ON THE EVOLUTION OF NONCOMMUTATIVE HARMONIC ANALYSIS

KENNETH I. GROSS

Dedicated to my dear friend Ernst Snapper, with admiration and respect

Introduction. This paper is an outgrowth of my preparation for an invited address delivered on January 31, 1977, to the Mathematical Association of America at its annual meeting in St. Louis, Missouri. It is a short, elementary exposition of the main themes that lead to the current frontier in noncommutative harmonic analysis. As such, it represents a written elaboration of what I felt could reasonably be presented in one hour to a general mathematical audience.

Quite briefly, noncommutative harmonic analysis is the meeting ground of group theory, analysis, and geometry. It is an area of mathematics that has been flourishing for fifty years. For those who practice the art, its hybrid nature is a source of uncommon beauty, depth, and difficulty. Most significantly, noncommutative harmonic analysis has been inspirational for new points of view in a wide range of applications that includes not only group theory, analysis, and geometry, but number theory, probability, ergodic theory, and modern physics.

My goals here are to give the general reader a feeling for the historical flow of ideas and to display some of the more readily accessible connections with classical mathematics and nonclassical physics. Although one principally writes for the reader, this has been a marvelous personal experience. For it is not the typical task of a research mathematician to attempt a nontechnical commentary on a subject which seems foreordained by nature as intrinsically technical.

Finally, it is not possible to enumerate by name all the people who have contributed to my point of view. However, special thanks are extended to my colleague Jonathan Brezin for stimulating and enjoyable conversations that helped organize my thoughts at crucial stages of the exposition.

Occasionally, a comment came to mind that seemed worthy of mention, but was either parenthetical to the main theme or else a slightly more technical supplement to the main theme. To help preserve the continuity of discourse, such remarks are set aside in smaller print.

1. Classical harmonic analysis: Fourier series. By reason of its importance in almost all aspects of harmonic analysis, as well as its primary historical position, we begin with an account of Fourier series. The fundamental ideas originate in differential equations, in particular with solutions of the classical partial differential equations of mathematical physics. Illustrative is the *one-dimensional wave equation*

$$\frac{\partial^2 u}{\partial x^2} = \frac{1}{a^2}\frac{\partial^2 u}{\partial t^2} \tag{1.1}$$

first derived by d'Alembert in 1747 as the law governing the motion of a vibrating string. In this equation, $u = u(x,t)$ represents the displacement from equilibrium at time t of the point x on the string, and the constant a^2 is determined from the physical characteristics of tension T and mass density ρ, according to the formula $a^2 = T/\rho$. The general solution of (1.1), again due to d'Alembert at that early date, is given by the expression

$$u(x,t) = f_1(x+at) + f_2(x-at) \tag{1.2}$$

where f_1 and f_2 are arbitrary twice differentiable functions of a real variable. This solution can be viewed as the superposition of two "waves" f_1 and f_2 propagating with speed a in opposite directions along the string.

Professor Gross did his graduate work at Brandeis University and Washington University under the direction of R. A. Kunze. He has taught at Tulane, Dartmouth, the University of California at Irvine, the University of North Carolina (since 1973), and the University of Utah (on leave from North Carolina). His principal fields of interest are harmonic analysis, representations of groups, and undergraduate education.—*Editors*

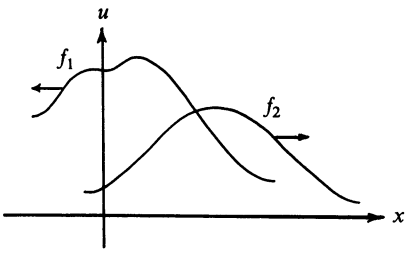

FIG. 1.

If, for example, one imposes upon a solution to (1.1) the *boundary conditions*

$$u(0,t)=u(L,t)=0 \tag{1.3}$$

for all t (with L fixed), as well as the *initial conditions*

$$u(x,0)=\phi(x) \quad \text{and} \quad \frac{\partial u}{\partial t}(x,0)=\psi(x)$$

for all x (with ϕ and ψ fixed), it is not difficult to see that the solution (1.2) then takes the form

$$u(x,t)=f(at+x)-f(at-x) \tag{1.4}$$

where the function f is periodic with period $2L$ and is uniquely determined, up to an arbitrary constant, as

$$f(x)=\frac{1}{2}\left(\phi(x)+\frac{1}{a}\int_0^x \psi(s)ds\right). \tag{1.5}$$

Of course, the initial data ϕ and ψ should be periodic with period $2L$. Physically, (1.4) and (1.5) describe the motion of a string of length L that is clamped at the end points and which at time $t=0$ is given initial displacement $\phi(x)$ and initial velocity $\psi(x)$.

The notion of *harmonic analysis* originates with the possibility of superimposing certain of the solutions (1.4); namely, the nth fundamental modes, or nth *harmonics*,

$$\sin(n\pi x/L)\cos(n\pi at/L) \quad \text{and} \quad \sin(n\pi x/L)\sin(n\pi at/L)$$

which correspond in (1.5) to $f(x)$ of the form $\sin(n\pi x/L)$ and $\cos(n\pi x/L)$, respectively. As expressed most colorfully by Daniel Bernoulli in 1753, "all sonorous bodies contain potentially an infinity of corresponding ways of making their regular vibrations." Rephrased more mathematically, $f(x)$ should be of the form

$$f(x)=a_0+\sum_{n=1}^{\infty}\{a_n\cos(n\pi x/L)+b_n\sin(n\pi x/L)\}. \tag{1.6}$$

The study of such series representations and related generalizations is referred to as *Fourier analysis*. It would appear that this reference to Fourier is well deserved, for not only does Fourier's 1807 investigation of heat flow give the first broad examination of the series (1.6), including the determination of the coefficients in terms of f, but it marks the beginning of modern real analysis. In particular, many of our most decisive mathematical concepts—for example, the *modern notion of function, contemporary set theory*, both the *Riemann* and *Lebesgue integrals*, and, in more recent years, *distribution theory*—were generated by problems in Fourier analysis.

A modern treatment of Fourier series would recast (1.6) in complex notation (with $L=\pi$) as

$$f(\theta)=\sum_{n=-\infty}^{\infty} c_n e^{in\theta} \tag{1.7}$$

where

$$c_n=\hat{f}(n)=\frac{1}{2\pi}\int_0^{2\pi} f(\theta)e^{-in\theta}d\theta. \tag{1.8}$$

Here, $0 \leqslant \theta \leqslant 2\pi$ parametrizes the unit circle $\mathbf{T}$ of complex numbers $e^{i\theta}$ of absolute value 1. The right side of (1.7) is called the *Fourier series* of f, and $\{\hat{f}(n)\}$ is the sequence of *Fourier coefficients* of f.

There are many techniques or, in the technical jargon, methods of summability, for obtaining convergence in (1.7). The most common and elementary procedure involves the introduction of the inner product

$$\langle f_1 | f_2 \rangle = \frac{1}{2\pi} \int_0^{2\pi} f_1(\theta)\, \overline{f_2(\theta)}\, d\theta, \tag{1.9}$$

and applies to functions f on $\mathbf{T}$ which are square-integrable; that is, the norm

$$\|f\| = \langle f | f \rangle^{1/2} = \left\{ \frac{1}{2\pi} \int_0^{2\pi} |f(\theta)|^2 d\theta \right\}^{1/2} \tag{1.10}$$

is finite. Thus, the functions e_n on $\mathbf{T}$ given by $e_n(\theta) = e^{in\theta}$ form a complete orthonormal system; (1.7) is quite simply the expansion of f in terms of this orthonormal basis (i.e., $\hat{f}(n) = \langle f | e_n \rangle$ for all n); and equality in (1.7) is interpreted in the mean-square sense as

$$\lim_{N \to \infty} \|f - \sum_{n=-N}^{N} c_n e_n\| = 0.$$

In other words,

$$\|f\|^2 = \sum_{n=-\infty}^{\infty} |\hat{f}(n)|^2. \tag{1.11}$$

Equation (1.11) is an example of a so-called *Parseval* or *Plancherel Formula*, the extension of which to more broad, general contexts will be a central theme of the sequel.

The previous paragraph describes what is known as the L^2-theory of Fourier series. In contrast to convergence in the mean-square sense, the general problem of *pointwise* convergence of Fourier Series is extraordinarily difficult. Thus, by an argument of Dirichlet (1837), it is not too hard to show that the Fourier series of a continuously differentiable function on the circle is absolutely convergent. Yet, it was not proved until 1966 that the Fourier series of a continuous function (more generally, square-integrable function) converges almost everywhere. The proof of this fact, due to L. Carleson, is very complicated (cf. [38]). A general heuristic principle can be stated: The smoother the function, the more rapidly its Fourier series converges, and vice-versa.

We shall give two interpretations of the Fourier series (1.7). The first illustrates the fundamental and intrinsic connection between Fourier series and differential equations on the circle. Let D denote differentiation with respect to θ. It is appropriate (although it may seem somewhat premature in this one-dimensional context) to call the operator $\Delta = D^2$ the *Laplacian* on $\mathbf{T}$. Upon the observation that $\Delta e_n = -n^2 e_n$ for all n, the Fourier series (1.7) takes on an entirely new appearance. For we see that $-\Delta$ *is a positive (and self-adjoint) differential operator on* $\mathbf{T}$, *and* (1.7) *is the expansion of* f *in eigenfunctions of the Laplacian.*

Moreover, from two integrations by parts,

$$(\Delta y)\hat{}\,(n) = -n^2 \hat{y}(n) \tag{1.12}$$

for any twice differentiable function y on $\mathbf{T}$ and for all n. Now consider the differential equation on $\mathbf{T}$

$$\Delta y = -f + \alpha, \tag{1.13}$$

where f is any fixed continuous (more generally, square-integrable) function on $\mathbf{T}$ and α is the constant

$$\alpha = \hat{f}(0) = \frac{1}{2\pi} \int_0^{2\pi} f(\theta) d\theta. \tag{1.14}$$

By (1.12), the solution y is uniquely determined up to an additive constant by the equations

$$\hat{y}(n) = \frac{1}{n^2} \hat{f}(n), \qquad n \neq 0, \tag{1.15}$$

for its Fourier coefficients.

The constant α appears in (1.13) for the following technical reason. By (1.12) with $n=0$, the left side of (1.13) is a function with mean value zero over **T**. Thus, (1.14) gives precisely the value of α for which the right side of (1.13) also has mean zero. Put another way, by integration of (1.13) one sees that this is the value of α such that $dy/d\theta$ is continuous on **T**.

Of course, one really wants the function y, not its Fourier coefficients. Posed in more generality, given functions f_1 and f_2, we want that function, denoted $f_1 * f_2$, for which the Fourier coefficients are

$$(f_1 * f_2)\hat{}\,(n) = \hat{f}_1(n)\hat{f}_2(n). \tag{1.16}$$

By means of formula (1.8) for $\hat{f}_1(n)$ and $\hat{f}_2(n)$, it is easy to verify that the desired function is given by the *convolution*

$$(f_1 * f_2)(\theta) = \frac{1}{2\pi}\int_0^{2\pi} f_1(\phi) f_2(\theta - \phi)\, d\phi \tag{1.17}$$

of f_1 and f_2.

From the algebraic point of view, a vector space V of functions on **T**, which is closed under convolution, has the structure of a commutative ring relative to convolution as multiplication. As examples of such spaces which arise in a variety of both abstract and concrete problems in analysis, we list the spaces $\mathcal{C}(\mathbf{T})$ of continuous functions, $\mathcal{C}^1(\mathbf{T})$ of continuously differentiable functions, $\mathcal{S}(\mathbf{T})$ of infinitely differentiable functions, $L^1(\mathbf{T})$ of Lebesgue integrable functions, and $L^2(\mathbf{T})$ of Lebesgue square-integrable functions, all of which are closed under convolution.

Thus, let g be that function on **T** such that

$$\hat{g}(n) = \begin{cases} 0 & \text{if } n = 0 \\ 1/n^2 & \text{if } n \neq 0. \end{cases} \tag{1.18}$$

Then the complete solution of the differential equation (1.13) (*nonhomogeneous Laplace's equation* on **T**, if you like) is given by

$$y = g * f + c \tag{1.19}$$

where c is an arbitrary constant. An elementary calculation shows that g is the function defined by $g(\theta) = (1/2)\theta^2 - \pi\theta + (1/3)\pi^2$ for $0 \leqslant \theta \leqslant 2\pi$.

The second interpretation of Fourier series makes explicit its group theoretic nature. In the usual way—that is, by means of addition modulo 2π in the parameter θ—the circle **T** is an abelian group. Moreover, the functions e_n are group homomorphisms, and these are all the continuous homomorphisms of the circle into the multiplicative group of nonzero complex numbers. Under pointwise multiplication as the law of composition, these homomorphisms themselves form a group, isomorphic to the additive group **Z** of integers, which is called the *dual group* of **T**. From this perspective, *the Fourier series* (1.7) *is the expansion of f in terms of the dual group.*

Although there will be many variations, the theme throughout the sequel is as follows: *Try to analyze, or decompose, spaces of functions on a group, or a set on which a group acts, in terms of the most "elementary" functions which mirror the group operation*; that is to say, *the fundamental harmonics*. As we shall see, these functions are constructed from suitable homomorphisms, called *irreducible representations*, of the group in question.

2. Algebraic harmonic analysis: representations of finite groups. The concept of an abstract group, progenitive of the modern axiomatic point of view in algebra, developed in the latter third of the past century out of concrete manifestations—such as permutation groups, Galois groups, and Lie groups—that had arisen in the theory of equations, number theory, and geometry. In particular, the study of finite groups (in the modern sense) was already well under way, when in the 1890's Frobenius invented the notion of a representation of an abstract group by matrices, thereby introducing into the subject the techniques of linear algebra.

In modern terms, a *representation* T of a finite group G is a function which associates to each element x of G a linear transformation $T(x)$ on a nonzero finite-dimensional complex vector space $V=V_T$, such that two properties hold: (1) $T(xy)=T(x)T(y)$ for all $x,y\in G$; and (2) $T(e)=I$, where e is the identity in G and I is the identity transformation on V. Clearly, $T(xx^{-1})=T(x)T(x^{-1})=I$, or $T(x)^{-1}=T(x^{-1})$ for all $x\in G$; so a representation of G is simply a homomorphism of G into the general linear group $GL(V)$ of invertible linear transformations on V.

More in line with the original development by Frobenius (more accurately, the simplified treatment by Schur), it is often useful to think concretely of a representation as having matrix values. For if we fix a basis $e_1, e_2,\ldots,e_d$ of V where $d=\dim V$ then the equations $T(x)e_j=\sum_{i=1}^{d}t_{ij}(x)e_i$ define the matrix $t(x)=t_{ij}(x)$ of $T(x)$, and the mapping $x\to t(x)$ is the matrix-valued realization of T relative to the indicated basis. In short, the group $GL(V)$ is isomorphic to the *general linear group* $GL(d,\mathbf{C})$ of nonsingular $d\times d$ complex matrices. The number $d=d_T$ is called the *degree* of T, and the d^2 functions t_{ij} on G are the *matrix entries of T in the given basis.*

Within a decade, most notably through the work of Frobenius, Schur, and Burnside, the major framework was all in place, and representation theory had already been shown to be among the most powerful devices for penetrating the structure of a finite group.

More than amusing, it is in many ways instructive to contrast the change over a short period of time in Burnside's perception of the power of representation theory. The first edition of his famous book [5], written in 1897, completely omits any discussion of representation theory, for, in Burnside's words, "it would be difficult to find a result [in the pure theory of abstract groups] that could be most directly obtained by the consideration of groups of linear transformations." However, the 1911 second edition relies heavily upon the methods of representation theory; and in disavowing his previous opinion, Burnside comments that "it is now more true to say that for further advances in the abstract theory one must look largely to the representation of a group as a group of linear substitutions [i.e., transformations]." Indeed, seven years earlier, Burnside had successfully employed representation theory to obtain his celebrated result that a group of order $p^\alpha q^\beta$ (p,q prime) is solvable. Thus, one may safely conclude that, whatever the gastronomical discomfort, rich mathematical rewards accompanied the eating of his words.

Here, our look into the representation theory of a finite group will be necessarily brief, for limited space and the topic at hand preclude both a detailed treatment and a level of generality appropriate from the standpoint of algebra. What we are looking for is a glimpse of noncommutative harmonic analysis in its algebraic essence, shorn of complicating analytic considerations. For, as we will indicate in the next section, the role played by representation theory of a finite group is far more substantive in the genesis of modern harmonic analysis than in that of Fourier series, the group theoretic aspect of which is a relatively recent observation.

To be more precise concerning the degree of generality in which the results of this section are valid, we could replace the complex field of scalars by any field the characteristic of which is relatively prime to the order of the finite group in question. This yields the classical "ordinary theory" of representations of a finite group. There is a "modular theory," due largely to R. Brauer, in which there is no such restriction on the characteristic of the field. However, this theory is much too technical to be included here.

As a highly illustrative example, consider the symmetric group $\mathcal{G}$ on three objects, defined to be the group of all permutations of the set $S=\{1,2,3\}$. For $g\in\mathcal{G}$ and $s\in S$, let sg denote the image of s under the permutation g. To proceed to a richer structure, form the three-dimensional vector space $C(S)$ of all complex-valued functions on S. Then the set-theoretic action of $\mathcal{G}$ on S lifts to a linear action R of $\mathcal{G}$ on $C(S)$, called the *standard representation of $\mathcal{G}$ on $C(S)$*. Specifically, to each $g\in\mathcal{G}$ is associated the linear transformation of $C(S)$ given by

$$(R(g)f)(s)=f(sg). \tag{2.1}$$

Clearly, R is a representation of $\mathcal{G}$. The fundamental problem in harmonic analysis on S can now be formulated: *Decompose $C(S)$, or equivalently R, in terms of those functions f on S whose translates $R(g)f$ under $\mathcal{G}$ span as small a subspace of $C(S)$ as possible.*

Some notation is needed. Given a representation T of a finite group G on a space $V=V_T$, we say that a subspace W of V is *invariant* under T if $T(g)W\subset W$ for all $g\in G$. For a non-zero invariant subspace W, restriction of the transformations $T(g)$ to W defines a representation T_W of G on the space W, called the *subrepresentation* of T on W. A non-zero invariant subspace is said to be *irreducible* if it has no proper invariant subspace. If the representation space V_T itself is irreducible, then T is termed *irreducible*. Evidently, either a representation is irreducible, or else it has an irreducible subrepresentation.

Returning to the example, we shall say that a function f on S, not identically zero, is a *fundamental harmonic*—or more simply, an *S-harmonic*—if its translates $R(g)f$ under $\mathcal{G}$ span an irreducible subspace of $C(S)$. With the help of the natural inner product

$$\langle f_1|f_2\rangle=\sum_{s=1}^{3} f_1(s)\overline{f_2(s)} \tag{2.2}$$

on $C(S)$, it is a simple matter to find the spaces of S-harmonics. For it is clear that the constant functions form a one-dimensional space, say W_1, of S-harmonics. On the other hand, its orthogonal complement W_2, consisting of functions of mean zero (i.e., $\langle f|1\rangle=f(1)+f(2)+f(3)=0$), is a two-dimensional invariant subspace, and it is not difficult to see that W_2 is irreducible. Thus,

$$C(S)=W_1\oplus W_2 \tag{2.3}$$

is the decomposition of $C(S)$ into spaces W_1 and W_2 of S-harmonics, and for all $g\in\mathcal{G}$ the transformation $R(g)$ decomposes as the direct sum

$$R(g)=1\oplus\lambda(g) \tag{2.4}$$

where 1 *denotes the* 1*-dimensional subrepresentation* ($1(g)=1$ *for all* g) *on* W_1, *and* λ *is the two-dimensional subrepresentation on* W_2.

The standard representation R, the two-dimensional representation λ, and the decomposition (2.4) all have geometric interpretations. To see this, let e_1,e_2,e_3 be the usual orthonormal basis for 3-dimensional space. Viewed as functions in $C(S)$, $e_j(k)=1$ when $j=k$ and $e_j(k)=0$ otherwise ($j,k=1,2,3$). Relative to this basis

$$R(\tau)=\begin{pmatrix}0&1&0\\1&0&0\\0&0&1\end{pmatrix}\quad\text{and}\quad R(\sigma)=\begin{pmatrix}0&1&0\\0&0&1\\1&0&0\end{pmatrix}$$

where τ is the transposition (12) and σ is the cyclic permutation (123). To obtain the matrix version of (2.4) change to a new orthonormal basis; namely, the basis E_1,E_2,E_3 where $E_1=(e_1-e_2)/\sqrt{2}$, $E_2=(e_1+e_2-2e_3)/\sqrt{6}$, and $E_3=(e_1+e_2+e_3)/\sqrt{3}$. Then E_3 spans the 1-dimensional subspace of $C(S)$ of constant functions, and E_2, E_3 span a 2-dimensional invariant subspace. Hence, with respect to this new basis of S-harmonics,

$$R(g)=\left[\begin{array}{cc|c}\multicolumn{2}{c|}{\lambda\ (g)} & 0\\ \hline 0 & 0 & 1\end{array}\right]\quad\text{for all } g \text{ in } \mathcal{G}.$$

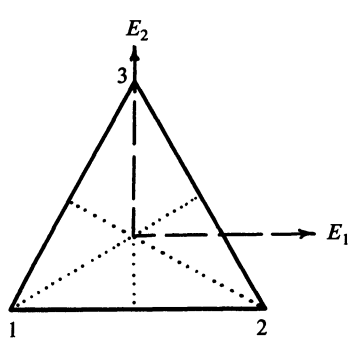

FIG. 2.

Moreover, λ is the familiar representation of $\mathcal{G}$ as symmetries of an equilateral triangle. For example, the matrices

$$\lambda(\tau)=\begin{pmatrix}-1&0\\0&1\end{pmatrix}\quad\text{and}\quad\lambda(\sigma)=\tfrac{1}{2}\begin{pmatrix}-1&\sqrt{3}\\-\sqrt{3}&-1\end{pmatrix}$$

correspond geometrically to reflection about the direction E_2 and rotation through the angle $2\pi/3$, respectively.

As was indicated in its construction, R can be thought of as the linear version of the definition of $\mathcal{G}$. However, not all irreducible representations appear in R, and there is a larger representation $\mathfrak{R}$ of $\mathcal{G}$ which more fully captures the structure of $\mathcal{G}$. Thus, let $C(\mathcal{G})$ be the six-dimensional space of all

complex-valued functions F on $\mathcal{G}$, and for $g \in \mathcal{G}$ define

$$(\mathfrak{R}(g)F)(g_1) = F(g_1 g) \tag{2.5}$$

for $F \in C(\mathcal{G})$. The representation $\mathfrak{R}$ is called the *right regular representation* of $\mathcal{G}$ on $C(\mathcal{G})$. Of course, it would be equivalent to work with the *left* regular representation $\mathfrak{L}$ defined on $C(\mathcal{G})$ by

$$(\mathfrak{L}(g)F)(g_1) = F(g^{-1}g_1). \tag{2.6}$$

We remark that the term "equivalent" has a technical meaning in the subject of group representations that is quite analogous to the notion of "similarity" in linear algebra. Two representations T_1 and T_2 of a finite group G are said to be *equivalent*, written $T_1 \cong T_2$, if there exists an invertible linear transformation $A: V_{T_1} \to V_{T_2}$ such that $AT_1(g)A^{-1} = T_2(g)$ for all $g \in G$. Alternatively, $T_1 \cong T_2$ if and only if there are bases of V_{T_1} and V_{T_2} relative to which the matrix-valued realizations of T_1 and T_2 are identical. For all intents and purposes, the abstract theory of group representations does not distinguish equivalent representations, and the basic concepts (e.g., invariant subspace, irreducibility, direct sum, etc.) apply to the equivalence classes as well as the representations themselves.

The harmonic analysis of $C(\mathcal{G})$ is easily accomplished with the help of the inner product

$$\langle F_1 | F_2 \rangle = \sum_{g \in \mathcal{G}} F_1(g)\overline{F_2(g)} \tag{2.7}$$

on $C(\mathcal{G})$, our knowledge of $C(S)$, and the following elementary but extremely powerful result known as SCHUR'S LEMMA. *Suppose T_1 and T_2 are irreducible representations of a group G and that A is a linear transformation from V_{T_1} to V_{T_2} such that $AT_1(g) = T_2(g)A$ for all $g \in G$. If T_1 and T_2 are inequivalent, then $A = 0$. If $T_1 \cong T_2$, then either $A = 0$ or A is invertible and unique up to a constant factor. In particular, if $T_1 = T_2$, then $A = cI$ for some complex number c.* It follows from Schur's Lemma that the matrix entries of inequivalent representations are orthogonal relative to (2.7).

The proof of Schur's Lemma follows immediately from the fact that the null space and range of A are invariant under T_1 and T_2, respectively.

To begin, note that the signature function, defined by $\operatorname{sgn}(g) = \pm 1$ according to g even or odd, and the constant function 1 are both homomorphisms of $\mathcal{G}$ into the group $\{\pm 1\}$. Each spans a one-dimensional invariant subspace of $C(\mathcal{G})$. It remains to find four other orthogonal functions.

Thus, bring λ into the picture in terms of the subspace $C_\lambda(\mathcal{G})$ of $C(\mathcal{G})$ spanned by the matrix entries $\lambda_{11}, \lambda_{12}, \lambda_{21}, \lambda_{22}$ relative to a choice of basis in W_2 (e.g., E_1 and E_2 in Fig. 2). It is not difficult to show (either by brute verification or a general argument based upon Schur's Lemma) that $\lambda_{11}, \lambda_{12}, \lambda_{21}$, and λ_{22} are mutually orthogonal, and $C_\lambda(\mathcal{G})$ is a four-dimensional invariant subspace. Moreover, although $C_\lambda(\mathcal{G})$ is not irreducible, the next best situation occurs. Namely, the subrepresentation of $\mathfrak{R}$ on $C_\lambda(\mathcal{G})$ is equivalent to $\lambda \oplus \lambda$. For if $C_\lambda^{(1)}(\mathcal{G})$ and $C_\lambda^{(2)}(\mathcal{G})$ are the two-dimensional subspaces of $C_\lambda(\mathcal{G})$ having bases $\{\lambda_{11}, \lambda_{12}\}$ and $\{\lambda_{21}, \lambda_{22}\}$, respectively, then these subspaces are invariant under $\mathfrak{R}$, $C_\lambda(\mathcal{G}) = C_\lambda^{(1)}(\mathcal{G}) \oplus C_\lambda^{(2)}(\mathcal{G})$, and by a straightforward calculation in the indicated bases, both subrepresentations on $C_\lambda^{(1)}(\mathcal{G})$ and $C_\lambda^{(2)}(\mathcal{G})$ are equivalent to λ.

Finally, we observe (again, either directly or from Schur's Lemma) that both 1 and sgn are orthogonal to $C_\lambda(\mathcal{G})$, and by a dimension count

$$C(\mathcal{G}) = \mathbf{C}1 \oplus \mathbf{C}\operatorname{sgn} \oplus C_\lambda(\mathcal{G}) \tag{2.8}$$

and

$$\mathfrak{R} \cong 1 \oplus \operatorname{sgn} \oplus 2\lambda \tag{2.9}$$

where we have set $\lambda \oplus \lambda = 2\lambda$.

In summary, *to within equivalence* 1, sgn, *and* λ *are the only irreducible representations of* $\mathcal{G}$, *and each appears in the right (or left) regular representation with multiplicity equal to its degree. In terms of* $C(\mathcal{G})$, *the matrix entries* 1, sgn, $\lambda_{11}, \lambda_{12}, \lambda_{21}, \lambda_{22}$ *of the irreducible representations are an orthogonal basis for* $C(\mathcal{G})$.

We have here the fundamental harmonic analysis—or, if you like, Fourier analysis—for the symmetric group $\mathcal{G}$. These results are specific examples from a general structure theory valid for any finite group G. The most prominent features are listed below.

1. *Unitarizability*. Given a representation T of G, there exists an inner product $\langle\cdot|\cdot\rangle$ on $V=V_T$ which is invariant under T; i.e., $\langle T(g)v|T(g)w\rangle=\langle v|w\rangle$ for all $g\in G$ and $v,w\in V$. To wit, if $(\cdot|\cdot)$ is any inner product on V, then the desired inner product is given by $\langle v|w\rangle=\sum_{g\in G}(T(g)v|T(g)w)$. In short, T can be assumed *unitary*; i.e., $T(g)$ is a unitary operator for each $g\in G$. In particular, $T(g)^*=T(g^{-1})$ for all $g\in G$. It follows that the orthogonal complement $W^\perp$ of an invariant subspace W is also invariant; and hence we have:

2. *Complete reducibility*. A representation T decomposes as an (orthogonal) direct sum of irreducible representations. Moreover, the irreducible representations that appear are unique to within equivalence.

3. *Uniqueness of decomposition*. Denote by $\hat{G}$ the collection of all equivalence classes $[\lambda]$ of irreducible representations λ of G. One calls $\hat{G}$ the *dual* or *dual object* of G. A class $[\lambda]\in\hat{G}$ is said to *occur in* T if T has a subrepresentation equivalent to λ, and the *multiplicity* n_λ of $[\lambda]$ in T is the number of times the class $[\lambda]$ appears in a complete reduction of T. Let $\lambda_1,\lambda_2,\ldots,\lambda_k$ be representatives for the distinct classes in $\hat{G}$ that appear in T. Then for each $j=1,\ldots,k$ there is a *unique* invariant subspace V_{λ_j} of $V=V_T$, on which T is equivalent to n_{λ_j} copies of λ_j; the spaces V_{λ_j} are mutually orthogonal; and

$$V=V_{\lambda_1}\oplus\cdots\oplus V_{\lambda_k}. \tag{2.10}$$

Equation (2.10) gives the *primary decomposition* of V, and the representation-theoretic version

$$T\cong n_1\lambda_1\oplus\cdots\oplus n_k\lambda_k \tag{2.11}$$

is the *primary decomposition* of T. The space V_{λ_j} is called the $[\lambda_j]$-*primary*, or $[\lambda_j]$-*isotypic, subspace* of V; or alternatively, *the subspace of T that transforms according to λ_j*.

4. *Schur orthogonality relations*. Let $\mathrm{C}(G)$ be the vector space of all functions $F:G\to\mathbf{C}$. The dimension of $\mathrm{C}(G)$ is n, where n is the order of G. Equip $\mathrm{C}(G)$ with the normalized inner product

$$\langle F_1|F_2\rangle=n^{-1}\sum_{g\in G}F_1(g)\overline{F_2(g)}. \tag{2.12}$$

For each class $[\lambda]\in\hat{G}$, choose a representation λ which is unitary, and let λ_{ij} denote the matrix entries relative to an orthonormal basis for V_λ. Then the Schur orthogonality relations (which follow directly from Schur's Lemma) state that the functions $e_{ij}^{(\lambda)}(g)=\sqrt{d_\lambda}\,\lambda_{ij}(g)$ on G, for $[\lambda]\in\hat{G}$ and $1\leqslant i, j\leqslant d_\lambda=\deg\lambda$, are an orthonormal subset of $\mathrm{C}(G)$. That this orthonormal set is complete is a consequence of the:

5. *Decomposition of the regular representation*. Let $\mathfrak{R}$ be the right regular representation of G on the space $C(G)$. Then each class $[\lambda]\in\hat{G}$ occurs in $\mathfrak{R}$ with multiplicity $d_\lambda=\deg\lambda$, and the $[\lambda]$-primary subspace of $C(G)$ is the space $C_\lambda(G)$ spanned by the matrix entries λ_{ij}. In short

$$C(G)=\sum_{[\lambda]\in\hat{G}}\oplus\, C_\lambda(G) \tag{2.13}$$

and

$$\mathfrak{R}\cong\mathfrak{L}\cong\sum_{[\lambda]\in\hat{G}}\oplus\, d_\lambda\lambda. \tag{2.14}$$

Note that by a dimension count, the order of G is related to the degrees of the irreducible representations by the formula

$$n=\sum_{[\lambda]\in\hat{G}}d_\lambda^2. \tag{2.15}$$

6. *The group algebra*. For $F_1, F_2 \in C(G)$, define their *convolution* by

$$(F_1 * F_2)(g) = \sum_{g_1 \in G} F_1(g_1) F_2(g_1^{-1} g). \tag{2.16}$$

With convolution as multiplication, $C(G)$ has the structure of an associative algebra, called the *group algebra* of G. From this point of view, the subspaces $C_\lambda(G)$, for $[\lambda] \in \hat{G}$, are two-sided ideals, and (2.13) is the decomposition of the group algebra into its minimal two-sided ideals. In particular, the representation theory of G is a special case of the Wedderburn theory of semi-simple rings.

In a common variant of the group algebra, the function $F \in C(G)$ is written as a formal sum, which in our case has the nonstandard form $\sum_{g \in G} F(g) g^{-1}$. By an elementary calculation, it is easy to see that in this formalism convolution is the "natural" multiplication

$$\left(\sum_{g_1 \in G} F_1(g_1) g_1^{-1}\right)\left(\sum_{g_2 \in G} F_2(g_2) g_2^{-1}\right) = \sum_{g_1 \in G} \sum_{g_2 \in G} (F_1(g_1) F_2(g_2)) g_1^{-1} g_2^{-1}.$$

7. *Character theory*. The character of a representation T is the complex-valued function $\chi_T(g) = tr(T(g))$ for $g \in G$. The importance of the notion of character, fundamental to the original work by Frobenius and Schur as well as to the modern point of view, lies in the fact that two representations are equivalent if and only if they have the same character. In other words, the trace is a complete invariant for the equivalence relation.

Finally, there is a Fourier analytic overview to the representation theory of G. That is the theme of the next section.

3. Compact harmonic analysis: The Peter–Weyl Theorem. Modern harmonic analysis begins in the 1920's with the confluence of two great streams of mathematical thought which had previously been developing concurrently but independently.

First, there is the role of group theory, rooted in large part in geometry through the philosophy (Klein's Erlanger program) of studying a space through its group of motions. In particular, the theory of geometric transformation groups, known today as *Lie groups* in honor of the Norwegian mathematician Sophus Lie, who created the subject over a century ago, arose from certain kinds of partial differential equations in rough analogy to the origin of Galois groups from algebraic equations.

On the other hand, a variety of purely analytic theories were developing in the late nineteenth century from very different aspects, such as boundary value problems, of differential equations. For example, if in the vibrating string problem described in Section 1 we do not assume that the tension and density are uniform across the string, then the wave equation (1.1) is replaced by the equation

$$\frac{\partial}{\partial x}\left(T(x)\frac{\partial u}{\partial x}\right) = \rho(x)\frac{\partial^2 u}{\partial t^2}. \tag{3.1}$$

The principle of superposition of solutions applied to more general equations such as (3.1) leads to eigenfunction expansions more general than Fourier series, to Sturm-Liouville theory, to spectral theory on Hilbert space, and ultimately to modern abstract functional analysis.

The genius for bringing together these two seemingly unrelated themes belongs to Hermann Weyl, who should be regarded as the father of modern harmonic analysis. The date of birth is 1927, and the official birth certificate is the remarkable paper **[14]** by Peter and Weyl, in which the structure theory for the representations of a finite group is carried over, essentially without change, to the context of compact Lie groups. For it turns out that it is not the finiteness of the group on which the validity of the properties (1)–(7) hinges, but rather on the existence of an averaging procedure over the group. That is to say, what is required is an *invariant integral* which assigns *finite volume* to the group. The decisive analytic idea, originating with Peter and Weyl and fundamental to essentially all the succeeding developments in noncommutative harmonic analysis, is the use of an *infinite dimensional*

representation and its decomposition by means of spectral theory for bounded operators on Hilbert space.

F. Peter was a schoolteacher who worked under Weyl for a short time. Upon completing this joint paper he apparently decided that the research life was not for him, and returned to high school teaching. Needless to say, it was a very good high school.

A *topological group* is a group G on which is given a topology with respect to which the group multiplication $(x,y)\to xy$ and inversion $x\to x^{-1}$ are continuous mappings. Clearly, each translation (e.g., right translation $x\to xa$ by a) is a homeomorphism of G, so the topology of G is completely determined by local behavior at the identity e. Thus, one says that G is *locally compact* if there exists a compact neighborhood of e; and still more restrictedly, G is *locally Euclidean* if there exists a neighborhood of e which is homeomorphic to an open subset of some Euclidean space $\mathbf{R}^n$. The most important topological groups are the *Lie groups*. These are locally Euclidean groups in which the group operations are infinitely differentiable mappings. For a long time it was an open question—*Hilbert's fifth problem*—as to whether a locally Euclidean group is necessarily a Lie group. That is to say, can continuity be replaced by differentiability? The affirmative resolution was given by Gleason, Montgomery, and Zippin in the 1950's.

As the terminology suggests, a topological group G, whether or not it is a Lie group, is called *compact* if G itself is a compact set. It is to the class of compact groups that the results of this section apply.

The first prerequisite for harmonic analysis is the availability of an *invariant integral* on the group. When dealing with noncommutative groups, an author is faced with the decision—completely arbitrary, of course—as to the use of left versus right translation. For reason of habit, we choose the side of right, and for the most part leave the sinister unsaid.

Throughout, we shall be extremely casual in our approach to integration. The reader familiar with measure theory will know what is meant by an "integral." Those less experienced will find it adequate to simply think of an invariant integral as a generalization to G of the ordinary Riemann integral on the real line.

An integral on the topological group G is said to be *right invariant* if

$$\int_G f(xa)dx = \int_G f(x)dx \tag{3.2}$$

for all $a\in G$. For a Lie group, it is rather easy to construct a right invariant integral. For locally compact groups, the existence of a right invariant integral was proved by Haar in 1933. Then von Neumann immediately established the uniqueness, up to a positive constant factor, of such an integral and obtained a number of important consequences.

Fix a right invariant integral on the locally compact group G. By its uniqueness, for each $a\in G$ there is a positive constant $\delta(a)$ defined by the identity $\int f(ax)dx=\delta(a)\int f(x)dx$. Then $\delta: G\to\mathbf{R}^+$ is a continuous homomorphism, called the *modular function* of G, which relates left and right invariance as well as inversion. Specifically, the formula $\int f(x)\delta(x)dx$ defines the corresponding left invariant integral, and $\int f(x)\delta(x)dx=\int f(x^{-1})dx$.

The group G is called *unimodular* if $\delta(a)=1$ for all a. Obviously, G is unimodular if and only if there exists an integral which is simultaneously left and right invariant, in which case it is also invariant under inversion. Evidently a compact group is unimodular, for the multiplicative group $\mathbf{R}^+$ has no compact subgroups other than $\{1\}$.

We note that in the same paper von Neumann used the existence of an invariant integral, together with the Peter–Weyl theorem, to give an elementary and short proof of Hilbert's fifth problem for compact groups.

Finally, in his celebrated book **[15]** in 1938, Weil provided a converse. Namely, local compactness is implied by the existence of a right invariant integral. In honor of its discoverer, an invariant integral is called a *Haar integral*.

Now fix a compact group G. From the compactness, it follows that G has *finite volume*. Therefore, we can assume that the invariant integral is so normalized that the volume of G is one; i.e.,

$$\mathrm{vol}(G)=\int_G 1\,dx=1. \tag{3.3}$$

For example, if $G=\mathbf{T}$, the circle group, then

$$\int_G f(x)dx=(2\pi)^{-1}\int_0^{2\pi} f(\theta)d\theta;$$

and if G is a finite group of order n, then

$$\int_G f(x)dx=n^{-1}\sum_{g\in G} f(g).$$

Armed with the invariant integral, motivated by the Fourier analysis of Section 1, and aware of the representation-theoretic facts in Section 2, we now describe the Peter–Weyl theory.

In view of the topological nature of G, we hereafter require all representations to be *continuous*; that is to say, the matrix entries are to be continuous functions on G. Let the dual object $\hat{G}$ be defined as in Section 2, and for each class $[\lambda]\in\hat{G}$ fix once and for all a unitary matrix-valued representation λ belonging to that class. Thus, for each x in G, $\lambda(x)$ is a $d_\lambda\times d_\lambda$ matrix, where $d_\lambda=\deg\lambda$.

THE PETER–WEYL THEOREM. *The normalized matrix entries* $e_{ij}^{(\lambda)}(x)=\sqrt{d_\lambda}\,\lambda_{ij}(x)$ for $[\lambda]\in\hat{G}$ and $1\leqslant i,j\leqslant d_\lambda$, *form a complete orthonormal system relative to the inner product*

$$\langle f_1|f_2\rangle=\int_G f_1(x)\,\overline{f_2(x)}\,dx. \tag{3.4}$$

The structure theory described for finite groups in Section 2 can now be transferred more or less directly to the context of compact groups. That is to say, there are complete analogs of properties (1) through (7), in which summation over a finite group is supplanted by integration over the compact group.

There are a few distinct features that ought to be mentioned when G is compact but not finite. The first concerns the introduction of infinite-dimensional spaces.

A *Hilbert space* is defined to be a (complex) vector space with inner product, which is complete relative to the norm determined from the inner product. In the case at hand, the vector space $\mathcal{C}(G)$ of all *continuous* functions on G with inner product (3.4) is not complete. Its completion is the Hilbert space $L^2(G)$ of all (Lebesgue) square-integrable functions on G. Thus, to be technically correct in the transition from finite groups to compact groups, one should replace $\mathcal{C}(G)$ by $L^2(G)$. Of course, $L^2(G)$ is infinite dimensional (if G is not finite), and the regular representations $\mathfrak{R}$ and $\mathfrak{L}$ are *infinite dimensional unitary representations* of G on the space $V=L^2(G)$. This is the highly original step taken by Peter and Weyl. Namely, they replaced the counting argument (cf. formula (2.15)) for a finite group by a spectral theoretic analysis of the infinite-dimensional representation space $L^2(G)$. Thus, in analogy to (2.14),

$$L^2(G)=\sum_{[\lambda]\in\hat{G}}\oplus\ \mathcal{C}_\lambda(G)$$

where each irreducible representation λ of G is finite dimensional, but in general $\hat{G}$ is an infinite set. Then (2.14) is valid as it stands; that is, each irreducible representation appears in the regular representation with multiplicity equal to its degree.

For the reader who has some experience with spectral theory on Hilbert space, we mention the simple idea at the heart of the proof of the Peter–Weyl Theorem. For f_1 and f_2 in $L^2(G)$, their convolution is given by

$$(f_1*f_2)(x)=\int_G f_1(y)f_2(y^{-1}x)dy$$

(cf. (2.16)). The operator $\mathfrak{L}(f)$ on $L^2(G)$ defined by $\mathfrak{L}(f)h=f*h$ for $h\in L^2(G)$, can be formally written as $\mathfrak{L}(f)=\int f(y)\mathfrak{L}(y)dy$. This operator $\mathfrak{L}(f)$ is called *left convolution* by f. Now, $\mathfrak{R}(a)\mathfrak{L}(f)=\mathfrak{L}(f)\mathfrak{R}(a)$ for all $a\in G$, so the primary subspaces of the right regular representation coincide with the eigenspaces of $\mathfrak{L}(f)$.

The Peter–Weyl Theorem is then readily reduced to the fact that the operators $\mathfrak{L}(f)$ are completely continuous.

The Peter–Weyl Theorem can be rephrased in terms of *Fourier analysis*. The Fourier series of a function f on G is defined as

$$f(x)=\sum_{[\lambda]\in\hat{G}} d_\lambda \sum_{i,j=1}^{d_\lambda} \hat{f}_{ij}(\lambda)\lambda_{ij}(x) \tag{3.5}$$

where the numbers

$$\hat{f}_{ij}(\lambda)=\langle f|\lambda_{ij}\rangle=\int_G f(x)\,\overline{\lambda_{ij}(x)}\,dx \tag{3.6}$$

are the *Fourier coefficients* of f. The series (3.5) applies to those functions f which are square-integrable in that the norm

$$\|f\|=\left\{\int_G |f(x)|^2 dx\right\}^{1/2} \tag{3.7}$$

is finite; and equality in (3.5) is in the mean-square sense of

$$\|f\|^2=\sum_{[\lambda]\in\hat{G}} d_\lambda \sum_{ij=1}^{d_\lambda} |\hat{f}_{ij}(\lambda)|^2. \tag{3.8}$$

Formula (3.8) is called the *Plancherel formula* for G.

Of course, when $G=\mathbf{T}$ is the circle group, the dual object is just the dual group of functions e_n, and formulas (3.5), (3.6), (3.7), and (3.8) reduce to the Fourier series formulas (1.7), (1.8), (1.10), and (1.11), respectively.

More generally, let G be any compact abelian group. Then by Schur's Lemma the irreducible representations are all one-dimensional, and the dual object $\hat{G}$ is the *dual group of* all continuous homomorphisms λ of G into $\mathbf{T}$. Then the Fourier coefficients $\hat{f}(\lambda)=\int f(x)\lambda(x)dx$ are numbers, and the Fourier series (3.5) is more simply given by $f(x)=\sum_{\lambda\in\hat{G}}\hat{f}(\lambda)\lambda(x)$.

Since the Plancherel formula is exceedingly important in harmonic analysis, we offer some further elaboration. Let $L^2(G)$ denote the vector space of square integrable functions on G. Motivated by formula (3.6) we define the $d_\lambda\times d_\lambda$ matrix

$$\hat{f}(\lambda)=\int_G f(x)\,\overline{\lambda(x)}\,dx, \tag{3.9}$$

the matrix entries of which are the Fourier coefficients. Then the Fourier series (3.5) takes the form

$$f(x)=\sum_{[\lambda]\in\hat{G}} d_\lambda tr(\hat{f}(\lambda)\lambda(x)^t) \tag{3.10}$$

and the Plancherel formula (3.8) is more simply written

$$\|f\|^2=\sum_{[\lambda]\in\hat{G}} \|\hat{f}(\lambda)\|^2_{d_\lambda}, \tag{3.11}$$

where the space $\mathbf{C}^{d_\lambda x d_\lambda}$ of all $d_\lambda\times d_\lambda$ matrices is given the inner product

$$\langle A|B\rangle_{d_\lambda}=d_\lambda tr(AB^*)=d_\lambda\sum_{ij=1}^{d_\lambda} a_{ij}\overline{b_{ij}},$$

and the corresponding norm $\|A\|^2_{d_\lambda}=d_\lambda tr(AA^*)$. Form the orthogonal direct sum

$$L^2(\hat{G})=\sum_{[\lambda]\in\hat{G}}\oplus\mathbf{C}^{d_\lambda\times d_\lambda}. \tag{3.12}$$

Then (3.11) implies the noncommutative version of the PLANCHEREL THEOREM: *The mapping* $\mathcal{F}: f \to \hat{f}$ *is a unitary transformation, the Plancherel transform, of* $L^2(G)$ *onto* $L^2(\hat{G})$. In particular, the Plancherel Theorem shows that noncommutativity in harmonic analysis resides in the noncommutativity of matrix multiplication.

The importance of the Plancherel transform lies in the fact that it quite explicitly gives the decomposition of the right (as well as left) regular representation $\mathcal{R}$. Indeed, the equivalent representation $\hat{\mathcal{R}}$ on $L^2(\hat{G})$, defined by $\hat{\mathcal{R}}(a) = \mathcal{F}\mathcal{R}(a)\mathcal{F}^{-1}$, is easily seen to be described in terms of right multiplication by $\lambda(a)^t$. Specifically, $(\hat{\mathcal{R}}(a)\hat{f})(\lambda) = \hat{f}(\lambda)\lambda(a)^t$.

Examples: As with all abstract mathematical theories, it is important to analyze the examples to which the abstraction applies and to enrich the theory by any additional structures which may be present. The Peter–Weyl Theory is a case in point, for it can only be fully appreciated when combined with the rich algebraic, geometric, and combinatoric processes that collectively comprise the Cartan–Weyl representation theory of the compact *Lie groups*. Now, there exists a complete structure theory for the compact Lie groups. Very roughly, a connected compact Lie group splits as a product, the individual factors of which are either abelian or "simple," and these are listed below.

1. A connected compact *abelian* Lie group is necessarily a *torus*, by which is meant a direct product of circles.
2. The *special orthogonal group* $SO(n)$ consists of all $n \times n$ real matrices g of determinant 1 such that $gg^t = 1$. In other words, $SO(n)$ is the connected component of the identity matrix in the group of all matrices that preserve the natural inner product on $\mathbf{R}^n$.
3. The *special unitary group* $SU(n)$ is the analog over the complex field of $SO(n)$. It consists of all complex $n \times n$ matrices of determinant 1 such that $gg^* = 1$.
4. The *compact symplectic group* $Sp(n)$ is the analog over the quaternionic division algebra $\mathbf{H}$ of $SO(n)$. Specifically, $Sp(n)$ consists of all $n \times n$ matrices over $\mathbf{H}$ which preserve the natural (real) inner product on $\mathbf{H}^n$.
5. The groups $SO(n)$, $SU(n)$, and $Sp(n)$ form the three families of so-called *classical* compact simple groups. They are *simple* in the sense that modulo a finite group (the center) there are no normal subgroups. Aside from these classical groups there are only five other compact simple groups, denoted E_6, E_7, E_8, F_4, and G_2. They are referred to as the *exceptional* compact simple groups, and have realizations in terms of certain nonassociative algebras such as the Cayley numbers.

It is beyond the scope of this elementary introduction to go into the explicit nature of the harmonic analysis of the compact Lie groups (although we touch upon it in the next section). Suffice it to say that the irreducible representations were constructed algebraically, or "infinitesimally," by Cartan, and the formulas for the characters were calculated by Weyl, who also constructed the representations "globally." Although all of this was known over fifty years ago, it is still of importance today to obtain new realizations of the representations of compact Lie groups that bear upon various aspects of analysis, geometry, arithmetic, and mathematical physics.

Compact groups which are not Lie groups arise by suitable "limits" from Lie groups. In particular, there are a variety of p-adic analogs of the compact simple groups, but the representation theory and harmonic analysis of such groups is at present highly incomplete.

4. Harmonic analysis and special functions: Spherical harmonics. The classical treatises on the special functions of mathematical physics—for example, the five-volume Bateman manuscript or Watson's encyclopedic *Treatise on Bessel Functions*—provide imposing monuments to the cleverness of our forebears. To the purist they are often impenetrable, and their use may represent a turn of events born of desperation. For there one finds what seems to be a tangled chaotic jungle of series and integral representations, recurrence relations, difference and differential equations, and assorted variations for a myriad of higher transcendental functions, if ruled by reason, clearly not by order.

On the other hand, among physicists, engineers, and those with an applied bent, there is a familiarity that can effect a great uplifting of the spirit. A vivid illustration is given by my erudite colleague and friend W. J. Holman, III, who upon breaking a problem in complex analysis on a symmetric space through esoteric methods in special functions exclaimed: "Is it not sublime! Meditate it until you can feel its exquisite diathrodial ginglimus in every joint of your body." An uncommon group theoretical allusion, *diathrodial ginglimus* refers to that property of the jaw, unique among all the joints of the body, that it can be both rotated and translated.

Nonetheless, from the standpoint of harmonic analysis, there is an orderly overview to a sizable part of the literature on special functions. This approach is founded upon the Peter–Weyl Theory, and originates in 1929 with E. Cartan's determination of the special functions associated with compact Riemannian symmetric spaces. We give two examples illustrative of the general theory. The first, the *harmonics of the circle*, is simply a rephrasing of the Fourier series considerations from Section 1, and is presented for motivation. The second is a group-theoretical description of *spherical harmonics*.

EXAMPLE 1. We view the circle in two different ways. First, it is the 1-dimensional unit sphere S^1 of points $x=(x_1,x_2)$ in the Euclidean plane $\mathbf{R}^2$ having unit norm. Second, it is the rotation group $SO(2)$ of matrices

$$g=g(\theta)=\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \tag{4.1}$$

acting on $\mathbf{R}^2$ by the formula

$$xg=(x_1\cos\theta-x_2\sin\theta, x_1\sin\theta+x_2\cos\theta). \tag{4.2}$$

Of course, the action of $SO(2)$ on $\mathbf{R}^2$ restricts to the subset S^1, for S^1 is the *orbit* under $SO(2)$ of the "north pole" $\mathbf{1}=(1,0)$. That is to say, S^1 consists precisely of the points $\mathbf{1}g=(\cos\theta,\sin\theta)$. Let R denote the regular representation

$$(R(g)f)(x)=f(xg) \tag{4.3}$$

of $SO(2)$ on the Hilbert space $L^2(S^1)$ of square-integrable functions on S^1. More familiarly, $(R(\theta)f)(\phi)=f(\phi+\theta)$ where $x=(\cos\phi,\sin\phi)\in S^1$ and $g=g(\theta)\in SO(2)$.

Now, the two-dimensional Laplacian $\nabla^2=\partial^2/\partial x_1^2+\partial^2/\partial x_2^2$ can be rewritten in polar coordinates as

$$\nabla^2=\frac{\partial^2}{\partial r^2}+\frac{1}{r}\frac{\partial}{\partial r}+\frac{1}{r^2}\frac{\partial^2}{\partial\theta^2} \tag{4.4}$$

from which one sees the justification in referring to the operator $\Delta=d^2/d\theta^2$ as the Laplacian on the circle. Namely, it is the "circular part" of ∇^2. By the results in Section 1, the Hilbert space $L^2(S^1)$ decomposes as the orthogonal direct sum

$$L^2(S^1)=\sum_{n=0}^{\infty}\oplus\ H_n \tag{4.5}$$

of the eigenspaces H_n of the Laplacian. Recall that H_0 is the 1-dimensional null-space of Δ and contains the constant functions; whereas for $n\geqslant 2$, H_n is the 2-dimensional eigenspace corresponding to the eigenvalue $-n^2$ of Δ and having the functions $\cos n\theta$ and $\sin n\theta$ as basis. *From this point of view, H_n could be called the space of circular harmonics of order n, and the variant*

$$f(\theta)=a_0+\sum_{n=1}^{\infty}(a_n\cos n\theta+b_n\sin n\theta) \tag{4.6}$$

of the Fourier series (1.7) *is the orthogonal expansion of f in the special functions associated to the action of $SO(2)$ on S^1*. The precise relationship with the representation theory of $SO(2)$ is expressed by the commutation relation

$$R(g)\Delta=\Delta R(g) \tag{4.7}$$

for $g\in SO(2)$.

EXAMPLE 2. The preceding example has a higher dimensional generalization to the $(k-1)$-dimensional unit sphere S^{k-1} in the Euclidean space $\mathbf{R}^k$. For simplicity, we restrict to the case $k=3$.

The 2-sphere S^2 is not a group. However, it is a *homogeneous space*. In fact, $S^2 \cong SO(3)/SO(2)$. More precisely,

$$S^2 = K \backslash G \tag{4.8}$$

that is, the space of right cosets Kg, where $G = SO(3)$ is the rotation group on $\mathbf{R}^3$ and K is the subgroup, isomorphic to $SO(2)$, of matrices

$$k = k(\theta) = \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{4.9}$$

In finer detail, $G = SO(3)$ is the group of proper isometries of $\mathbf{R}^3$; i.e., the 3×3 matrix g is in G if and only if $\|xg\| = \|x\|$ for all $x \in \mathbf{R}^3$ (alternatively, $gg^t = 1$) and $\det g = 1$. Here, $\|x\| = (x_1^2 + x_2^2 + x_3^2)^{1/2}$ is the norm of the point $x = (x_1, x_2, x_3) \in \mathbf{R}^3$. Then the unit sphere $S^2 = \{x \in \mathbf{R}^3 : \|x\| = 1\}$ is the *orbit* of the north pole $\mathbf{1} = (0,0,1)$ under G; i.e., $S^2 = \{\mathbf{1}g : g \in G\}$. The subgroup of G that leaves the point $\mathbf{1}$ fixed (alternatively, the subgroup of rotations about the x_3-axis) is easily seen to be the above group K, and the mapping $x = \mathbf{1}g \to Kg$ identifies S^2 with the space $K\backslash G$ of right cosets Kg in such a way that the action by rotation of G on S^2 corresponds to the natural right action of G on $K\backslash G$.

Next, the 3-dimensional Laplacian $\nabla^2 = \partial^2/\partial x_1^2 + \partial^2/\partial x_2^2 + \partial^2/\partial x_3^2$ can be rewritten in spherical coordinates as

$$\nabla^2 = \frac{\partial^2}{\partial r^2} + \frac{2}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\left[\frac{1}{\sin\phi}\frac{\partial}{\partial\phi}\left(\sin\phi\frac{\partial}{\partial\phi}\right) + \frac{1}{\sin^2\phi}\frac{\partial^2}{\partial\theta^2}\right] \tag{4.10}$$

where $r = \|x\|$ and ϕ and θ are latitudinal and longitudinal coordinates on S^2, respectively. In particular, *the Laplacian on the sphere* is

$$\Delta = \frac{1}{\sin\phi}\frac{\partial}{\partial\phi}\left(\sin\phi\frac{\partial}{\partial\phi}\right) + \frac{1}{\sin^2\phi}\frac{\partial^2}{\partial\theta^2}. \tag{4.11}$$

As in the example of the circle, Δ is a negative-definite self-adjoint differential operator on S^2, and the Hilbert space $L^2(S^2)$ of square-integrable functions on the sphere (relative to the usual surface integral) decomposes as the orthogonal direct sum

$$L^2(S^2) = \sum_{n=0}^{\infty} \oplus\ H_n \tag{4.12}$$

of the eigenspaces H_n of Δ. Here, the eigenvalue of Δ for H_n is $-n(n+1)$, and $\dim H_n = 2n+1$.

The decomposition (4.12) also admits a group-theoretic interpretation, for Δ commutes with the right regular representation R of $SO(3)$ on $L^2(S^2)$ in complete analogy to formulas (4.3) and (4.7). This dual relationship can be summed up as follows:

Formula (4.12) *gives both the spectral decomposition of the Laplacian on the sphere as well as the primary decomposition of the regular representation of* $SO(3)$ *on* $L^2(S^2)$. *In particular,* H_n *is both an eigenspace of* Δ, *and an irreducible invariant subspace of* $L^2(S^2)$. *The functions in* H_n *are called the spherical harmonics of degree* n, *and they are the special functions associated to the action of* $SO(3)$ *on* S^2.

Denote by $\lambda^{(n)}$ the subrepresentation of R on the space H_n. Then $\lambda^{(n)}$ is an irreducible representation of $SO(3)$ of degree $2n+1$. These representations are all inequivalent and, in fact, exhaust the dual object $SO(3)\hat{}$. Now, $SO(3)$ is homeomorphic to 3-dimensional real projective space, so the matrix entries $(2n+1)^{1/2}\lambda_{ij}^{(n)}$ form the Peter–Weyl basis for the Hilbert space of square-integrable functions on the projective space. For the general case $k>3$, the analogous representations $\lambda^{(n)}$ of $SO(k)$ (i.e., those that appear in $L^2(S^{k-1})$) are irreducible and inequivalent, but they comprise only a small part of the full dual $SO(k)\hat{}$.

The deeper aspects of the decomposition (4.12), and the connection with classical special functions, are revealed by a characteristic and most striking feature. Namely, in each space H_n of spherical harmonics there exists a particular function f_n, unique up to a constant factor, which is *invariant* under K; that is to say, $R(k)f_n = f_n$ for all $k \in K$. In other words, f_n is left fixed by all rotations about the x_3-axis; or equivalently, f_n is a function on the sphere which is constant on parallels of latitude. Classically, f_n is called the *zonal spherical function of degree n*, but in modern terminology it is known simply as a *spherical function*. The spherical function uniquely determines the entire space H_n, for the translates $R(g)f_n$ under $SO(3)$ span H_n.

We can now reveal the classical identity of the spherical functions. For from (4.11) and the fact that f_n is independent of θ (it depends only upon latitude ϕ), f_n satisfies the differential equation

$$\frac{d^2y}{d\phi^2} + (\cot\phi)\frac{dy}{d\phi} = -n(n+1)y \tag{4.13}$$

which upon the subsitution $t = \cos\phi$ becomes *Legendre's equation*

$$(1-t^2)\frac{d^2y}{dt^2} - 2t\frac{dy}{dt} + n(n+1)y = 0. \tag{4.14}$$

Thus,

$$f_n(\theta,\phi) = P_n(\cos\phi) \tag{4.15}$$

where P_n is the *Legendre polynomial* of degree n. It follows that the Legendre polynomials have a purely group theoretical formulation, so it should not be surprising that the wealth of classical analysis relating to these special functions is put into place by means of harmonic analysis.

To hint at the general theory, the previous example is a special case of a phenomenon that applies to certain Riemannian manifolds (compact symmetric spaces) S for which G is a transitive group of isometries, K is the subgroup leaving fixed a base point, and Δ is the Laplace–Beltrami operator on S. However, the connection with special functions persists even more generally; and as one varies the group, space, and action through the more familiar examples, one accounts for the theory of Bessel functions, certain other cases of the hypergeometric function, and Jacobi, Gegenbauer, Hermite, and Laguerre polynomials. (The spherical functions for the action of $SO(k)$ on $S^{k-1} \cong SO(k)/SO(k-1)$ are Gegenbauer polynomials.) For more complicated groups, the associated special functions need not be classical. For example, for the collection of (non-compact!) real semi-simple Lie groups, Harish-Chandra has calculated the spherical functions. To be sure, the Harish-Chandra spherical functions form only a small part of the special functions which are crucial to his profound and impressive theory of harmonic analysis.

5. Noncompact classical analysis: Fourier integrals. The foregoing exposition has dealt exclusively with harmonic analysis on *compact* domains. However, major emphasis during the past four decades has been placed upon the *noncompact* noncommutative theory, a subject which is fundamentally more difficult than its compact counterpart and which should be viewed realistically as occupying an extensive expanse of the current frontier in harmonic analysis.

It is to noncompact problems that we now turn, beginning with the classical commutative case. Even in this well-known and comparatively elementary context, the first of the two major difficulties in noncompact harmonic analysis arises. Namely, *the fundamental harmonics do not occur discretely and the function spaces do not decompose as direct sums. Rather, the harmonics form a continuum, and the natural function spaces decompose as continuous smearings, called in the technical jargon "direct integrals."*

Historically, the noncompact theory originates, as with the compact theory, in the partial differential equations of mathematical physics; in particular, with the heat equation for an object, such as a metal rod, of infinite extent. Once more, the crucial idea is due to Fourier, who, in 1811, replaced the series representation of a solution by an integral representation, and thereby initiated the study of *Fourier integrals*.

Formally, the *Fourier integral*, or *Fourier transform*, $\hat{f}$ of a function f on the real line is defined for all real numbers λ by

$$\hat{f}(\lambda)=(2\pi)^{-1/2}\int_{-\infty}^{\infty} f(x)e^{-i\lambda x}dx, \tag{5.1}$$

and f is recaptured from $\hat{f}$ by the *Fourier inversion formula*

$$f(x)=(2\pi)^{-1/2}\int_{-\infty}^{\infty} \hat{f}(\lambda)e^{i\lambda x}d\lambda. \tag{5.2}$$

Clearly, these formulas bear strong analogy to formulas (1.8) and (1.7), respectively, for Fourier coefficients and Fourier series of a function on the circle.

The constant $(2\pi)^{-1/2}$ in (5.1) appears as a matter of convenience, in that one avoids the appearance of a constant in the important formula (5.3) below. Group theoretically, this constant can be viewed as a suitable normalization of the Haar integral on $\mathbf{R}$.

The L^1-theory of the Fourier integral treats the case in which f is integrable over the real line. Then $\hat{f}$ is well defined. Nonetheless, the integrability of f does not imply the integrability of $\hat{f}$ (even in the Lebesgue sense), and generalized "methods of summability" are required to give rigorous sense to the integral in (5.2).

More to our purposes is the L^2-theory. For if f is both integrable and square-integrable (i.e., $f\in L^1(\mathbf{R})\cap L^2(\mathbf{R})$), then $\hat{f}$ is square-integrable. Moreover, in complete analogy to the case of Fourier series, equality in (5.2) holds in the mean-square sense and the *Plancherel formula*

$$\int_{-\infty}^{\infty} |\hat{f}(\lambda)|^2 d\lambda=\int_{-\infty}^{\infty} |f(x)|^2 dx \tag{5.3}$$

is valid. The L^2-theory is summarized by the PLANCHEREL THEOREM FOR THE REAL LINE: *The mapping $f\to\hat{f}$, originally defined on $L^1(\mathbf{R})\cap L^2(\mathbf{R})$, extends uniquely to a unitary operator $\mathcal{F}$, called the Plancherel transform, from the Hilbert space $L^2(\mathbf{R})$ to itself.*

Let us view the L^2-theory group-theoretically. The real line $\mathbf{R}$ under addition is a locally compact abelian group, and the regular representation R of $\mathbf{R}$ on $L^2(\mathbf{R})$, defined by

$$(R(a)f)(x)=f(x+a), \tag{5.4}$$

is unitary. Furthermore, the functions $e_\lambda(x)=e^{i\lambda x}$ exhaust the *dual group* of all continuous homomorphisms of $\mathbf{R}$ into the circle group $\mathbf{T}$. To facilitate the "spectral analysis" of R, we consider the equivalent Plancherel-transformed representation $\hat{R}$, given by $\hat{R}(a)=\mathcal{F}R(a)\mathcal{F}^{-1}$. A straightforward calculation shows that for any $a\in\mathbf{R}$

$$(\hat{R}(a)\hat{f})(\lambda)=e^{i\lambda a}\hat{f}(\lambda) \tag{5.5}$$

for all $\hat{f}\in L^2(\mathbf{R})$. Notice, in contrast to the compact case, that λ is a *continuous* variable. Consequently, formula (5.5) represents a *continuous decomposition* of $\hat{R}$ into its irreducible constituents e_λ. In symbols, we write

$$\hat{R}=\int_{\mathbf{R}}^{\oplus} e_\lambda d\lambda \tag{5.6}$$

(in analogy to the notation for a direct sum) and say that $\hat{R}$ is the *direct integral* of the one-dimensional representations e_λ. In particular, each specific value of λ is measure-theoretically irrelevant, so e_λ is not a subrepresentation of $\hat{R}$ in the usual sense. At best, it is an "infinitesimal" component. As a consequence, *we are led to conclude that $\hat{R}$, or equivalently the regular representation of $\mathbf{R}$, has no irreducible subrepresentations. Alternatively, the Hilbert space $L^2(\mathbf{R})$ of square-integrable functions on $\mathbf{R}$ possesses no irreducible invariant subspaces.*

As one might expect from previous considerations, the above group theory can be replaced by spectral theory of the Laplacian $\Delta=d^2/dx^2$ on $\mathbf{R}$. For Δ is negative-definite (and self-adjoint) and commutes with the regular representation. In this context, e_λ is an eigenfunction (specifically,

$\Delta e_\lambda = -\lambda^2 e_\lambda$), but these eigenfunctions are not square-integrable. Thus, we see that the noncompactness of $\mathbf{R}$ is manifested in a continuous spectrum for Δ, and the direct integral decomposition

$$L^2(\mathbf{R}) = \int_{\mathbf{R}}^{\oplus} \mathbf{C} e_\lambda \, d\lambda, \tag{5.7}$$

underlies the spectral decomposition

$$\Delta = \int_{\mathbf{R}}^{\oplus} (-\lambda^2) d\lambda \tag{5.8}$$

as well as the representation-theoretic formulation (5.6).

The theory of Fourier series and integrals has a generalization, due to Weil [15], to an arbitrary locally compact abelian group G. The *dual group* $\hat{G}$ consists of the continuous homomorphisms of G into the circle group $\mathbf{T}$, and $\hat{G}$ is itself a locally compact abelian group relative to pointwise multiplication as the law of composition and the compact-open topology. Then the Fourier transform

$$\hat{f}(\lambda) = \int_G f(x) \overline{\lambda(x)} \, dx \tag{5.9}$$

is defined for all $\lambda \in \hat{G}$, and the *inversion formula* takes the form

$$f(x) = \int_{\hat{G}} \hat{f}(\lambda) \lambda(x) d\lambda \tag{5.10}$$

where dx and $d\lambda$ are suitably normalized Haar integrals on G and $\hat{G}$, respectively. The classical theory of Fourier analysis—exclusive of the connection with differentiation—has an extension to this general context. Of course, when $G = \mathbf{T}$, the mapping $e_n \to n$ identifies $\hat{G}$ with the group $\mathbf{Z}$; when $G = \mathbf{R}$, the mapping $e_\lambda \to \lambda$ identifies $\hat{G}$ with $\mathbf{R}$ (so $\mathbf{R}$ is *self-dual*); and the formulas (5.9) and (5.10) reduce to those for Fourier series and Fourier integrals, respectively.

Finally, we can at least mention the *second* major difficulty in noncompact harmonic analysis, a complication that does *not* arise in the commutative theory. As we have indicated, the irreducible representations of a locally compact *abelian* group are all one-dimensional. For *compact* groups we have seen that the irreducible representations are finite-dimensional, and if the group is non-abelian they need not be of dimension 1. However, *for a group which is locally compact, but neither compact nor abelian, one is forced to deal with irreducible representations which are infinite-dimensional.*

6. Quantum mechanics and infinite-dimensional harmonic analysis. In our discussion of classical harmonic analysis, we have hinted at the interaction with classical physics. The most striking rapport is obtained with noncommutative harmonic analysis and quantum physics. For within a few years of the appearance in 1925 of Heisenberg's tour de force on "matrix mechanics," with which the floodgates were opened for mathematical quantum theory, the operator-theoretic basis for quantum mechanics was well established. Moreover, it had already been shown that the quantum numbers were parameters for representations of appropriate symmetry groups, and Weyl's classic treatment [13] of the connection with group theory had been published.

For our purpose of introducing infinite-dimensional representation theory, we shall examine the intimate relationship between the HEISENBERG UNCERTAINTY PRINCIPLE

$$PQ - QP = -iI \tag{6.1}$$

and the noncompact noncommutative group G of all 3×3 real matrices of the form

$$g(r_1, r_2, r_3) = \begin{bmatrix} 1 & 0 & 0 \\ r_2 & 1 & 0 \\ r_3 & r_1 & 1 \end{bmatrix}. \tag{6.2}$$

By reason of this association, G is known as the HEISENBERG GROUP.

Formula (6.1) can be given a brief mathematical description as follows. The physical states of a quantum mechanical system are to be viewed as vectors in an infinite-dimensional Hilbert space H,

and the observables are Hermitian linear transformations, typically unbounded, on H. In particular, for a dynamical system with one degree of freedom, there is the position observable Q and the momentum observable P which satisfy the fundamental commutation relation (6.1). This relation is interpreted physically as an expression of the uncertainty (we have normalized Planck's constant $\hbar$ to 1) inherent in simultaneous measurement of position and momentum. Now, Heisenberg realized P and Q as infinite matrices, the entries of which represented transitions between energy states. Shortly thereafter, Schrödinger, working with wave mechanics, realized the state space H (in modern terminology) as $L^2(\mathbf{R})$ and the operators P and Q as $-id/dx$ and multiplication by x, respectively. These competing mathematical models for quantum theory generated the obvious question: What are all possible solutions to the Heisenberg relation (6.1)? To this question Weyl, Stone, and von Neumann (independently) gave a *group theoretical* answer: Up to multiplicity and equivalence, Schrödinger's solution is unique. The group theory enters as follows.

In the above-mentioned book, Weyl observes that (6.1) can be converted into a group theoretical form if one replaces P and Q by the one-parameter groups of unitary operators that they generate, namely, $U(\sigma)=e^{i\sigma P}$ and $V(\tau)=e^{i\tau Q}$, respectively. Then *U and V are unitary representations of the additive group* $\mathbf{R}$ *on the Hilbert space H such that*

$$U(\sigma)V(\tau)U(\sigma)^{-1}V(\tau)^{-1}=e^{i\sigma\tau}I \tag{6.3}$$

for all $\sigma,\tau\in\mathbf{R}$. Formula (6.3), which is the "global version" of the Heisenberg relation, is called the *Weyl equation*. If we use script letters for Schrödinger's solution, so $(\mathcal{P}f)(x)=-if'(x)$ and $(\mathcal{Q}f)(x)=xf(x)$ for sufficiently smooth functions in $L^2(\mathbf{R})$, the corresponding solution of (6.3) is given by

$$(\mathcal{U}(\sigma)f)(x)=f(x+\sigma) \quad\text{and}\quad (\mathcal{V}(\tau)f)(x)=e^{i\tau x}f(x). \tag{6.4}$$

The term *global* is used in contrast to the term *infinitesimal*, the latter being referenced to differentiation. Thus, P and Q are recaptured from U and V by differentiation; specifically, $P=U'(0)$ and $Q=V'(0)$.

The uniqueness theorem mentioned above can now be stated with precision.

Let U and V be unitary representations of $\mathbf{R}$ *on a Hilbert space H such that* (6.3) *holds. Then H is the orthogonal direct sum of subspaces, say H_α, and there is a unitary mapping A from H_α to $L^2(\mathbf{R})$ such that*

$$AU(\sigma)A^{-1}=\mathcal{U}(\sigma) \quad\text{and}\quad AV(\tau)A^{-1}=\mathcal{V}(\tau) \tag{6.5}$$

for all $\sigma,\tau\in\mathbf{R}$, *where $\mathcal{U}$ and $\mathcal{V}$ are the specific representations given by* (6.4).

This result is sometimes called the Uniqueness Theorem for the Heisenberg Relations, but more often is referred to as the Stone–von Neumann Theorem. In short, the Schrödinger formulas (6.4) give the only representations of $\mathbf{R}$ that satisfy the Weyl equation, up to multiplicity and (unitary) equivalence.

We turn to the Heisenberg group G and its relationship to the previous theory. Notice from (6.2) that there are three *one-parameter subgroups* of G consisting of the matrices $u(r_1)=g(r_1,0,0)$, $v(r_2)=g(0,r_2,0)$, and $w(r_3)=g(0,0,r_3)$, respectively. Moreover, the former two subgroups generate all of G, for

$$g=w(r_3)v(r_2)u(r_1) \tag{6.6}$$

for all $g=g(r_1,r_2,r_3)\in G$, and

$$u(\sigma)v(\tau)u(\sigma)^{-1}v(\tau)^{-1}=w(-\sigma\tau) \tag{6.7}$$

for all $\sigma,\tau\in\mathbf{R}$.

A *one-parameter subgroup* of a topological group G is a homomorphism of $\mathbf{R}$ into G. Note that the one-parameter subgroup w is the center of G; i.e., $w(r)g=gw(r)$ for all $r\in\mathbf{R}$ and $g\in G$.

Evidently, the Weyl equation mirrors in operators the commutation relation (6.7) in G. In view of (6.6) it is natural to define

$$T(g)=e^{ir_3}\mathcal{V}(r_2)\mathcal{U}(r_1) \tag{6.8}$$

for $g \in G$, where $\mathcal{V}$ and $\mathcal{U}$ are given by (6.4). Now, from (6.2) and (6.4) one can easily verify that equation (6.8) defines a unitary representation T of G on the Hilbert space $L^2(\mathbf{R})$. Furthermore, *this representation illustrates the complication in noncompact noncommutative harmonic analysis alluded to at the end of the preceding section. For T is infinite-dimensional, but also irreducible.*

If T is an infinite-dimensional representation (i.e., dim $V=\infty$, where $V=V_T$), then we must refine the notion of irreducibility. Namely, T is irreducible if there are no proper *closed* invariant subspaces. In the situation at hand, although there are a number of *dense* subspaces of $L^2(\mathbf{R})$ that are invariant under T (e.g., the space of all continuous functions on $\mathbf{R}$ which vanish off some finite interval), $\{0\}$ and $L^2(\mathbf{R})$ itself are the only closed invariant subspaces.

From the basic representation T, one can construct a family of representations T_λ of G parametrized by all real numbers $\lambda \neq 0$. These are defined by the formula $T_\lambda(g)=T(g_\lambda)$ where $g_\lambda(r_1,r_2,r_3)=g(r_1,\lambda r_2,\lambda r_3)$ for all $g \in G$. In more detail,

$$(T_\lambda(g)f)(x)=e^{i\lambda(r_3+xr_2)}f(x+r_1) \tag{6.9}$$

for all $g \in G$ and $f \in L^2(\mathbf{R})$.

The mapping $g \to g_\lambda$ is an automorphism of G for each $\lambda \neq 0$. Alternatively, T_λ is that version of T which arises from the normalization $\hbar=\lambda$ of Planck's constant.

We can now restate the Stone–von Neumann Theorem as a characterization of the dual object $\hat{G}$ of the Heisenberg group: *With the exception of a set of* 1-*dimensional representations* (*which are substantially uninteresting in the harmonic analysis of G; cf. formulas* (6.10) *through* (6.13)), *all the irreducible representations of G are infinite-dimensional. More specifically, the representations* T_λ *are unitary, irreducible, mutually inequivalent, and to within equivalence exhaust all the* (*non* 1-*dimensional*) *irreducible unitary representations of G.*

Note that there is a special emphasis on *unitary* representations. In general, a noncompact group G will have representations (on a Hilbert space) which are not unitarizable. For such representations, there is no fully adequate theory of harmonic analysis. Hence, one usually restricts attention to unitary representations. In particular, $\hat{G}$ is the *unitary dual*, composed of the equivalence classes of irreducible unitary representations.

At this point, the better part of discretion might lead one to end the exposition. For to go further with the harmonic analysis of G becomes quite technical. However, the Heisenberg group is in a sense "almost abelian," and as such it is a relatively elementary noncompact nonabelian group. Thus, we cannot resist the temptation to illustrate the infinite-dimensional, operator-theoretic nature of noncompact harmonic analysis. Our comments will be brief, and necessarily vague.

The Fourier transform $\hat{f}$ of a function f on G is given formally for all nonzero real numbers λ by the operator

$$\hat{f}(\lambda)=\int_G f(g)T_\lambda(g)dg, \tag{6.10}$$

where $dg=dr_1dr_2dr_3$ is (bi-invariant) Haar measure on G, and f is recaptured by the inversion formula

$$f(g)=(2\pi)^{-2}\int_{\hat{G}} tr\left(\hat{f}(\lambda)T_\lambda(g)^{-1}\right)|\lambda|d\lambda. \tag{6.11}$$

Using classical Fourier analysis, one can show that for sufficiently smooth functions f, the trace in (6.11) exists, and these formulas make rigorous sense. The measure $dm(\lambda)=(2\pi)^{-2}|\lambda|d\lambda$ on $\hat{G}$ is called *Plancherel measure* for G, and the formula (which follows from (6.11))

$$\int_G |f(g)|^2dg=\int_{\hat{G}}\|\hat{f}(\lambda)\|^2dm(\lambda) \tag{6.12}$$

is the *Plancherel formula* for G. Here, $\|\hat{f}(\lambda)\|$ denotes the so-called Hilbert–Schmidt norm of the operator $\hat{f}(\lambda)$. Finally, by means of the Plancherel formula, one obtains the decomposition

$$R \cong \int_{\hat{G}}^{\oplus} \infty T_\lambda \, dm(\lambda) \tag{6.13}$$

of the regular representation R of G as a direct integral in which each element T_λ of the infinite-dimensional unitary dual appears with infinite multiplicity.

7. Concluding remarks. The preceding exposition takes us to the brink of the postwar era in harmonic analysis. Indeed, the past three decades have witnessed a remarkable growth in our understanding of infinite-dimensional group representations and the associated noncompact harmonic analysis. Once again, in the interest of simplicity we shall be brief and mention but a few of the highpoints. Here, of course, my own prejudices will be revealed. I apologize for omissions of perhaps equally important contemporary developments that would be mentioned in a longer article.

The thrust of much of the recent work in harmonic analysis on locally compact topological groups has been directed toward solutions to the following two fundamental problems for a noncompact nonabelian group G.

1. Find all irreducible unitary representations of G, up to equivalence. That is, what is $\hat{G}$? This is the problem of existence of fundamental harmonics.

2. Decompose the regular representation of G on $L^2(G)$. Alternatively, find the Plancherel formula.

In stark deviation from the compact or abelian theory, there is essentially no hope for reasonable solutions in terms of the *generic* noncompact nonabelian group G.

Groups for which there exists a satisfactory decomposition theory are called *Type I*, or less commonly but more fittingly, *tame*. The definition is too technical to give here. The general abstract representation theory of such groups has been extensively studied and is well developed. For Type I unimodular groups, there is a general existence theorem for the Plancherel formula, due to I. E. Segal. However, for a specific group, the abstract theory gives no clue as to the construction of the Plancherel measure, nor to the structure of the dual.

Nonetheless, as one sees for the Heisenberg group, all is not so bleak if one considers more restrictive classes of such groups. Regrettably, the resultant harmonic analysis is most often extremely complicated. At the very least, one is forced to deal with such discomforts as interwoven combinations of discrete and continuous decompositions, infinite-dimensional irreducible representations, and the fact that not all elements of $\hat{G}$ appear in the Plancherel formula. We mention some of the high points.

Mackey theory. In the early 1950's, the algebraic and measure-theoretic foundations for infinite-dimensional representation theory were put in place by George Mackey. The central notion is that of an *induced representation*. Very roughly, Mackey's theory reduces the representation theory of a group to that of a normal subgroup, modulo questions of group extensions.

Harish-Chandra theory. Here one is concerned with harmonic analysis on noncompact real semi-simple Lie groups. The early work, especially for complex groups, was done by Gelfand and various collaborators. However, the subject is now dominated by the work of Harish-Chandra, whose numerous papers over a span of two decades finally culminated in 1968 in the Plancherel formula for real semi-simple groups.

Kirillov theory. In his doctoral thesis in 1962, A. A. Kirillov supplied a powerful geometric idea which completely described the harmonic analysis of nilpotent Lie groups, and which now dominates harmonic analysis on solvable groups. Roughly speaking, Kirillov's theory associates to each element of the dual of a Lie group, a so-called *orbit* which has the structure of a symplectic manifold. Extensions of Kirillov's method, by L. Auslander and B. Kostant, yielded the decomposition of the regular representation for all tame solvable groups, as well as the more general geometric theory, due to Kostant, of "quantization."

Typically, although not exclusively, the important applications of harmonic analysis (e.g., those in physics, number theory, and noncommutative real and complex analysis) arise from the decomposition of $L^2(X)$ under the action of a group of symmetries of the space X. For such applications, the reader is referred to the literature.

8. A Guide to Further Readings. *Section 1.* A more substantial introduction to Fourier series appears in the monograph by Weiss [**1**]. A treatment of the subject which brings out the diverse applications of Fourier series is given by Dym and McKean [**4**]. For a distribution-theoretic presentation of Fourier series, see the text by Beals [**2**]. Our historical citations are drawn from Grattan-Guiness [**3**] and the references therein. See also the exposition by Zygmund in [**38**]. The two volumes [**6**] by Zygmund form the definitive scholarly work in the field of Fourier series.

Section 2. Wussing's book [**8**] presents the historical development of group theory. (For a start, see the review by W. Waterhouse, Bulletin Am. Math. Soc., 78 (1972) 385-391.) Of course there are innumerable texts on the subject of finite groups. The book by Boerner [**7**] contains an elementary treatment of the representation theory.

Section 3. A mathematically substantive account of the development of Lie theory appears in the Historical Notes of [**9**].* Tidbits of philosophy and history can also be found in Weyl's important book [**11**]. See also the Historical Notes in [**10**] for information on the genesis of topological groups and Haar measure. For a short contemporaneous account of the Peter-Weyl Theorem from the pen of the master, see Weyl's lecture [**12**] to the Swiss Mathematical Society in 1927. For a current point of view, see the expository paper by G. Weiss in [**38**]. Weil's book [**15**] contains a more extensive treatment. The Cartan-Weyl theory is developed in Wallach [**31**].

Section 4. Vilenkin's book [**16**] provides extensive coverage of special functions from the harmonic analysis point of view. We refer to Stein and Weiss [**17**] for a classical treatment of spherical harmonics, and to Coifman and Weiss [**32**] for a group theoretic presentation. The elementary article [**18**] gives a somewhat more general point of view. The last chapter of Helgason's book [**29**] contains the general theory.

Section 5. Of course, there are many expositions of the classical theory, e.g., [**1**], [**17**], [**20**]. The encyclopedic treatise is Titchmarsh's book [**19**]. For the abstract theory, see [**15**], [**21**], and Graham's article in [**38**]. A development that centers around the Stone–von Neumann Theorem is given by Segal and Kunze [**22**].

Section 6. See the original article by von Neumann [**33**], the exposition by Pukanszky [**20**], or [**22**] for the Stone–von Neumann Theorem.

Section 7. There is no royal road through contemporary infinite-dimensional representation theory.‡ Most treatments or surveys are highly technical. Mackey's article [**24**] surveys the field through 1963,† and the A.M.S. symposium [**27**] brings matters up to 1972. The textbook [**23**] by Kirillov, which gives a broad introduction to current practice, contains a relatively elementary introduction to his orbit theory. The expository paper [**38**] by P. Sally describes the most easily accessible example in which Harish-Chandra's theory applies. One might also want to take a look at the survey [**30**] by Harish-Chandra, and the memoir [**28**] by Auslander and Moore. [**34**] treats

*After this paper was written, the preprint [**41**] by S. Helgason appeared. This is a written version of an invited address to the A.M.S. The introductory section is a highly attractive summary of the origins of Lie theory.

†It has been brought to my attention that the reissue in 1976 of Mackey's 1955 lectures at the University of Chicago contains a lengthy appendix that brings [**24**] up to date. In this addendum one finds a survey that touches not only upon "Mackey theory" itself, but nilpotent, solvable, and semi-simple representation theory, as well as applications in number theory, ergodic theory, and quantum mechanics.

‡Cf. the review by R. A. Kunze, Bull. Amer. Math. Soc., 84 (1978), 73–75.

number-theoretical applications, [35] concerns those in probability, and [36] is one among many treatments of applications to physics. Dyson's elementary article in [39] is the first place to look for group theory in physics. See also the expositions [25] and [26] by Mackey. Extensive bibliographies appear in [24], [27], [23], and Warner's treatise [37].

References

1. G. Weiss, Harmonic Analysis, MAA Studies in Mathematics, vol. 3, I. I. Hirschman, Jr., ed., Prentice-Hall, Englewood Cliffs, N.J., 1965, 124–178.
2. R. Beals, Advanced Mathematical Analysis, Springer Graduate Texts in Mathematics, 1973.
3. I. Grattan-Guiness, The Development of the Foundations of Mathematical Analysis from Euler to Riemann, MIT Press, Cambridge, Mass., 1970.
4. H. Dym and H. P. McKean, Fourier Series and Integrals, Academic Press, New York, 1972.
5. W. Burnside, Theory of Groups of Finite Order, Cambridge University Press, 1911 (Dover reissue, 1955).
6. A. Zygmund, Trigonometric Series, Cambridge University Press, New York, 1959.
7. H. Boerner, Representations of Groups, North-Holland, Amsterdam, 1963.
8. W. Wussing, Die Genesis des abstrakten Gruppenbegriffes, VED Deutscher Verlag der Wissenschaften, Berlin, 1969.
9. N. Bourbaki, Elements of Mathematics: Lie Groups and Lie Algebras, Pt. 1, Addison-Wesley, Reading, Mass., 1975, chapters 1-3.
10. ———, Éléments de Mathématique: Intégration, vol. 6, Hermann, Paris, 1963, chapters 7, 8.
11. H. Weyl, The Classical Groups, Princeton University Press, Princeton, N.J., 1946.
12. ———, Sur la représentation des groupes continus, L'Enseignement Mathématique, 26 (1927) 226–239.
13. ———, The Theory of Groups and Quantum Mechanics, Dover, New York, 1950.
14. F. Peter and H. Weyl, Die Vollständigkeit der primitiven Darstellungen einer geschlossenen kontinuierlichen Gruppe, Math. Ann., 97 (1927) 737–755.
15. A. Weil, L'Intégration dans les Groupes Topologiques et ses Applications, Hermann, Paris, 1938.
16. N. J. Vilenkin, Special Functions and the Theory of Group Representations, American Mathematical Society, Providence, R.I., 1968.
17. E. Stein and G. Weiss, Introduction to Fourier Analysis on Euclidean Space, Princeton University Press, Princeton, N.J., 1971.
18. K. I. Gross and R. A. Kunze, Fourier decompositions of certain representations, in Symmetric Spaces, W. Boothby and G. Weiss, eds., Marcel Dekker, New York, 1972.
19. E. C. Titchmarsh, Introduction to the Theory of Fourier Integrals, Clarendon Press, Oxford, 1962.
20. L. Pukanszky, Leçons sur les Représentations des Groupes, Dunod, Paris, 1967.
21. W. Rudin, Fourier Analysis on Groups, Interscience, New York, 1962.
22. I. Segal and R. Kunze, Integrals and Operators, McGraw-Hill, New York, 1968.
23. A. A. Kirillov, Elements of the Theory of Representations, Springer, 1976.
24. G. Mackey, Infinite-dimensional group representations, Bull. Amer. Math. Soc., 69 (1963) 628–686.
25. ———, Induced representations of locally compact groups and applications, Functional Analysis and Related Fields, F. Browder, ed., Springer, 1970.
26. ———, Induced Representations and Quantum Mechanics, Benjamin, New York, 1968.
27. Harmonic Analysis on Homogeneous Spaces, Proceedings of Symposia in Pure Mathematics, 26, C. C. Moore, ed., American Mathematical Society, Providence, R.I., 1972.
28. L. Auslander and C. C. Moore, Unitary representations of solvable Lie groups, Mem. Amer. Math. Soc., 62 (1966).
29. S. Helgason, Differential Geometry and Symmetric Spaces, Academic Press, New York, 1962.
30. Harish-Chandra, Harmonic analysis on semi-simple Lie groups, Bull. Amer. Math. Soc., 76 (1970) 529–551.
31. N. Wallach, Harmonic Analysis on Homogeneous Spaces, Marcel Dekker, New York, 1973.
32. R. Coifman and G. Weiss, Analyse Harmonique Non-Commutative sur Certains Espaces Homogènes, Springer Lecture Notes 242, 1972.
33. J. von Neumann, Die Eindeutigkeit de Schrödingerschen Operatoren, Math. Ann., 104 (1931) 570–578.
34. I. M. Gelfand, M. I. Graev, I. I. Pyatetskii-Shapiro, Representation Theory and Automorphic Functions, Saunders, Philadelphia, 1969.
35. E. J. Hannan, Group Representations and Applied Probability, Methuen Monographs, London, 1965.
36. M. Hamermesh, Group Theory and Its Applications to Physical Problems, Addison-Wesley, Reading, Mass., 1962.
37. G. Warner, Harmonic Analysis on Semi-simple Lie Groups, I and II, Springer, 1972.
38. Studies in Harmonic Analysis, MAA Studies in Mathematics 13, J. M. Ash, ed., Mathematical Association of America, 1976.

39. Mathematics in the Modern World, Readings from Scientific American, W. H. Freeman, San Francisco, 1968.

40. G. W. Mackey, The Theory of Unitary Group Representations, University of Chicago Press, 1976.

41. S. Helgason, Invariant differential equations on homogeneous spaces, Bull. Am. Math. Soc. 83 (1977).

Added in page proofs: G. W. Mackey has informed me that he has a lengthy article related to the theme of this paper to appear in 1979 in Rice University Studies, Proceedings of the Conference on History of Analysis.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL, NC 27514.

GEOMETRICAL OPTICS AND THE SINGING OF WHALES

CATHLEEN SYNGE MORAWETZ

I must begin by confessing that the title of this talk is quite a come on. I am just as fascinated as anyone by the huge mammals that cavort in the sea and by their long mysterious and repetitive songs, but I am not at all an authority on how they make these sounds or what they are trying to communicate. What captured my fancy was the enormous distances sound travels in water and that whales appear to take advantage of this.

However, even more than that I felt that here is a simple example from nature which should arouse the applied mathematician in all of us to understanding, the analyst to rigorizing and, I hope, the teacher to explaining. It is, therefore, to the theoretical problems of mathematical physics that I will address myself.

That is the gist of my sermon and I am sorry that I cannot reproduce in print the beautiful "hymns" of the whales which have been so excellently recorded.* Certain features are characteristic. Early recordings sound like a lot of cocktail party noise and are as hard to analyze. Single whales in bays demonstrate that they repeat their sounds over and over. It is not hard to imagine that they are fascinated by the echoes from the walls of the bay. They also may be trying to send messages and we shall see how. The most intriguing sound of all is the low frequency note that they produce which can be transmitted for hundreds of miles, unlike the high frequency tones which are absorbed by the water. This note sounds more like a motor boat than a sound made by anything so human as a whale.

Turning to the mathematics, it is necessary to begin somewhere. We plunge in with the differential equation for the disturbance in the pressure:

$$\phi_{tt} = c^2 \Delta \phi,$$

here ϕ is a function of all the space variables and time, $\Delta = \text{div grad}$ and c is the speed of sound given in terms of pressure p as a function of density ρ by $\sqrt{dp/d\rho}$. Since water is almost incompressible, this speed is much greater in the ocean than in air. In the limit it seems $\Delta\phi = 0$. What in fact does this familiar limit mean? We shall come back to this later, but meanwhile we must justify the quantity c as the speed of propagation of something and also ask how much of the something actually goes with it.

To draw these conclusions from this equation is fairly complicated and the simplest way to go about it is to study high frequency disturbances. It is their behavior which is given by "geometrical optics." The helpful fact is that qualitatively they also describe not so high frequency behavior.

*The best example is "Songs of the Humpback Whale" by Roger S. Payne, Capital Records Stereo ST-620, and early recordings were made by W. E. Chaville and W. A. Watkins at the Woods Hole Oceanographic Institute. See also [7].

The author received her Ph.D. from New York University (Courant Institute) under the direction of K. O. Friedrichs and (except for a year at M.I.T.) has been there ever since. Her interests are in the applications of partial differential equations, especially in fluid dynamics, transsonic flow, diffraction theory, etc. This article was an invited address to the MAA at the 1976 Summer Meeting.—*Editors*