq C|x|^{-n-1}$ with C independent of ϵ. Moreover, the results of the preceding paragraph imply that the operator $f \mapsto f * K_\epsilon$ (which maps $\mathbb{C}$-valued functions to $\mathcal{H}$-valued functions) is bounded from $L^2(\mathbb{R}^n)$ to $L^2(\mathbb{R}^n, \mathcal{H})$ uniformly in ϵ. It follows that

$$\left(\int \|f * K_\epsilon(x)\|_{\mathcal{H}}^p\right)^{1/p} = \|f * K_\epsilon\|_{L^p(\mathbb{R}^n,\mathcal{H})} \leq B_p \|f\|_p$$

with B_p independent of ϵ. But $g_f(x) = \lim_{\epsilon\to 0} \|f * K_\epsilon(x)\|_{\mathcal{H}}$, so $\|g_f\|_p \leq B_p\|f\|_p$ as desired.

The reverse L^p estimates follow from the preceding results by an easy duality argument. Here it is for g_f^1: polarization of the identity $\|g_f^1\|_2^2 = \frac{1}{4}\|f\|_2^2$ yields

$$4\int_{\mathbb{R}^n}\int_0^\infty \frac{\partial u_1}{\partial t}(t,x)\overline{\frac{\partial u_2}{\partial t}}(t,x)\,t\,dt\,dx = \int_{\mathbb{R}^n} f_1(x)\overline{f_2(x)}\,dx,$$

so by the Cauchy-Schwarz inequality (in t) and Hölder's inequality (in x),

$$\left|\int f_1(x)\overline{f_2(x)}\,dx\right| \leq 4\int g_{f_1}^1(x)g_{f_2}^1(x)\,dx \leq 4\|g_{f_1}^1\|_p\|g_{f_2}^1\|_q,$$

where p and q are conjugates. Taking the supremum over all f_2 with $\|f_2\|_q \leq 1$, for q (and hence p) in $(1,\infty)$, we obtain $\|f_1\|_p \leq 4B_q\|g_{f_1}\|_p$, and we are done.

There are several other nonlinear operations related to g-functions that satisfy similar estimates. We shall not discuss them in detail, but we should at least mention one of the most important, the so-called "Lusin area integral"

$$S_f(x) = \left[\iint_{|y-x|<t<\infty} |\nabla_{t,y}u(t,y)|^2 t^{1-n}\,dy\,dt\right]^{1/2},$$

which is to g_f as nontangential maximal functions are to radial maximal functions. The analogue of S_f for functions on $\mathbb{T}$ (where the term "area integral" is directly appropriate) goes back to a 1930 paper of N. N. Lusin [55]; the theory on $\mathbb{R}^n$ — in particular, the analogue of Theorem 31 for S_f — was developed by Stein [80] (see also [81]).

The g-functions and their relatives are a powerful tool for proving estimates for various classes of linear operators. For example, they are used in establishing the L^p boundedness of the multi-parameter singular integrals described at the end of §9. A more classical application is to the study of *Fourier multipliers*, that is, operators T on $L^2(\mathbb{R}^n)$ of the form $(Tf)\widehat{\ } = m\widehat{f}$ where m is a bounded measurable function. There is a group of related theorems that give general conditions on m under which T is also bounded on L^p for all $p \in (1,\infty)$. Here is one (a minor variation on results of Mikhlin and Hörmander):

***Theorem* 32** — *Suppose m is of class $C^{(k)}$ on $\mathbb{R}^n \setminus \{0\}$ and that $|\partial^\alpha m(\xi)| \leq A|\xi|^{-|\alpha|}$ for $|\alpha| \leq k$, where $k > n/2$. Then the operator T defined by $(Tf)\widehat{\ } = m\widehat{f}$ satisfies $\|Tf\|_p \leq BA\|f\|_p$ for $1 < p < \infty$, where B depends only on p and n.*

A proof of this result can be found in Stein [81], where the main point is to estimate g_{Tf} in terms of a more complicated g-function of f (an analogue of the Littlewood-Paley g_f^*). Here, to give the flavor, we shall give a simple proof of the important special case where m is radial and of "Laplace transform type":

$$m(\xi) = m_0(|\xi|) \quad \text{where} \quad m_0(\lambda) = \lambda \int_0^\infty e^{-\lambda t}\phi(t)\,dt \text{ for some } \phi \in L^\infty(0,\infty.)$$

(Note that this implies that m satisfies the hypotheses of Theorem 32 for all k.) In view of Theorem 31 it is enough to show that $g^1_{Tf} \leq Mg^2_f$. Setting $t = 2\pi s$ and $\psi(s) = \phi(2\pi s)$ in the formula for m_0, one sees that $(Tf)\widehat{\ }(\xi) = -\int_0^\infty (\partial/\partial s)[e^{-2\pi s|\xi|}]\psi(s)\widehat{f}(\xi)\,ds$ and hence

$$(P_t * T_f)\widehat{\ }(\xi) = -\int_0^\infty \frac{\partial}{\partial s}[e^{-2\pi(t+s)|\xi|}\widehat{f}(\xi)]\psi(s)\,ds.$$

In other words, if u and U are the Poisson integrals of f and Tf, respectively,

$$U(t,x) = -\int_0^\infty \frac{\partial u}{\partial s}(t+s,x)\psi(s)\,ds,$$

so with $M = \|\phi\|_\infty$,

$$\left|\frac{\partial U}{\partial t}(t,x)\right| = \left|-\int_0^\infty \frac{\partial^2 u}{\partial s^2}(t+s,x)\psi(s)\,ds\right| \leq M\int_0^\infty \left|\frac{\partial^2 u}{\partial s^2}(t+s,x)\right|\,ds.$$

Hence by the Cauchy-Schwarz inequality,

$$\left|\frac{\partial U}{\partial t}(t,x)\right|^2 \leq \left[M\int_t^\infty \left|\frac{\partial^2 u}{\partial s^2}(s,x)\right|\,ds\right]^2 \leq \frac{M^2}{t}\int_t^\infty \left|\frac{\partial^2 u}{\partial s^2}(s,x)\right|^2 s^2\,ds.$$

Now an integration in t and an application of Fubini's theorem yields $g^1_{Tf}(x)^2 \leq M^2 g^2_f(x)^2$ as desired.

As an immediate application of Theorem 32, we can extend Proposition 16 to the L^p Sobolev spaces for $1 < p < \infty$. As the discussion following that theorem indicates, the missing ingredient is to show that Λ^{iy} is bounded on L^p with a bound that does not grow too rapidly in y. But $(\Lambda^{iy}f)\widehat{\ } = m_y\widehat{f}$ with $m_y(\xi) = (1+|\xi|^2)^{iy/2}$, and it is easy to check that m_y satisfies the hypotheses of Theorem 32 with $A = C(1+|y|)^k$, which is more than sufficient.

Another application of Theorem 32 is to establish an analogue of Theorem 28 for functions on $\mathbb{R}$. (There are also versions on $\mathbb{R}^n$; see [81].) For $k \in \mathbb{Z}$, let χ_{2k} and χ_{2k+1} be the characteristic functions of $[2^k, 2^{k+1}]$ and $[-2^{k+1}, -2^k]$, respectively, and let $(\Delta_j f)\widehat{\ } = \chi_j\widehat{f}$; thus $f = \sum_{-\infty}^\infty \Delta_j f$ for $f \in L^2(\mathbb{R})$.

***Theorem* 33** — *With notation as above, for $1 < p < \infty$ there are constants A_p and B_p such that $A_p\|f\|_p \leq \|\sum_{-\infty}^\infty (|\Delta_j f|^2)^{1/2}\|_p \leq B_p\|f\|_p$ for all $f \in L^p(\mathbb{R})$.*

The idea of the proof is as follows (see [81] for details). Fix $\phi \in C^\infty(\mathbb{R})$ with $\phi(\xi) = 1$ if $1 \le \xi \le 2$ and $\phi(\xi) = 0$ if $\xi \le \frac{1}{2}$ or $\xi \ge 4$. For $k \in \mathbb{Z}$ let $\phi_{2k}(\xi) = \phi(2^{-k}\xi)$ and $\phi_{2k+1}(\xi) = \phi(-2^{-k}\xi)$, so that ϕ_j is a smoothed-out version of χ_j, and let $(S_j f)\widehat{\ } = \phi_j \widehat{f}$. Also, let r_j be the jth Rademacher function (14). It is easy to check that for all $t \in [0,1]$ the functions $m_t(\xi) = \sum_0^\infty [r_{2j}(t)\phi_j(\xi) + r_{2j+1}(t)\phi_{-j}(\xi)]$ satisfy the hypotheses of Theorem 32, with the constant A independent of t, and hence the corresponding multiplier operators are uniformly bounded on L^p, $1 < p < \infty$. The conclusion of Theorem 33, with Δ_j replaced by S_j, then follows by the same clever application of Lemma 11 as in the proof of Theorem 10.

This last conclusion says that the map $f \mapsto (S_j f)_{-\infty}^\infty$ is bounded from $L^p(\mathbb{R})$ to $L^p(\mathbb{R}, l^2)$. Also, by using a vector-valued version of the Hilbert transform and arguing as in the proof of Corollary 9, it is not hard to see that the map $(f_j) \mapsto (\Delta_j f_j)$ is bounded on $L^p(\mathbb{R}, l^2)$. Combining these facts with the observation that $\Delta_j S_j = \Delta_j$, Theorem 33 follows.

Various aspects of Littlewood-Paley theory have been developed and applied in contexts other than $\mathbb{R}^n$ with its standard geometry. The broadest generalization, due to Stein [82], is to the setting of symmetric diffusion semigroups of which the Poisson integral is a paradigmatic example. For another historical review of aspects of the Littlewood-Paley theory, see Stein [83].

11. Harmonic Analysis on Groups

The theory of Fourier series and Fourier transforms constitutes the analysis of functions on $\mathbb{T}$ and $\mathbb{R}$ in terms of the basic functions $e_k(\theta) = e^{ik\theta}$ and $e_\xi(x) = e^{2\pi i \xi x}$, which are precisely the continuous homomorphisms from $\mathbb{T}$ and $\mathbb{R}$ into $\mathbb{T}$. In this section we briefly sketch some of the history of the analogous theories for functions on other types of groups. The story we tell here is seriously incomplete, but at least all parts of it are connected to each other! For more complete accounts of the general theories discussed here with additional historical references we refer to Folland [30] and Mackey [58].

The ideas underlying these theories can be traced back to the work of Gauss on number theory, where (from the modern perspective) he made use of Fourier analysis on the group of integers modulo n. The general picture emerged from the theory of Lie groups, begun by Sophus Lie in the early 1870s and further developed by others — notably Friedrich Engel, Wilhelm Killing, and Élie Cartan — over the succeeding half-century, and the theory of general topological groups, which dates from the 1930s. The crucial prerequisite for analysis is the existence on any locally compact topological group of a (right) *Haar measure*, that is, a Radon measure, unique up to scalar multiples, that is invariant under right translations. (This immediately yields another Radon measure that is invariant under left translations. The two coincide for all the classes of groups we shall discuss below: Abelian, compact, and nilpotent Lie.) For Lie groups this can easily be established by differential-geometric constructions; it was proved by Haar [34] for second countable locally compact groups and by André Weil [94] in general. In what follows, we assume that each locally compact group G is equipped with a fixed Haar measure μ, and we denote $L^p(G,\mu)$ and the volume element $d\mu(x)$ simply by $L^p(G)$ and dx.

With Haar measure in hand, there is a Fourier analysis on any locally compact Abelian group G that directly generalizes the classical theory on $\mathbb{T}$ and $\mathbb{R}$. To begin with, we define the *dual group* $\widehat{G}$ to be the set of all continuous homomorphisms from G to $\mathbb{T}$, which — equipped with pointwise multiplication and the topology of uniform convergence on compact sets — is also a locally compact Abelian group. For $x \in G$ and $\xi \in \widehat{G}$, we denote the action of ξ on x by $\langle \xi, x\rangle$. (When $G = \mathbb{R}^n$, we identify $\widehat{G}$ with $\mathbb{R}^n$ by the pairing $\langle \xi, x\rangle = e^{2\pi i \xi\cdot x}$). Each $x \in G$ defines an element $\widetilde{x}$ of the double dual $\widehat{\widehat{G}}$ by $\langle \widetilde{x}, \xi\rangle = \langle \xi, x\rangle$, and the map $x \mapsto \widetilde{x}$ is an isomorphism of topological groups (the *Pontrjagin duality theorem*). Moreover, G is compact if and only if $\widehat{G}$ is discrete, and vice versa.

The *Fourier transform* of $f \in L^1(G)$ is the bounded continuous function $\widehat{f}$ on $\widehat{G}$ defined by

$$\widehat{f}(\xi) = \int_G f(x)\overline{\langle \xi, x\rangle}\,dx.$$

As on $\mathbb{T}$ and $\mathbb{R}$, the Fourier transform turns convolution into pointwise multiplication and translation by x_0 into multiplication by $\langle \xi, x_0\rangle$; it maps $L^1(G)$ into $C_0(\widehat{G})$ (the Riemann-Lebesgue lemma) and extends to a map from $L^2(G)$ to $L^2(\widehat{G})$ that is unitary if Haar measure on $\widehat{G}$ is suitably normalized; moreover, with the same normalization one has the inversion formula

$$f(x) = \int_{\widehat{G}} \widehat{f}(\xi)\langle \xi, x\rangle\,d\xi$$

(to be taken with a grain of salt unless $\widehat{f} \in L^1(\widehat{G})$). This was worked out first by Weil [94] and then in a more elegant form by Henri Cartan and Roger Godement [7].

For non-Abelian groups G the homomorphisms into $\mathbb{T}$ do not suffice to analyze functions on G, as they are all trivial on the commutator subgroup. The appropriate generalization is found by recognizing that $\mathbb{T}$ may be regarded as the group of unitary 1×1 matrices and considering homomorphisms into higher-dimensional unitary groups instead, that is, unitary representations of G. In the greatest generality, a *unitary representation* of G on a Hilbert space $\mathcal{H}$ is a homomorphism π from G into the group $\mathcal{U}(\mathcal{H})$ of unitary operators on $\mathcal{H}$ that is strongly continuous, i.e., the map $x \mapsto \pi(x)u$ is continuous from G to $\mathcal{H}$ for each $u \in \mathcal{H}$. (Henceforth, when we say "representation" we shall *always* mean "unitary representation".) A representation π is called *irreducible* if the only closed subspaces of $\mathcal{H}$ that are invariant under $\pi(x)$ for all $x \in G$ are $\{0\}$ and $\mathcal{H}$. If π and π' are representations on $\mathcal{H}$ and $\mathcal{H}'$, respectively, a bounded linear map $T : \mathcal{H} \to \mathcal{H}'$ such that $T\pi(x) = \pi'(x)T$ for all $x \in G$ is said to *intertwine* π and π', and two representations are called *equivalent* if there is a unitary map that intertwines them.

We are allowing representations on arbitrary Hilbert spaces from the outset, as a matter of efficiency. However, it should be noted that nobody thought of studying infinite-dimensional representations as such (though a few special cases were well known) until about 1930, when they arose in the context of quantum mechanics.

One of the fundamental results of representation theory, known as *Schur's lemma*, is that if π is an irreducible representation of G on $\mathcal{H}$, the only bounded operators on $\mathcal{H}$ that commute with π (i.e., that intertwine π with itself) are scalar multiples of the identity. (The converse is also true: if π is reducible, the orthogonal projection onto a nontrivial invariant subspace commutes with π.) This has the following consequences. First, if π is irreducible, $\pi(x)$ must be a multiple of the identity whenever x is in the center of G. From this it follows further that if G is Abelian, every irreducible representation is one-dimensional, and hence that $\widehat{G}$ can be regarded as the set of equivalence classes of irreducible representations of G. Also, if G is compact, every irreducible unitary representation of G is finite-dimensional. (The key here is that for any representation π of G on $\mathcal{H}$ and any nonzero $v \in \mathcal{H}$, the operator $T_v(u) = \int_G \langle u, \pi(x)v\rangle \pi(x)v \, dx$ is nonzero, self-adjoint, and compact, and it commutes with π, so its eigenspaces with nonzero eigenvalues are finite-dimensional and invariant under π.)

The representation theory of finite groups was developed by Ferdinand Georg Frobenius, William Burnside, and Frobenius's student Issai Schur beginning about 1890. The historical evolution of the basic concepts and results was rather different from the way the subject would normally be presented now, but that is a story to be told elsewhere; see Curtis [12] and Mackey [58]. For analysts the real starting point is the fundamental paper of Hermann Weyl and his student Fritz Peter [64] in which they showed that certain aspects of this representation theory could be generalized to arbitrary compact groups to yield a Fourier analysis on such groups. (They assumed that their groups were Lie groups solely in order to have an invariant measure available, as their paper antedates Haar [34] by a few years.) Their main theorem is as follows:

***Theorem* 34** *(Peter-Weyl,* 1927*) — Let G be a compact group, with Haar measure normalized so that the measure of G is 1. Let $\widehat{G}$ be a set of irreducible (necessarily finite-dimensional) representations of G containing exactly one member of each equivalence class. For $\pi \in \widehat{G}$, let d_π be the* *dimension of the Hilbert space* $\mathcal{H}_\pi$ *on which* π *acts, and for* $x \in G$ *let* $(\pi_{ij}(x))$ *be the matrix of* $\pi(x)$ *with respect to a fixed orthonormal basis of* $\mathcal{H}_\pi$*. Then*

$$\left\{\sqrt{d_\pi}\pi_{ij} : \pi \in \widehat{G},\ i,j = 1,\dots,d_\pi\right\}$$

is an orthonormal basis for $L^2(G)$*.*

The main contribution of Peter and Weyl — the point where they had to go beyond the algebraic reasoning that had been developed for finite groups — was in proving the completeness of the π_{ij}'s, which they did by showing that their linear combinations are dense in $C(G)$ in the uniform norm.

The theorem can be reformulated in a way that avoids a choice of orthonormal basis for $\mathcal{H}_\pi$. Namely, for $\pi \in \widehat{G}$ and $f \in L^2(G)$, let $\widehat{f}(\pi)$ be the operator on $\mathcal{H}_\pi$ defined by

$$\widehat{f}(\pi) = \int_G f(x)\pi(x^{-1})\,dx = \int_G f(x)\pi(x)^*\,dx.$$

Then the expansion $f = \sum_{\pi,i,j} d_\pi \langle f, \pi_{ij}\rangle \pi_{ij}$ can be restated as

$$f(\cdot) = \sum_{\pi\in\widehat{G}} d_\pi \operatorname{tr}[\widehat{f}(\pi)\pi(\cdot)]$$

where the convergence is in the L^2 norm; and the Parseval identity $\|f\|_2^2 = \sum_{\pi,i,j} d_\pi |\langle f, \pi_{ij}\rangle|^2$ can be restated as

$$\|f\|_2^2 = \sum_{\pi\in\widehat{G}} d_\pi \operatorname{tr}[\widehat{f}(\pi)^*\widehat{f}(\pi)].$$

(The trace of a matrix is invariant under conjugation, so it makes sense to speak of the trace of an operator on a finite-dimensional space.)

There is another important aspect to the Peter-Weyl theorem. Any compact group G acts on $L^2(G)$ by right translations, giving a unitary representation R of G on $L^2(G)$ defined by $[R(x)f](y) = f(yx)$. It is known as the (*right*) *regular representation* of G. With notation as in the Peter-Weyl theorem, it is easy to verify that for each $\pi \in \widehat{G}$ and each $i = 1,\dots,d_\pi$, the subspace of $L^2(G)$ spanned by $\pi_{i1},\dots,\pi_{id_\pi}$ (the ith row of the matrix (π_{ij})) is invariant under R, and that the subrepresentation of R on this subspace is equivalent to π. Hence:

Corollary 35 — The regular representation of a compact group G is a direct sum of irreducible subrepresentations, and for each $\pi \in \widehat{G}$ the equivalence class of π occurs in this direct sum with multiplicity d_π.

The example of $\mathbb{R}$ already shows that compactness is needed for the first assertion of this corollary: there are no one-dimensional subspaces of $L^2(\mathbb{R})$ that are invariant under translations and hence no irreducible subrepresentations of the regular representation of $\mathbb{R}$. Rather, the Fourier inversion formula shows how to synthesize functions in $L^2(\mathbb{R})$ out of an *integral* of the irreducible representations $x \mapsto e^{2\pi i\xi x}$. In 1930 Marshall Stone [88] showed how to generalize this to arbitrary unitary representations of $\mathbb{R}$, and about fourteen years later Godement, Warren Ambrose, and M. A. Naimark independently and almost simultaneously generalized Stone's theorem to arbitrary locally compact Abelian groups.

To state this result we need to recall a definition. If $\mathcal{H}$ is a Hilbert space and X is a set equipped with a σ-algebra of subsets $\mathcal{M}$, an *$\mathcal{H}$-projection-valued measure* on X is a map P from $\mathcal{M}$ to the set of orthogonal projections on $\mathcal{H}$ such that $P(\varnothing) = 0$, $P(X) = I$, $P(E \cap F) = P(E)P(F)$ for all $E, F \in \mathcal{M}$, and $P(\bigcup E_j) = \sum P(E_j)$ (convergence in the strong operator topology) for all finite or infinite sequences $\{E_j\}$ of disjoint sets in $\mathcal{M}$. Any such P determines a $*$-algebra homomorphism from the algebra of bounded measurable functions on X to the algebra of bounded normal operators on $\mathcal{H}$ denoted by $f \mapsto \int_X f(x)\, dP(x)$. (See [30] for details.)

***Theorem* 36** *(Stone,* 1930*; Ambrose, Godement, Naimark,* 1944*) — If π is a unitary representation of a locally compact Abelian group G on a Hilbert space $\mathcal{H}$, there is an $\mathcal{H}$-projection-valued Borel measure P on $\widehat{G}$ such that $\pi(x) = \int_{\widehat{G}} \langle \xi, x\rangle\, dP(\xi)$.*

When π is the regular representation R of G, the measure P is given by $[P(E)f]\widehat{\ } = \chi_E \widehat{f}$. Stone's original theorem ($G = \mathbb{R}$) is often stated in the form $\pi(x) = e^{2\pi i x A}$ where A is a (perhaps unbounded) self-adjoint operator on $\mathcal{H}$; the relation with our formulation is that P is the spectral measure of A, so that $A = \int \xi\, dP(\xi)$.

There is a related result that has influenced many later developments. One has the regular representation R and the "modulation" representation M of $\mathbb{R}^n$ on $L^2(\mathbb{R}^n)$, defined by

$$[R(x)f](y) = f(y + x), \qquad [M(\xi)f](y) = e^{2\pi i\xi\cdot y} f(y), \tag{43}$$

which are intertwined by the Fourier transform. These representations are jointly irreducible, that is, there are no nontrivial closed subspaces of $L^2(\mathbb{R}^n)$ that are invariant under all $R(x)$ and all $M(\xi)$. (If $\mathfrak{X}$ is a closed invariant subspace, f is a nonzero element of $\mathfrak{X}$, and $g \perp \mathfrak{X}$, then $0 = \langle M(\xi)R(x)f, g\rangle = \int e^{2\pi i\xi\cdot y} f(y - x)\overline{g(y)}\, dy$ for all x and ξ; hence $f(\cdot - x)g(\cdot) = 0$ a.e. for all x; hence $f = 0$ or $g = 0$.) It is easily computed that R and M satisfy

$$R(x)M(\xi) = e^{2\pi i\xi\cdot x} M(\xi)R(x.) \tag{44}$$

This is the integrated form of the "canonical commutation relations" of quantum mechanics, and it is important to determine how many different pairs of representations there might be that satisfy this relation. In fact, assuming irreducibility, up to unitary equivalence there is only one; this is the celebrated Stone-von Neumann theorem first announced by Stone [88] (who never published his proof in full) and proved in detail by John von Neumann [93]:

***Theorem* 37** *(Stone,* 1930*; von Neumann,* 1931*) — If π and ρ are unitary representations of $\mathbb{R}^n$ on a Hilbert space that are jointly irreducible and satisfy*

$$\pi(x)\rho(\xi) = e^{2\pi i\xi\cdot x}\rho(\xi)\pi(x), \tag{45}$$

then there is a unitary map $U : \mathcal{H} \to L^2(\mathbb{R}^n)$ such that $U\pi(x)U^{-1} = R(x)$ and $U\rho(\xi)U^{-1} = M(\xi)$, where R and M are defined by (43).

Von Neumann's elegant proof of this can also be found in [27].

By this point it should be clear that one of the main tasks for anyone wanting to do harmonic analysis on a locally compact group G is to classify the irreducible representations of G up to equivalence. For Abelian groups this means describing the dual group $\widehat{G}$ explicitly; this is a well-studied matter. There are also several ways of constructing the irreducible representations of the classical compact matrix groups; see, for example, Hall [35]. For non-Abelian, noncompact groups, however, the irreducible representations are in general infinite-dimensional, and their classification requires a host of new techniques depending on the nature of the group in question. It did not really get under way, except for a few special cases, until after World War II.

In fact, the first complete classification of the irreducible representations of a group for which the finite-dimensional ones do not separate points is implicitly contained in the Stone-von Neumann theorem, although this apparently was not explicitly realized until much later. The relation (44) implies that the operators of the form $e^{2\pi it}M(\xi)R(x)$ ($x, \xi \in \mathbb{R}^n$, $t \in \mathbb{R}$) form a group whose abstract structure is given as follows:

$$\begin{gathered}(x,\xi,t)\cdot(x',\xi',t') = (x+x',\, \xi+\xi',\, t+t'+x\cdot\xi'),\\ (x,\xi,t)^{-1} = (-x,-\xi,-t+\xi\cdot x.)\end{gathered} \tag{46}$$

In other words, the space $\mathbb{R}^n\times\mathbb{R}^n\times\mathbb{R}$, equipped with the operations (46), is a group, now called the *Heisenberg group* H_n, and the map $S : H_n \to \mathcal{U}(L^2(\mathbb{R}^n))$ defined by

$$[S(x,\xi,t)f](y) = [e^{2\pi it}M(\xi)R(x)f](y) = e^{2\pi i(t+\xi\cdot y)}f(y+x)$$

is an irreducible representation of H_n on $L^2(\mathbb{R}^n)$.

Now suppose Π is an arbitrary irreducible representation of H_n on a Hilbert space $\mathcal{H}$. The center of H_n is easily seen to be $Z = \{(0,0,t) : t \in \mathbb{R}\}$, so by Schur's lemma we must have

$\Pi(0,0,t) = e^{2\pi i a t} I$ for some $a \in \mathbb{R}$. If $a = 0$, Π factors through the Abelian group $H_n/Z \cong \mathbb{R}^{2n}$, so it is one-dimensional and of the form $\sigma_{\alpha,\beta}(x,\xi,t) = e^{2\pi i(\alpha\cdot x+\beta\cdot\xi)}$. If $a \neq 0$, let $\Pi^a(x,\xi,t) = \Pi(x, a^{-1}\xi, a^{-1}t)$; then Π^a is again a representation of H^n, and its restrictions $\pi(x) = \Pi^a(x,0,0)$ and $\rho(\xi) = \Pi^a(0,\xi,0)$ to the subgroups $\{(x,0,0) : \xi \in \mathbb{R}^n\}$ and $\{(0,\xi,0) : \xi \in \mathbb{R}^n\}$ are easily seen to satisfy (45). But then by Theorem 37, Π^a is equivalent to S, and hence Π is equivalent to $S_a(x,\xi,t) = S(x, a\xi, at)$. Moreover, the S_a's are all inequivalent to one another: they are already inequivalent on the center Z.

In short, the representations $\sigma_{\alpha,\beta}$ ($\alpha,\beta \in \mathbb{R}^n$) and S_a ($a \in \mathbb{R} \setminus \{0\}$) form a complete set of inequivalent irreducible representations of H_n. But this early achievement in non-Abelian, non-compact representation theory went unremarked for many years; the earliest explicit acknowledgment I have found is in a 1958 paper of Dixmier [17], and it is ignored in the historical survey [58]. And not until the 1970s was the ubiquity of H_n sufficiently appreciated that the name "Heisenberg group" became common usage.

One of the most important devices for constructing representations of a group is the *inducing process*, due originally to Frobenius in the context of finite groups. One starts with a locally compact group G, a closed subgroup H, and a unitary representation σ of H on a Hilbert space $\mathcal{H}_\sigma$. Let G/H be the homogeneous space of right H-cosets and $q : G \to G/H$ the quotient map, $q(x) = Hx$; G/H carries the locally compact topology in which E is open precisely when $q^{-1}(E)$ is open in G. Let $\mathcal{F}_0$ be the space of continuous $\mathcal{H}_\sigma$-valued functions f on G such that (i) $f(hx) = \sigma(h)f(x)$ for $x \in G$ and $h \in H$, and (ii) $q(\mathrm{supp}(f))$ is compact. By (i), for $f \in \mathcal{F}_0$ the norm $\|f(x)\|_{\mathcal{H}_\sigma}$ depends only on $q(x)$, so if G/H admits a G-invariant measure μ, we can form the Hilbert space completion $\mathcal{F}$ of $\mathcal{F}_0$ with respect to the norm $\|f\|^2_{\mathcal{F}} = \int_{G/H} \|f(x)\|^2_{\mathcal{H}_\sigma}\, d\mu(q(x))$, and the action of G on $\mathcal{F}_0$ by right translation, $[\pi(x)f](y) = f(yx)$, extends to a unitary representation of G on $\mathcal{F}$. It is called the *representation of G induced by σ* and denoted by $\mathrm{ind}_H^G(\sigma)$. There is also a modification of this construction that works when G/H has no G-invariant measure; see, for example, [30] (where, however, G/H is taken to be the space of *left* H-cosets and the action of G on $\mathcal{F}_0$ is given by *left* translation).

The simplest example: If H is the trivial subgroup $\{1\}$ of G and σ is the trivial representation of H on $\mathbb{C}$, then $\mathrm{ind}_H^G(\sigma)$ is the regular representation of G.

One of the most far-reaching theorems of representation theory is the so-called Mackey imprimitivity theorem, which George Mackey discovered by generalizing the Stone-von Neumann theorem in three steps, the first two of them in [56] and the last (and most substantial) in [57]. To explore the applications of this theorem would take us too far afield, but as the path to it is fairly short and uses the results discussed above, it is worth sketching here.

First, one generalizes Theorem 37 from $\mathbb{R}^n$ to an arbitrary locally compact Abelian group G. To wit, if π and ρ are jointly irreducible representations of G and $\widehat{G}$ on a Hilbert space $\mathcal{H}$ that satisfy $\pi(x)\rho(\xi) = \langle \xi, x\rangle\rho(\xi)\pi(x)$, there is a unitary map $U : \mathcal{H} \to L^2(G)$ such that $[U^{-1}\pi(x)Uf](y) = f(yx)$ and $[U^{-1}\rho(\xi)Uf](y) = \langle \xi, y\rangle f(y)$.

This may be restated as follows. By Theorem 36 plus Pontrjagin duality, there is an $\mathcal{H}$-projection-valued measure P on G such that $\rho(\xi) = \int_G \langle \xi, x\rangle dP(x)$, and the commutation relation $\pi(x)\rho(\xi) = \langle \xi, x\rangle\rho(\xi)\pi(x)$ is equivalent to $\pi(x)P(E) = P(Ex^{-1})\pi(x)$. Second generalization: when reformulated this way, the preceding result is valid also for non-Abelian groups. That is, if π is a representation of a locally compact group G on $\mathcal{H}$ and P is an $\mathcal{H}$-projection-valued measure on G such that $\pi(x)P(E) = P(Ex^{-1})\pi(x)$ for $x \in G$, $E \subset G$, and π and P are jointly irreducible, there is a unitary map $U : \mathcal{H} \to L^2(G)$ such that $[U^{-1}\pi(x)Uf](y) = f(xy)$ and $U^{-1}P(E)Uf = \chi_E f$.

The final generalization tells what happens if we are given a π and a P as above where P lives not on G but on a homogeneous space G/H. Here is the answer, whose broad scope obviates the irreducibility hypothesis:

***Theorem* 38** *(Mackey,* 1949*) — Suppose G is a locally compact group, H a closed subgroup, π a unitary representation of G on $\mathcal{H}$, and P an $\mathcal{H}$-projection-valued measure on G/H such that $\pi(x)P(E) = P(Ex^{-1})\pi(x)$ for $x \in G$ and $E \subset G/H$. Then there is a unitary representation σ of H (uniquely determined by π and P up to equivalence) and a unitary map $U : \mathcal{H} \to \mathcal{F}$, where $\mathcal{F}$ is the Hilbert space of $\mathrm{ind}_H^G(\sigma)$, such that $U^{-1}\pi(x)U = [\mathrm{ind}_H^G(\sigma)](x)$ for $x \in G$ and $U^{-1}P(E)Uf = (\chi_E \circ q)f$ for $f \in \mathcal{F}$.*

Induced representations are a major source of irreducible representations of non-Abelian, noncompact groups. For example, the representations S_a of the Heisenberg group H_n discussed above are equivalent to representations induced from the one-dimensional representations of the subgroup $\{(0, \xi, t) : \xi \in \mathbb{R}^n,\ t \in \mathbb{R}\}$. Another class of examples comes from the noncompact semisimple Lie groups G, which possess a family of representations known as the "principal series." These are representations induced from the finite-dimensional irreducible representations of subgroups B of a certain type ("Borel subgroups"). We now say a little more about them under the hypothesis that G has "real rank one" (which we need not explain here), with an eye to making connections with the topics discussed earlier; details can be found in [44].

There is a certain nilpotent subgroup N of G that meets each coset of B (except for a set of measure zero) in one point, so the principal series representations can be realized as acting on certain spaces of vector-valued functions on it. N has the following properties: First, it can be identified as a set with $\mathbb{R}^n$ (for suitable n) in such a way that the origin is the group identity, the group operations are given by polynomials in the coordinates, and Lebesgue measure is a (left and right) Haar measure. Second, it is naturally equipped with a one-parameter family $\{\delta_r : r > 0\}$ of group automorphisms called *dilations* which, under the identification with $\mathbb{R}^n$, are of the form

$$\delta_r(x_1, \ldots, x_n) = (r^{\alpha_1} x_1, \ldots, r^{\alpha_n} x_n) \qquad (\alpha_1, \ldots, \alpha_n > 0.) \tag{47}$$

We shall call groups with these two properties *homogeneous groups*. (They are also often referred to as "nilpotent groups with dilations," as they are, in fact, all nilpotent.) The classic examples are the Heisenberg groups H_n, whose canonical dilations are given by $\delta_r(x, \xi, t) = (rx, r\xi, r^2 t)$. ($H_n$ is isomorphic to the N for $G = SU(n+1, 1)$.)

Principal series representations may or may not be irreducible and inequivalent to each other. The question of determining when these conditions hold amounts to the study of the intertwining operators between a representation and itself or between two different representations. It turns out that these operators can be realized as singular integral operators on N. More precisely, they are formally defined as convolution operators,

$$Tf(x) = \int_N f(xy^{-1}) K(y)\, dy,$$

where K is a smooth function on $N \setminus \{0\}$ that has a certain "mean zero" property and satisfies $K(\delta_r(x)) = r^{-Q} K(x)$ where Q is the sum of the exponents α_j in (47). (This degree of homogeneity with respect to the dilations δ_r is precisely the one that puts K just on the borderline of integrability near the origin and near infinity.) In other words, these operators closely resemble the singular integral operators of Theorem 19 except that the translation structure is non-commutative (xy^{-1} rather than $x - y$) and the dilations in question are non-isotropic. Like the latter, they are bounded on $L^p(N)$ for $1 < p < \infty$, but here the L^2 boundedness must be established by an application of the Cotlar-Stein lemma before the Calderón-Zygmund machine can be used to establish the boundedness on other L^p spaces. In fact, this was the original application of the Cotlar-Stein lemma, and the study of intertwining operators by Knapp and Stein [44] was the original application of these singular integrals.

It was realized not long afterward, however, that singular integrals on homogeneous groups, like their classical counterparts, can tell us much about certain kinds of differential operators, particularly those constructed out of non-commuting vector fields where the non-commutativity plays an essential role in their behavior. The first substantial step in this direction was taken in Folland and Stein [31], where singular integrals on the Heisenberg group H_n were used to obtain sharp estimates for certain operators arising in complex analysis in several variables. (The connection with the latter subject comes from the fact that $SU(n+1,1)$ is isomorphic to the group of biholomorphic transformations of a certain domain $D \subset \mathbb{C}^{n+1}$ that is an analogue of the upper half plane in $\mathbb{C}$, and the boundary of D can be naturally identified with H_n.) Other applications involving more general homogeneous groups soon followed; see [26] for a concise survey.

Various other aspects of classical harmonic analysis, including real-variable H^p spaces and Littlewood-Paley theory, have analogues on homogeneous groups; see [32].

12. WAVELETS

Recall the Haar basis for $L^2([0,1])$ defined by (7) and (8). We can translate all its elements by an integer k to obtain a basis for $L^2([k,k+1])$, and then combine all these to obtain an orthonormal basis for $L^2(\mathbb{R})$. With a slight change of notation, this basis is $\{\phi_k : k \in \mathbb{Z}\} \cup \{\psi_{jk} : j \geq 0,\ k \in \mathbb{Z}\}$, where

$$\phi_k = \chi_{[k,k+1)}, \qquad \psi_{jk}(x) = 2^{j/2}\psi(2^j x - k) \text{ with } \psi = \chi_{[0,1/2)} - \chi_{(1/2,1]}. \tag{48}$$

In the expansion $f = \sum_{k\in\mathbb{Z}} \langle f, \phi_k \rangle \phi_k + \sum_{j=0}^{\infty} \sum_{k\in\mathbb{Z}} \langle f, \psi_{jk} \rangle \psi_{jk}$, the first sum $\sum_{k\in\mathbb{Z}} \langle f, \phi_k \rangle \phi_k$ provides a first (probably crude) approximation to f by functions that are constant on each interval $[k,k+1)$, and then the sums $\sum_{k\in\mathbb{Z}} \langle f, \psi_{jk} \rangle \psi_{jk}$ for $j = 0,1,2,\ldots$ provide successively finer levels of detail.

Of course, there is no reason why one has to start with intervals of length 1 as the "base level"; one could dilate everything by a factor of 2^J to start with intervals of length 2^J instead. In the limit as $J \to \infty$ the need for the "base layer" $\{\phi_k\}$ in the basis disappears, and one can see without difficulty that the functions ψ_{jk} defined in (48), but now with both j and k arbitrary integers, constitute an orthonormal basis for $L^2(\mathbb{R})$.

The idea of manufacturing a basis for L^2 out of the translates and dilates of a single function offers many interesting possibilities, and it works also in n dimensions. In what follows, we shall employ the notational convention that if $f \in L^2(\mathbb{R}^n)$, the functions f_{jk} are defined by

$$f_{jk}(x) = 2^{jn/2} f(2^j x - k) \qquad (j \in \mathbb{Z},\ k \in \mathbb{Z}^n.) \tag{49}$$

The factors $2^{jn/2}$ are there to ensure that $\|f_{jk}\|_2$ is independent of j and k. For $n = 1$, an orthonormal basis for $L^2(\mathbb{R})$ of the form $\{f_{jk} : j,k \in \mathbb{Z}\}$ will be called a *wavelet basis*, and the function f will be called the *mother wavelet* of the basis. (For $n > 1$ this needs to be modified slightly, as we shall explain below.)

The mother wavelet ψ of the Haar basis is precisely localized in the sense that it vanishes outside $[0, 1]$ (an advantage), but it is not smooth (a disadvantage). The latter condition is reflected in the fact that its Fourier transform $\widehat{\psi}(\xi) = (1 - e^{-\pi i\xi})^2/2\pi i\xi$ decays only slowly at infinity and hence is *poorly* localized.

It is not hard to find another such basis with the advantage and disadvantage switched. To wit, let $\Psi(x) = (\sin 2\pi x - \sin \pi x)/\pi x$, whose Fourier transform is $\widehat{\Psi}(\xi) = \chi_{[-1,-1/2]}(\xi) + \chi_{[1/2,1]}(\xi)$. It is easy to see that for each j, the Fourier transforms of the functions Ψ_{jk} ($k \in \mathbb{Z}$) are all supported on $A_j = [-2^j, -2^{j-1}] \cup [2^{j-1}, 2^j]$ and constitute an orthonormal basis for $L^2(A_j)$ (the Fourier basis, essentially, once one recognizes that A_j is congruent to $[0, 2^j]$ mod 2^j). It follows that $\{\Psi_{jk} : j, k \in \mathbb{Z}\}$ is a wavelet basis for $L^2(\mathbb{R})$ consisting of slowly decaying functions whose Fourier transforms are precisely localized. Moreover, the localizaton in Fourier space is something we have seen before: the projections $P_j f = \sum_k \langle f, \Psi_{jk}\rangle \Psi_{jk}$ onto the spaces $\{f : \widehat{f} = 0 \text{ outside } A_j\}$ are essentially the Littlewood-Paley projections $\Delta_j f$ of Theorem 33. (In fact, $P_j = \Delta_{2k-2} + \Delta_{2k-1}$). Thus the expansions in terms of the basis $\{\Psi_{jk}\}$ are connected with Littlewood-Paley theory.

The question now arises: can we find wavelet bases whose mother wavelet ψ is both smooth (at least of class C^m for some specified $m \in \mathbb{Z}^+$) and rapidly decaying at infinity (at least faster than polynomially) — in other words, such that ψ and $\widehat{\psi}$ are both well localized? It came as a pleasant surprise to find that the answer is *yes*. The first examples (exponentially decaying and of class C^m, for any finite m) were constructed by Jan-Olov Strömberg [90] in 1982, but they were a bit before their time and had little immediate impact. The next examples (with $\widehat{\psi} \in C_c^\infty$, hence $\psi \in \mathcal{S}$) were discovered by Yves Meyer [60] in 1985 and immediately led to an explosive development of the subject over the next five years or so, incuding the construction by Ingrid Daubechies [13] of compactly supported wavelets of class C^m, for any finite m. Our brief presentation here will be limited to a sketch of the most basic results and their connections with the topics discussed earlier in this paper. Among the excellent books to which the reader may refer for a full account of the results mentioned here and additional material are Daubechies [14], Hernández and Weiss [38], and Meyer [61, 62].

The appropriate general setting for constructing wavelets was worked out by Stéphane Mallat and Meyer in 1986; it works in $\mathbb{R}^n$ for any n. To wit, a *multi-resolution analysis* or *MRA* on $\mathbb{R}^n$ is a family $\{V_j : j \in \mathbb{Z}\}$ of closed subspaces of $L^2(\mathbb{R}^n)$ with the following properties:

(i) There is a function $\phi \in V_0$, called the *scaling function*, such that the functions $\phi_k(x) = \phi(x-k)$ ($k \in \mathbb{Z}^n$) form an orthonormal basis for V_0.

(ii) $V_{j+1} = \{f \circ \delta : f \in V_j\}$ for all j, where $\delta(x) = 2x$.

(iii) $V_j \subset V_{j+1}$ for all j, $\bigcap_{j\in\mathbb{Z}} V_j = \{0\}$, and $\bigcup_{j\in\mathbb{Z}} V_j$ is dense in $L^2(\mathbb{R}^n)$.

We observe that (i) and (ii) imply that for each j, $\{\phi_{jk} : k \in \mathbb{Z}^n\}$ is an orthonormal basis for V_j, where ϕ_{jk} is defined by (49).

To construct an MRA, one generally begins with a function ϕ such that the functions $\phi_k(x) = \phi(x-k)$ are orthonormal, defines V_0 to be their closed linear span and V_j to be the set of all f of the form $f(x) = g(2^j x)$ with $g \in V_0$ (so that (i) and (ii) are satisfied), and then investigates whether (iii) is also satisfied. To this end, it is convenient to reformulate these conditions in terms of the Fourier transform. To avoid pathologies, in what follows we shall assume that the scaling function ϕ satisfies

$$|\phi(x)| \leq C(1+|x|)^{-n-1}, \qquad |\widehat{\phi}(\xi)| \leq C(1+|\xi|)^{-n-1},$$

so that ϕ and $\widehat{\phi}$ are integrable and continuous. Under these conditions, it is not hard to see that the ϕ_k are orthonormal if and only if

$$\sum_{k\in\mathbb{Z}^n} |\widehat{\phi}(\xi+k)|^2 \equiv 1. \tag{50}$$

In this case, with V_j defined as just indicated, we have $V_j \subset V_{j+1}$ if and only if $\phi \in V_1$, and since $\{\phi_{1k} : k \in \mathbb{Z}^n\}$ is an orthonormal basis for V_1, that happens if and only if

$$\begin{gathered} \phi = \sum a_k \phi_{1k} \text{ with } \sum |a_k|^2 < \infty; \quad \text{or equivalently,} \\ \widehat{\phi}(\xi) = m(\xi/2)\widehat{\phi}(\xi/2) \text{ where } m(\xi) = 2^{-n/2}\sum a_k e^{-2\pi i k\cdot\xi}. \end{gathered} \tag{51}$$

The condition $\bigcap V_j = \{0\}$ is always satisfied under the preceding assumptions, and the condition $\overline{\bigcup V_j} = L^2$ turns out to be satisfied precisely when $|\widehat{\phi}(0)| = 1$. (This says more than one might think at first; by (50), it implies that $\widehat{\phi}(k) = 0$ for all nonzero $k \in \mathbb{Z}^n$.) After multiplying ϕ by a constant, then, we may and shall assume that $\widehat{\phi}(0) = 1$.

The relation (51) can be iterated — $\widehat{\phi}(\xi) = m(\xi/2)m(\xi/4)\widehat{\phi}(\xi/4)$, etc. — so, with $\widehat{\phi}(0) = 1$, under mild regularity assumptions one can pass to the limit and obtain $\widehat{\phi}(\xi) = \prod_1^\infty m(\xi/2^j)$. This suggests that one way to construct an MRA is to start with a suitable periodic function m and define $\widehat{\phi}$ to be this infinite product, which automatically yields the relation (51). This is the method used by Daubechies to construct compactly supported wavelets on $\mathbb{R}$ (for which m is a trigonometric polynomial).

In dimension 1, there is a canonical way to pass from an MRA to a wavelet basis. Given an MRA $\{V_j\}$ on $\mathbb{R}$ with scaling function ϕ, let

$$W_j = V_{j+1} \cap V_j^{\perp};$$

thus, by (iii), we have $L^2(\mathbb{R}) = \bigoplus_{-\infty}^{\infty} W_j$. Moreover, with a_k and m as in (51), define $\psi \in V_1$ by

$$\psi = \sum (-1)^{k+1} \overline{a}_{1-k} \phi_{1k}, \quad \text{that is,} \quad \widehat{\psi}(\xi) = e^{-\pi i \xi} \overline{m((\xi+1)/2)} \widehat{\phi}(\xi/2.) \tag{52}$$

***Theorem* 39** — *Given an MRA $\{V_j\}$ on $\mathbb{R}$ with scaling function ϕ, define ψ by (52) and then ψ_{jk} by (49) (with $n = 1$). Then $\{\psi_{jk} : k \in \mathbb{Z}\}$ is an orthonormal basis for W_j for each j, and hence $\{\psi_{jk} : j, k \in \mathbb{Z}\}$ is a wavelet basis for $L^2(\mathbb{R})$.*

For example, the Haar basis arises from this construction by taking the scaling function ϕ to be $\chi_{[0,1)}$ (so V_j is the space of L^2 functions that are constant on each interval $[k2^{-j}, (k+1)2^{-j})$), and the Littlewood-Paley-type basis $\{\Psi_{jk}\}$ presented above arises by taking $\Phi(x) = (\sin \pi x)/\pi x$ (so $\widehat{\Phi} = \chi_{[-1/2,1/2]}$ and V_j is the space of $f \in L^2$ such that $\widehat{f} = 0$ outside $[-2^{j-1}, 2^{j-1}]$). (The basis $\{\Phi_{jk} : k \in \mathbb{Z}\}$ for V_j is the one that arises in the Shannon sampling theorem, and Ψ is sometimes called the *Shannon wavelet* for this reason.)

It is easy to construct n-dimensional MRAs and wavelets out of 1-dimensional ones by using tensor products. For simplicity, let $n = 2$. Given an MRA $\{V_j\}$ on $\mathbb{R}$ with scaling function ϕ and associated mother wavelet ψ, we can define an MRA $\{\mathcal{V}_j\}$ on $\mathbb{R}^2$ by $\mathcal{V}_j = V_j \otimes V_j$; that is, $\mathcal{V}_j$ is the closed linear span of the functions of the form $f(x,y) = g(x)h(y)$ with $g, h \in V_j$; its scaling function is $\Phi = \phi \otimes \phi$, that is, $\Phi(x,y) = \phi(x)\phi(y)$. But now there is a small complication: since $V_{j+1} = V_j \oplus W_j$, we have

$$\mathcal{V}_{j+1} = (V_j \otimes V_j) \oplus (W_j \otimes V_j) \oplus (V_j \otimes W_j) \oplus (W_j \otimes W_j),$$

so $\mathcal{W}_j = \mathcal{V}_{j+1} \cap \mathcal{V}_j^{\perp}$ is given by

$$\mathcal{W}_j = (W_j \otimes V_j) \oplus (V_j \otimes W_j) \oplus (W_j \otimes W_j.)$$

To obtain an orthonormal basis for this we need not one but three mother wavelets: $\Psi^1 = \psi \otimes \phi$, $\Psi^2 = \phi \otimes \psi$, and $\Psi^3 = \psi \otimes \psi$. And indeed, one easily checks that the functions Ψ^{ϵ}_{jk} $(= (\Psi^{\epsilon})_{jk})$ with $j \in \mathbb{Z}$, $k \in \mathbb{Z}^2$, and $\epsilon = 1, 2, 3$ constitute an orthonormal basis for $L^2(\mathbb{R}^2)$. The generalization to $\mathbb{R}^n$ is obvious: one needs $2^n - 1$ mother wavelets to generate a basis.

This phenomenon persists even for MRAs that are not of tensor product type: a *wavelet basis* for $L^2(\mathbb{R}^n)$ must have the form

$$\{\psi^{\epsilon}_{jk} : j \in \mathbb{Z},\ k \in \mathbb{Z}^n,\ \epsilon = 1, \ldots, 2^n - 1\},$$

generated by $2^n - 1$ mother wavelets ψ^ϵ. (The reason, in a nutshell, is that the group $(2\mathbb{Z})^n$ has index 2^n in $\mathbb{Z}^n$.) To simplify the notation in what follows, we define

$$\mathcal{N} = \mathbb{Z} \times \mathbb{Z}^n \times \{1, \ldots, 2^n - 1\}, \qquad \psi_\nu = \psi^\epsilon_{jk} \text{ for } \nu = (j, k, \epsilon) \in \mathcal{N}. \tag{53}$$

An important feature of smooth wavelets is that they have many vanishing moments. To be precise:

***Theorem* 40** — *Suppose ψ is a function of class C^m on $\mathbb{R}^n$ such that $|\psi(x)| \le C(1 + |x|)^{-(m+n+\epsilon)}$ for some $C, \epsilon > 0$ and $\{\psi_{jk} : j \in \mathbb{Z},\ k \in \mathbb{Z}^n\}$ is an orthonormal set. Then $\int P(x)\psi(x)\,dx = 0$ for every polynomial P of degree $\le m$.*

The idea of the proof is as follows: For $j_0 \gg 0$ and $x_0 \in \mathbb{Z}^n$, there is a $k_0 \in \mathbb{Z}^n$ so that the mass of $\psi_{j_0k_0}$ is concentrated in a small ball about x_0. For $j \ll j_0$, ψ_{jk} can be well approximated on this ball by its Taylor polynomial P_{m,x_0} of degree m centered at x_0, so $\int \psi_{j_0k_0}\overline{P}_{m,x_0} \approx \langle \psi_{j_0,k_0}, \psi_{jk}\rangle = 0$, and this approximation improves as $j_0 - j$ increases. By rescaling and varying x_0, one deduces that $\int \psi P = 0$ for a family of polynomials P of degree m that spans the whole space of such polynomials.

Theorem 40 says that if ψ is of class C^m then $\widehat{\psi}$ vanishes to order m at the origin, since $\partial^\alpha\widehat{\psi}(0) = \int(-2\pi i x)^\alpha\psi(x)\,dx$. If ψ and its derivatives up to order m are in L^1, then $\widehat{\psi}$ also vanishes to order m at infinity, so most of the mass of $\widehat{\psi}$ is concentrated in a spherical shell $0 < a < |\xi| < b < \infty$. Hence, for any $f \in L^2$ the sum $S^f_j = \sum_k \langle f, \psi_{jk}\rangle\widehat{\psi_{jk}}(\xi) = \sum_k \langle f, \psi_{jk}\rangle e^{2\pi i k\cdot\xi/2^j}\widehat{\psi}(\xi/2^j)$ is concentrated in the shell $2^j a < |\xi| < 2^j b$, and the expansion $f = \sum_j S^f_j$ provides a decomposition of f into terms concentrated in various frequency bands — not as precisely as with the Shannon wavelet, but still in the spirit of Littlewood-Paley.

Another immediate consequence of Theorem 40 is that *there are no C^∞ wavelets with exponential decay*. Indeed, if $\psi \in C^\infty$, then $\widehat{\psi}$ vanishes to infinite order at the origin, whereas if $|\psi(x)| \le Ce^{-c|x|}$, then $\widehat{\psi}$ extends holomorphically to the strip $|\operatorname{Im}\xi| < c$; the two conditions are incompatible for $\psi \ne 0$. Hence the smoothness-plus-decay properties of the Strömberg, Meyer, and Daubechies wavelets are more or less optimal.

Many of the common function spaces, including L^p and L^p_s $(1 < p < \infty)$, H^p, and BMO, have characterizations in terms of wavelet expansions; we discuss just a couple of these. Perhaps the simplest is the one for L^2 Sobolev spaces. In view of the remarks two paragraphs earlier, the reader should have no difficulty in appreciating the plausibility of the following result:

***Theorem* 41** — *Suppose $\{\psi^\epsilon_{jk}\}$ is a wavelet basis for $L^2(\mathbb{R}^n)$ of class C^m. For $0 \le s < m$, $f \in L^2_s$ if and only if $\sum_{j,k,\epsilon} |\langle f, \psi^\epsilon_{jk}\rangle|^2(1 + 2^{2js}) < \infty$.*

As far as L^p spaces and their relatives are concerned, the essential point is the connection between wavelets and singular integrals. In what follows we employ the terminology of (weak) Calderón-Zygmund operators introduced in §9 in connection with the $T(1)$ theorem (Theorem 26). We also employ the notation (53).

***Theorem* 42** — *Let $\{\psi_\nu : \nu \in \mathcal{N}\}$ be a wavelet basis where the ψ_ν's are of class C^1 and rapidly decaying at infinity. If $\{c_\nu : \nu \in \mathcal{N}\}$ is any bounded set of complex numbers, the operator T defined on L^2 by $T(\sum a_\nu \psi_\nu) = \sum c_\nu a_\nu \psi_\nu$ is a Calderón-Zygmund operator and hence is bounded on L^p for $1 < p < \infty$; moreover, the operator norm of T on L^p depends only on p and $\sup_\nu |c_\nu|$.*

As the operator norm of T on L^2 is $\sup_\nu |c_\nu|$, the point here is that T is associated to the kernel $K(x,y) = \sum c_\nu \psi_\nu(x)\overline{\psi}_\nu(y)$, and one uses the properties of the ψ's to show that K is standard.

In particular, one can take $c_\nu = \pm 1$ for each ν to conclude that the operators $\sum a_\nu \psi_\nu \mapsto \sum \pm a_\nu \psi_\nu$ are uniformly bounded on L^p ($1 < p < \infty$) and hence that $\{\psi_\nu\}$ is an unconditional basis for L^p. More precisely, one can show that $f \in L^p$ if and only if $[\sum |\langle f, \psi_\nu\rangle \psi_\nu|^2]^{1/2} \in L^p$, with equivalence of norms. (Again, there is a resonance with Theorem 28.)

Moreover, the argument that proves Theorem 42 also shows that if $\{\psi_\nu\}$ and $\{\widetilde{\psi}_\nu\}$ are two different wavelet bases satisfying the hypotheses of the theorem, the map $T(\sum a_\nu \psi_\nu) = \sum a_\nu \widetilde{\psi}_\nu$ (which is of course unitary on L^2) is a Calderón-Zygmund operator. This addresses many of the issues that might arise from the nonuniqueness of wavelet bases.

Theorem 42 concerns operators that are diagonal with respect to a wavelet basis, but it is pretty obvious that the result extends to operators that are only "approximately diagonal" — that is, operators T whose matrix elements $\langle T\psi_\nu, \psi_{\nu'}\rangle$ are uniformly bounded and tend to zero sufficiently rapidly as the distance between ν and ν' grows in $\mathcal{N}$. The precise decay conditions are a bit technical, and we shall not state them here.

We conclude by sketching a proof of the $T(1)$ theorem by using wavelets (see [62] for details). Recall that a "weakly bounded" operator T is assumed initially to map $\mathcal{S}$ into $\mathcal{S}'$ but extends to a continuous map from C_c^K (the space of C^K functions of compact support) to its dual space, for suitable $K \in \mathbb{Z}^+$. Thus, if $\{\psi_\nu\}$ is a wavelet basis whose elements belong to C_c^K, the matrix elements $\langle T\psi_\nu, \psi_{\nu'}\rangle$ make sense. One can show that if $T(1) = T^*(1) = 0$, these matrix elements satisfy estimates of the sort referred to in the preceding paragraph; hence T is approximately diagonal and is therefore a Calderón-Zygmund operator. This proves the $T(1)$ theorem for the special case $T(1) = T^*(1) = 0$.

The reduction of the general case to the special one is accomplished as follows. Again let $\{\psi_\nu\}$ be a wavelet basis of class C_c^K. Since $\int \psi_\nu = 0$ (Theorem 40), the numbers $a_\nu = \langle T(1), \psi_\nu \rangle$ and $b_\nu = \langle T^*(1), \psi_\nu \rangle$ are well defined. Choose $\zeta \in C_c^\infty$ with $\int \zeta = 1$, and for $\nu = (j, k, \epsilon)$ let $\zeta^\nu(x) = 2^{jn}\zeta(2^j x - k)$ (independent of ϵ and normalized so that $\int \zeta^\nu = 1$), and define

$$S_1 f = \sum a_\nu \langle f, \zeta^\nu \rangle \psi_\nu, \qquad S_2 f = \sum b_\nu \langle f, \psi_\nu \rangle \zeta^\nu.$$

Then we have $S_1(1) = \sum a_\nu (\int \zeta^\nu) \psi_\nu = \sum a_\nu \psi_\nu = T(1)$ and $S_2(1) = \sum b_\nu (\int \psi_\nu) \zeta^\nu = 0$; since S_1^* and S_2^* are defined by the same sums with ζ^ν and ψ_ν switched, we likewise have $S_1^*(1) = 0$ and $S_2^*(1) = T^*(1)$. (It is not hard to justify these formal calculations rigorously.) An argument similar to the proof of Theorem 42 shows that S_1 and S_2 are associated to standard kernels, and the hypothesis that $T(1)$ and $T^*(1)$ are in BMO can be used to yield estimates on a_ν and b_ν that imply that S_1 and S_2 are bounded on L^2. Hence S_1 and S_2 are Calderón-Zygmund operators, and the preceding argument applies to $T - S_1 - S_2$, so the proof is complete.

References

1. S. Bernstein, Sur la convergence absolue des séries trigonométriques, *C. R. Acad. Sci. Paris*, **158** (1914), 1661-1664.
2. D. L. Burkholder, R. F. Gundy and M. L. Silverstein, A maximal function characterization of the class H^p, *Trans. Amer. Math. Soc.*, **157** (1971), 137-153.
3. A. P. Calderón and R. Vaillancourt, A class of bounded pseudo-differential operators, *Proc. Nat. Acad. Sci. USA*, **69** (1972), 1185-1187; also pp. 491-493 in Calderón's *Selected Papers*, American Mathematical Society, Providence, RI, 2008.
4. A. P. Calderón and A. Zygmund, On the existence of certain singular integrals, *Acta Math.*, **88** (1952), 85-139; also pp. 23-77 in Calderón's *Selected Papers*, American Mathematical Society, Providence, RI, 2008, and pp. 19-73 in Zygmund's *Selected Papers*, vol. 3, Kluwer, Dordrecht, 1989.
5. A. P. Calderón and A. Zygmund, Singular integral operators and differential equations, *Amer. J. Math.*, **79** (1957), 901-921; also pp. 161-181 in Calderón's *Selected Papers*, American Mathematical Society, Providence, RI, 2008, and pp. 257-277 in Zygmund's *Selected Papers*, vol. 3, Kluwer, Dordrecht, 1989.
6. L. Carleson, On convergence and growth of partial sums of Fourier series, *Acta Math.*, **116** (1966), 135-157.
7. H. Cartan and R. Godement, Théorie de la dualité et analyse harmonique dans les groupes Abéliens localement compacts, *Ann. Sci. Ecole Norm. Sup.*, **64**(3) (1947), 79-99; also pp. 1203-1223 in Cartan's *Oeuvres*, vol. III, Springer, Berlin, 1979.

8. M. Christ, A. Nagel, E. M. Stein and S. Wainger, Singular and maximal Radon transforms: Analysis and geometry, *Ann. of Math.*, **150** (1999), 489-577.

9. R. R. Coifman, A real variable characterization of H^p, *Studia Math.*, **51** (1974), 269-274.

10. R. R. Coifman and G. Weiss, *Analyse harmonique non-commutative sur certains espaces homogènes*, Lecture Notes in Math. no. 242, Springer, Berlin, 1971.

11. R. R. Coifman and G. Weiss, Extensions of Hardy spaces and their use in analysis, *Bull. Amer. Math. Soc.*, **83** (1977), 569-645.

12. C. W. Curtis, *Pioneers of representation theory: Frobenius, Burnside, Schur and Brauer*, American Mathematical Society, Providence, RI, 1999.

13. I. Daubechies, Orthogonal bases of compactly supported wavelets, *Comm. Pure Appl. Math.*, **41** (1988), 909-996.

14. I. Daubechies, *Ten lectures on wavelets*, Society for Industrial and Applied Mathematics, Philadelphia, 1992.

15. G. David and J.-L. Journé, A boundedness criterion for generalized Calderón-Zygmund operators, *Ann. of Math.*, **120** (1984), 371-397.

16. P. G. L. Dirichlet, Sur la convergence des séries trigonométriques qui servent à representer une fonction arbitraire entre des limites données, *J. Reine Angew. Math.*, **4** (1829), 157-169.

17. J. Dixmier, Sur les représentations unitaires des groupes de Lie nilpotents III, *Canad. J. Math.*, **10** (1958), 321-348.

18. P. du Bois-Reymond, Über die Fourierschen Reihen, *Nachr. Königl. Ges. Wiss. Göttingen* (1873), 571-582.

19. P. Fatou, Séries trigonométriques et séries de Taylor, *Math. Ann.*, **30** (1906), 335-400.

20. C. Fefferman, Characterizations of bounded mean oscillation, *Bull. Amer. Math. Soc.*, **77** (1971), 587-588.

21. C. Fefferman, Pointwise convergence of Fourier series, *Ann. of Math.*, **98** (1973), 551-571.

22. C. Fefferman and E. M. Stein, H^p spaces of several variables, *Acta Math.*, **129** (1972), 137-193.

23. R. Fefferman and E. M. Stein, Singular integrals on product spaces, *Adv. in Math.*, **45** (1982), 117-143.

24. L. Fejér, Untersuchungen Über Fouriersche Reihen, *Math. Ann.*, **58** (1903), 51-69.

25. E. Fischer, Sur la convergence en moyenne, *C. R. Acad. Sci. Paris*, **144** (1907), 1022-1024.

26. G. B. Folland, Applications of analysis on nilpotent groups to partial differential equations, *Bull. Amer. Math. Soc.*, **83** (1977), 912-930.

27. G. B. Folland, *Harmonic analysis in phase space*, Princeton University Press, Princeton, NJ. 1989.

28. G. B. Folland, *Introduction to partial differential equations* (2nd ed.), Princeton University Press, Princeton, NJ, 1995.

29. G. B. Folland, *Real Analysis* (2nd ed.), John Wiley, New York, 1999.

30. G. B. Folland, *A course in abstract harmonic analysis* (2nd ed.), CRC Press, Boca Raton, FL, 2015.

31. G. B. Folland and E. M. Stein, Estimates for the $\overline{\partial}_b$ complex and analysis on the Heisenberg group, *Comm. Pure Appl. Math.*, **27** (1974), 429-522.

32. G. B. Folland and E. M. Stein, *Hardy spaces on homogeneous groups*, Princeton University Press, Princeton, NJ, 1982.

33. A. Haar, Zur Theorie der orthogonalen Funktionensysteme, *Math. Ann.*, **69** (1910), 331-371; also pp. 47-87 in Haar's *Gesammelte Arbeiten*, Akadémiai Kiadó, Budapest, 1959.

34. A. Haar, Der massbegriff in der theorie der kontinuerlichen Gruppen, *Ann. of Math.*, **34** (1933), 147-169; also pp. 600-622 in Haar's *Gesammelte Arbeiten*, Akadémiai Kiadó, Budapest, 1959.

35. B. C. Hall, *Lie groups, Lie algebras, and representations*, Springer, New York, 2003.

36. G. H. Hardy, The mean value of the modulus of an analytic function, *Proc. London Math. Soc.*, **14**(2) (1915), 269-277; also pp. 549-557 in Hardy's *Collected Papers*, vol. III, Oxford U. Press, Oxford, 1969.

37. G. H. Hardy and J. E. Littlewood, A maximal theorem with function-theoretic applications, *Acta Math.*, **54** (1930), 81-116; also pp. 509-545 in Hardy's *Collected Papers*, vol. II, Oxford University Press, Oxford, 1967.

38. E. Hernández and G. Weiss, *A first course on wavelets*, CRC Press, Boca Raton, FL, 1996.

39. L. Hörmander, Pseudo-differential operators, *Comm. Pure Appl. Math.* **18** (1965), 501-517.

40. R. A. Hunt, On the convergence of Fourier series, pp. 235-255 in D. T. Haimo (ed.), *Orthogonal Expansions and their Continuous Analogues*, Southern Illinois University Press, Carbondale, IL, 1968.

41. F. John and L. Nirenberg, On functions of bounded mean oscillation, *Comm. Pure Appl. Math.*, **14** (1961), 415-426.

42. M. I. Kadets and V. M. Kadets, *Series in Banach spaces*, Birkhäuser, Basel, 1997.

43. A. W. Knapp and E. M. Stein, Singular integrals and the principal series, *Proc. Nat. Acad. Sci. USA*, **63** (1969), 281-284.

44. A. W. Knapp and E. M. Stein, Intertwining operators for semisimple groups, *Ann. of Math.*, **93** (1971), 489-578.

45. J. J. Kohn and L. Nirenberg, An algebra of pseudo-differential operators, *Comm. Pure Appl. Math.*, **18** (1965), 269-305.

46. A. N. Kolmogorov, Une série de Fourier-Lebesgue divergente presque partout, *Fund. Math.*, **4** (1923), 324-328; also pp. 1-5 in Kolmogorov's *Selected Works* (English ed.), vol I, Kluwer, Dordrecht, 1991.

47. A. N. Kolmogorov, Une série de Fourier-Lebesgue divergente partout, *C. R. Acad. Sci. Paris*, **183** (1926), 1327-1328; also pp. 74-75 in Kolmogorov's *Selected Works* (English ed.), vol I, Kluwer, Dordrecht, 1991.

48. A. N. Kolmogorov, Sur les fonctions harmoniques conjuguées et les séries de Fourier, *Fund. Math.*, **7** (1925), 23-28; also pp. 35-39 in Kolmogorov's *Selected Works* (English ed), vol I, Kluwer, Dordrecht, 1991.

49. A. Korányi and S. Vagi, Singular integrals in homogeneous spaces and some problems of classical analysis, *Ann. Scuola Norm. Sup. Pisa*, **25** (1971), 575-648.

50. R. H. Latter, A characterization of $H^p(\mathbb{R}^n)$ in terms of atoms, *Studia Math.*, **62** (1978), 93-101.

51. H. Lebesgue, Intégrale, longueur, aire, *Annali Mat. Pura Appl.*, **7**(3) (1902), 231-359; also pp. 201-331 in Lebesgue's *Oeuvres Scientifiques*, vol. I, L'Enseignement Mathématique, Geneva, 1972.

52. J. E. Littlewood, On mean values of power series (II), *J. London Math. Soc.*, **5** (1930), 179-82; also pp. 715-719 in Littlewood's *Collected Papers*, vol. I, Oxford University Press, Oxford, 1982.

53. J. E. Littlewood and R. E. A. C. Paley, Theorems on Fourier series and power series, *J. London Math. Soc.*, **6** (1931), 230-234; also pp. 599-603 in Littlewood's *Collected Papers*, vol. I, Oxford University Press, Oxford, 1982.

54. J. E. Littlewood and R. E. A. C. Paley, Theorems on Fourier series and power series (II), *Proc. London Math. Soc.*, **42** (1936), 52-89; also pp. 604-641 in Littlewood's *Collected Papers*, vol. I, Oxford University Press, Oxford, 1982.

55. N. Lusin, Sur une propriété des fonctions a carré sommable, *Bull. Calcutta Math. Soc.*, **20** (1930), 139-154.

56. G. W. Mackey, A theorem of Stone and von Neumann, *Duke Math. J.*, **16** (1949), 313-326.

57. G. W. Mackey, Imprimitivity for representations of locally compact groups I, *Proc. Nat. Acad. Sci. USA*, **35** (1949), 537-545.

58. G. W. Mackey, Harmonic analysis as the exploitation of symmetry — A historical survey, *Bull. Amer. Math. Soc. (N.S).*, **3** (1980), 543-698.

59. J. Marcinkiewicz, Sur l'interpolation des opérateurs, *C. R. Acad. Sci. Paris*, **208** (1939), 1272-1273.

60. Y. Meyer, Principe d'incertitude, bases hilbertiennes, et algébres d'opérateurs, *Seminaire Bourbaki*, **28** (1985-6), 209-223.

61. Y. Meyer, *Ondelettes et Opérateurs I: Ondelettes*, Hermann, Paris, 1990.

62. Y. Meyer, *Ondelettes et Opérateurs II: Opérateurs de Calderón-Zygmund*, Hermann, Paris, 1990.

63. R. E. A. C. Paley and A. Zygmund, On some series of functions (1), *Proc. Cambridge Phil. Soc.*, **26** (1930), 337-357; also pp. 337-357 in Zygmund's *Selected Papers*, vol. 1, Kluwer. Dordrecht, 1989.

64. F. Peter and H. Weyl, Die Vollstandigkeit der primitiven Darstellungen einer geschlossenen kontinuerlichen Gruppe, *Math. Ann.*, **97** (1927), 737-755; also pp. 58-75 in Weyl's *Gesammelte Abhandlungen*, vol. III, Springer, Berlin, 1968.

65. D. H. Phong and E. M. Stein, Hilbert integrals, singular integrals, and Radon transforms I, *Acta Math.*, **157** (1986), 99-157.

66. D. H. Phong and E. M. Stein, Singular Radon transforms and oscillatory integrals, *Duke Math. J.*, **58** (1989), 347-369.

67. M. Plancherel, Contribution à l'étude de la représentation d'une fonction arbitraire par des intégrales definies, *Rend. Circolo Mat. Palermo*, **30** (1910), 289-335.

68. H. Rademacher, Einige Sätze über Reihen von allgemeinen Orthogonalfunktionen, *Math. Annalen*, **87** (1922), 112-138.

69. B. Riemann, Über die Darstellbarkeit einer Funktion durch eine trigonometrische Reihe, *Abh. Königl. Ges. Wiss. Göttingen*, **13** (1867); also pp. 227-271 in Riemann's *Gesammelte Mathematische Werke* (2nd ed), Teubner, Leipzig, 1892 (reprinted with additional commentaries by Springer, Berlin, 1990).

70. F. Riesz, Sur les systèmes orthogonaux de fonctions, *C. R. Acad. Sci. Paris*, **144** (1907), 615-619; also pp. 378–381 in Riesz's *Oeuvres Complètes*, Akadémiai Kiadó, Budapest, 1960.

71. F. Riesz, Untersuchungen über Systeme integrierbarer Funktionen, *Math. Annalen*, **69** (1910), 449-497; also pp. 441-489 in Riesz's *Oeuvres Complètes*, Akadémiai Kiadó, Budapest, 1960.

72. F. Riesz, Über die Randwerte einer analytischen Funktion, *Math. Zeit.*, **18** (1923), 117-124; also pp. 645-653 in Riesz's *Oeuvres Complètes*, Akadémiai Kiadó, Budapest, 1960.

73. F. Riesz and M. Riesz, Über die Randwerte einer analytischen Funktion, *Comptes Rendus* 4^{e} *Congrès Math. Scand. Stockholm*, (1916), 27-44; also pp. 537–554 in F. Riesz's *Oeuvres Complètes*, Akadémiai Kiadó, Budapest, 1960, and pp. 195-212 in M. Riesz's *Collected Papers*, Springer, Berlin, 1988.

74. M. Riesz, Les fonctions conjuguées et les séries de Fourier, *C. R. Acad. Sci. Paris*, **178** (1924), 1464-1467; also pp. 360-362 in Riesz's *Collected Papers*, Springer, Berlin, 1988.

75. M. Riesz, Sur les fonctions conjuguées, *Math. Zeit.*, **27** (1927), 218-244; also pp. 410-436 in Riesz's *Collected Papers*, Springer, Berlin, 1988.

76. M. Riesz, Sur les maxima des formes bilinéaires et sur les fonctionelles linéaires, *Acta Math.*, **49** (1927), 465-497; also pp. 377-409 in Riesz's *Collected Papers*, Springer, Berlin, 1988.

77. L. Schwartz, Généralisation de la notion de fonction, de dérivation, de transformation de Fourier et applications mathématiques et physiques, *Ann. Univ. Grenoble Sect. Sci. Math. Phys.*, **21** (1945), 57-74.

78. L. Schwartz, Théorie des distributions et transformation de Fourier, *Ann. Univ. Grenoble Sect. Sci. Math. Phys.*, **23** (1948), 7-24.

79. E. M. Stein, Interpolation of linear operators, *Trans. Amer. Math. Soc.*, **83** (1956), 482-492.

80. E. M. Stein, On the functions of Littlewood-Paley, Lusin, and Marcinkiewicz, *Trans. Amer. Math. Soc.*, **88** (1958), 430-466.

81. E. M. Stein, *Singular integrals and differentiability properties of functions*, Princeton University Press, Princeton, NJ, 1970.

82. E. M. Stein, *Topics in harmonic analysis related to the Littlewood-Paley theory*, Princeton University Press, Princeton, NJ, 1970.

83. E. M. Stein, The development of square functions in the work of A. Zygmund, *Bull. Amer. Math. Soc. (N. S.)*, **7** (1982), 359-376.

84. E. M. Stein, *Harmonic analysis: Real-variable methods, orthogonality, and oscillatory integrals*, Princeton University Press, Princeton, NJ, 1993.

85. E. M. Stein and S. Wainger, Problems in harmonic analysis related to curvature, *Bull. Amer. Math. Soc.*, **84** (1978), 1239-1295.

86. E. M. Stein and G. Weiss, On the theory of harmonic functions of several variables, I: The theory of H^p spaces, *Acta Math.*, **103** (1960), 26-62.

87. E. M. Stein and G. Weiss, *Introduction to Fourier analysis on Euclidean spaces*, Princeton University Press, Princeton, NJ, 1971.

88. M. H. Stone, Linear transformations in Hilbert space III: Operational methods and group theory, *Proc. Nat. Acad. Sci. USA*, **16** (1930), 172-175.

89. B. Street, *Multi-parameter singular integrals*, Princeton University Press, Princeton, NJ, 2014.

90. J.-O. Strömberg, A modified Franklin system and higher order spline systems on $\mathbb{R}^n$ as unconditional bases for Hardy spaces, pp. 475-494 in W. Beckner *et al*., (eds.), *Conference on Harmonic Analysis in Honor of Antoni Zygmund*, Wadsworth, Belmont, CA, 1983.

91. M. E. Taylor, *Pseudodifferential operators*, Princeton University Press, Princeton, NJ, 1981.

92. G. O. Thorin, An extension of a convexity theorem due to M. Riesz, *Kungl. Fysiografiska Saellskapets i Lund Förhandlingar*, **8**(14) (1939).

93. J. von Neumann, Die Eindeutigkeit der Schrödingerschen Operatoren, *Math. Ann.*, **104** (1931), 570-578; also pp. 221–229 in von Neumann's *Collected Works*, vol. II, Pergamon Press, New York, 1961.

94. A. Weil, *L'Intégration dans les Groupes Topologiques et ses applications*, Hermann, Paris, 1940.

95. A. Zygmund, On a theorem of Marcinkiewicz concerning interpolation of operators, *J. Math. Pures Appl.*, **35**(9) (1956), 223-248; also pp. 214-239 in Zygmund's *Selected Papers*, vol. 3, Kluwer, Dordrecht, 1989.

96. A. Zygmund, *Trigonometric series* (2 volumes, reprinted as a single volume), Cambridge University Press, Cambridge, 1968.