# A TECHNICAL SURVEY OF HARMONIC ANALYSIS

TERENCE TAO

Abstract. A rather technical description of some of the active areas of research in modern real-variable harmonic analysis, and some of the techniques being developed. Note: references are completely missing; I may update this later.

## 1. General overview

Very broadly speaking, harmonic analysis is centered around the analysis (in particular, quantitative estimates) of functions (and transforms of those functions) on such domains as Euclidean spaces $\mathbf{R}^d$, integer lattices $\mathbf{Z}^d$, periodic tori $\mathbf{T}^d = \mathbf{R}^d/\mathbf{Z}^d$, or other groups $G$ (which could be anything from Lie groups, to finite groups, to infinite dimensional groups), or manifolds $M$, or subdomains $\Omega$ of these spaces, or perhaps more exotic sets such as fractals. These functions typically take values in the real numbers $\mathbf{R}$ or complex numbers $\mathbf{C}$, or in other finite-dimensional spaces (e.g. one may consider $k$-forms on a manifold $M$, although strictly speaking these are sections of a bundle rather than functions); they may even take values in infinite-dimensional Hilbert or Banach spaces. The range space is usually a Banach space[1]; this allows one to take finite or infinite linear combinations of these functions, and be able to take limits of sequences of functions. Rather than focusing on very special examples of functions (e.g. explicit algebraic functions), harmonic analysis is often more concerned with *generic* functions in a certain regularity class (e.g. in $C^k(M)$, the space of $k$ times continuously differentiable functions on a manifold $M$) or a certain integrability class (e.g. $L^p(M)$, the class of $p^{th}$-power integrable functions on $M$). As such, it provides a useful set of tools when dealing with functions appearing in some mathematical context in which some regularity or integrability is known on the function, but the function cannot be solved explicitly. As generic functions usually do not satisfy any interesting identities, harmonic analysis is often more concerned instead with obtaining *estimates* on these functions and on various transformations of these functions, for instance obtaining quantitative inequalities connecting various norms of such functions. (More qualitative properties such as continuity, differentiability, convergence, etc. are also of interest, but often these are proven by first establishing some relevant quantitative estimates and then applying some sort of limiting argument; the proof of the estimates often

[1] Functions which take values in other sets, such as manifolds embedded in a Euclidean space, are also of interest, although the analysis here has been more difficult, especially if one demands that the analysis remain completely intrinsic (independent of the choice of embedding).

absorbs most of the effort, with the limiting argument at the end usually being quite routine once the estimates are obtained.)

Harmonic analysis overlaps and interacts (quite fruitfully) with many other fields of mathematics, to the extent that it is sometimes difficult to draw a sharp dividing line between harmonic analysis and neighboring fields. For instance, the study of decomposing a function on a highly symmetric space (e.g. a Lie group) into components or "harmonics" overlaps with the representation theory of groups and algebras such as Lie groups and Lie algebras; the analysis of boundary behavior of a holomorphic function on a disk, half-plane, or other domain overlaps with complex variables and partial differential equations (PDE); estimates on exponential sums overlaps with analytic number theory; analyzing the structure of a collection of geometric objects (balls, disks, tubes, etc.) overlaps with geometric measure theory, combinatorial incidence geometry, and additive combinatorics; analysis of linear operators such as the shift map can lead to spectral and scattering theory on one hand, and ergodic theory on the other; and the study of very rough sets and functions on abstract measure spaces can lead one into real analysis, measure theory, and probability; conversely, probabilistic methods are a powerful tool that can be applied to certain problems in harmonic analysis. If one focuses instead on the function spaces corresponding to various regularities, as well as the spaces of linear transformations between them, harmonic analysis begins to blend with functional analysis, operator algebras, and interpolation theory. Finally, linear, multilinear and even non-linear estimates of various differential or integral operators (possibly with oscillation and singularities) are intimately tied up with the analysis of linear and non-linear partial differential equations. Via its ubiquituous appearance in PDE, harmonic analysis is then indirectly linked to even more fields, notably differential geometry and mathematical physics.

In addition to these theoretical connections, harmonic analysis also has many applications to applied and numerical mathematics. For instance, given the usefulness of the Fourier transform (on various groups) in many applications such as electrical engineering, physics, and information science, it is of interest to use harmonic analysis to discover algorithms which compute the Fourier transform (and associated operators such as Fourier multipliers) efficiently and robustly. More recently, other transforms, notably the class of wavelet transforms and their variants, have been developed for a number of applications, from signal processing to simulation of PDE, and also play an important role in approximation theory, giving rise to a very active applied branch of harmonic analysis.

Given the general scope of the field, and its many interconnections to other fields, it is not surprising that there are many complementary viewpoints to take in this field. For instance, one could approach harmonic analysis algebraically, viewing the Fourier transform, its symmetries, and the action of this transform and related operators on special functions such as harmonics or plane waves; this approach fits well with the representation theory aspect of harmonic analysis, though of course one still has to do a fair bit of analysis to justify convergence on these groups, especially if they are infinite, or perhaps even infinite-dimensional. This part of the field is sometimes referred to as *abstract harmonic analysis*, or *Fourier analysis on* *groups*, but also connects with other fields such as functional analysis and representation theory. Related to this, one could proceed via the spectral resolution of the Laplacian (which is closely related to the Fourier transform), leading to a spectral theory approach to the subject. Or one could rephrase harmonic analysis questions in a complex analysis setting (replacing Fourier series by Taylor or Laurent series, etc.); historically this was the primary way to approach this subject, and is still active today, although much of the theory here has been vastly generalized (e.g. replacing the Cauchy-Riemann equations by more general elliptic PDE in higher dimensions, or working with several complex variables instead of just one). Then there is the *real-variable* approach, which grew out of the complex variable approach but has a much greater emphasis on tools such as smooth cutoff functions (in both space and frequency) to localize the functions of interest into more manageable components, even at the expense of properties such as complex analyticity. This approach was developed by many people including Calderón, Zygmund, and Stein, and is perhaps one of the more prominent perspectives in the field today. Related to the real-variable approach are more geometric approaches, such as variational methods, heat kernel methods, or the viewpoint of *microlocal analysis*. Here the emphasis is on studying the distribution of functions either in physical space, or in frequency space (via the Fourier transform), or even simultaneously in a product space known as *phase space*. Here one runs into the well known obstruction given by the *uncertainty principle*, in that there is a limit as to how much one can localize in both physical and frequency space simultaneously, nevertheless in many applications it is possible to accept this limitation and still perform a viable analysis on phase space. This phase space also has a natural symplectic structure, and there are many connections between the analysis of functions in phase space and of the underlying symplectic geometry, which is closely related to the connections between quantum mechanics and classical (Hamiltonian) mechanics in physics (via *geometric quantization*). This viewpoint seems especially useful for solving PDE, especially time-dependent PDE. Once one transforms the problem to a phase space setting, one is sometimes left with the geometric combinatorics task of controlling the overlap of various objects such as balls, rectangles, tubes, light rays, or disks in physical space, frequency space, or phase space; while such tasks are perhaps strictly speaking not inside the realm of harmonic analysis, they do seem to be increasingly essential to it.

As mentioned earlier, harmonic analysis has been a very fruitful tool in the analysis of PDE (most obviously in the study of *harmonic functions*, which are solutions to the PDE $\Delta u = 0$, but in fact a very wide class of PDE is amenable to study by harmonic analysis tools), and has also found application in analytic number theory, as many functions in analytic number theory (e.g. the Möbius function $\mu(n)$) have such a "random" behavior that they can to some extent be treated as a generic function). In practice, though, when applying harmonic analysis methods to other fields, a purely harmonic analysis approach is not always the optimal one, as the functions that arise in other fields often have additional structure (e.g. they solve a PDE, or obey some geometric constraints, or satisfy some arithmetic relations, or have some positivity or symmetry properties, etc.) which must be exploited in conjunction with the harmonic analysis techniques to obtain sharp results. Thus harmonic analysis can often be viewed not a self-contained theory, but rather as a collection of useful tools and principles for estimating functions and various transforms of those functions, which can then be incorporated (together with arguments from other fields of mathematics) to control various mathematical objects arising in applications.

One characteristic of harmonic analysis is that it tends to be *local* - studying characteristics of functions which depend mostly on nearby values of the function and not so much on distant values. Because of this, the results in harmonic analysis are often rather insensitive to global features such as the topology of the underlying manifold or the degree of the map being studied, although there are a handful of methods (notably heat kernel methods) which are striking counterexamples to this general trait, and there are some recent results which begin to combine these global considerations with the more local ones of traditional harmonic analysis. This emphasis on localization also shows up in the application of harmonic analysis to PDE; typically harmonic analysis is only used to control the PDE *locally*, and other methods (e.g. using conserved integrals of motion) are then used to extend this control globally. Related to this emphasis on local phenomena is the widespread use in harmonic analysis of *cutoff functions* - smooth, compactly supported functions which one uses to localize the functions being studied to specific regions of space (or frequency). Thus the techniques and strategies in this field tend to have a "reductionistic" philosophy; for instance, if one wishes to control an integral operator $Tf(x) := \int K(x,y)f(y)\,dy$ which exhibits both singularity and oscillation in its kernel $K(x,y)$, then one might approach this problem by dividing the kernel $K$ into two (or more pieces), one of which is singular but not particularly oscillatory, while the other is oscillatory but not particularly singular, and estimate the two components separately. Another common method in a similar spirit is that of "dyadic decomposition" - splitting an expression (usually some sort of integral) into a countable number of pieces, depending on which "dyadic shell" a certain parameter (e.g. the frequency magnitude $|\xi|$, or the radial variable $|x|$, or the magnitude of a function $|f(x)|$) falls into, estimating each piece separately, and then somehow recombining the dyadic components efficiently to estimate the whole. While this type of "divide and conquer" strategy can lead to somewhat lengthy, inelegant and computationally intensive arguments, it does seem to be rather effective and can give reasonably good bounds, as long as one is prepared to concede a loss of a constant (or perhaps a logarithmic factor) in the final estimate.

It would be impossible to give the entire field justice in a survey as brief as this, and we shall therefore have to make a number of sacrifices. First of all we shall restrict our attention primarily to the harmonic analysis of a fixed Euclidean space $\mathbf{R}^d$. In doing so we forego any number of important directions of research in the field - for instance, harmonic analysis on discrete, compact, finite, or non-abelian groups, or on manifolds, or in very general measure spaces (e.g. spaces of homogeneous type), or in one or more complex variables, or on very large or infinite-dimensional spaces; however the Euclidean spaces often serve as a simple model or starting point for these more general situations, and so they seem to be the best context to discuss. For a related reason, we shall emphasize the real-variable approach to harmonic analysis, since it is well adapted to dealing with the Euclidean space setting, but as discussed above it is clearly not the only way to approach the subject. Thirdly, we shall deal mainly with scalar (real-valued or complex-valued) functions, and the various interesting transforms that take one or more such functions to others (typically in a linear or multilinear fashion); again, this excludes several important directions of study, notably vector-valued or manifold-valued functions, as well as the study of geometric objects such as fractals (which, like functions, have quantitative notions of regularity, etc. attached to them). Again, we choose this because the scalar functions serve as a starting point for these other, more complicated situations. Finally, we shall be concerned primarily with the task of obtaining *estimates* on these transforms; there are a number of other important questions one can ask concerning these transforms (invertibility, convergence, asymptotic behavior, etc.) but many of these questions require that one first establish estimates, and so we shall focus on these. As we shall see, even with such limitations in scope, this is still a very large subject to cover.

## 2. Operators

There are several main actors in a harmonic analysis problem, including the underlying domain (which may be a measure space, a metric space, a manifold, etc.), some spaces (often Banach spaces) of functions on that domain, and some operators (typically linear, sub-linear, or multi-linear) on those spaces. In this section we focus on the latter type of object, and briefly discuss some typical examples of operators which arise in harmonic analysis problems of interest. It is of course impossible to survey all of the operators that have been considered in the literature, and so we shall only restrict our attention to a few of the most intensively studied ones. For simplicity we shall only work in the model setting of a fixed Euclidean space $\mathbf{R}^d$ (and avoid discussing the interesting issue on the uniformity of estimates in the high-dimensional limit $d \to \infty$), endowed with the usual metric $|x - y|$ and measure $dx$, although many of the operators we discuss here have analogues on many other types of domains, or with less standard metrics or measures. It is impossible to list all the possible generalizations here, though one should remark that the study of Fourier series on the circle $\mathbf{T} = \mathbf{R}/\mathbf{Z}$ is also an important component of the field, especially from a historical perspective and with regard to the connections with complex analysis and with number theory (e.g. via the Hardy-Littlewood circle method). However in many circumstances there are *transference principles* available which allow one to convert a question about Fourier series on $\mathbf{R}/\mathbf{Z}$ to one about Fourier integrals on $\mathbf{R}$, or vice versa. We should also mention that there are very conceptually useful *dyadic models* of Euclidean space, for instance replacing the Fourier transform by the Fourier-Walsh transform, and replacing continuous wavelets by Haar-type wavelets, which while not as directly related to as many applications as the Euclidean problems, serve as much cleaner models in which to test and develop techniques and tools which one can later adapt (or in some cases directly transfer) to the Euclidean setting.

While the qualitative properties of these transforms (convergence, invertibility, regularity, etc.) are of interest, it is the *quantitative* properties (in particular, estimates for the transformed function in terms of the original function, or occasionally vice versa) which are often the primary focus of investigation, with the qualitative properties then being obtained as a by-product of the quantitative estimates ("hard analysis") via some general limiting arguments or functional analysis ("soft analysis"). For instance, if one is interested in the convergence of some operation applied to rather rough functions (e.g. square-integrable functions), one might proceed by first proving an *a priori* estimate applied to *test functions* - smooth functions which are compactly supported, or *Schwartz functions* - smooth functions whose derivatives are all rapidly decreasing[2], and then applying some sort of limiting procedure to obtain the desired convergence for rough functions (the a priori estimate being used to make the passage to the limit rigorous). Note that while the a priori estimate is only proven for smooth functions, it is often important that the actual estimate itself relies only on rougher norms of these functions such as $L^p$ norms, in order to justify the passage to the limit. In the rest of this section, therefore, we shall not concern ourselves with questions of convergence, integrability, etc. of various integrals and transforms, assuming instead that there is always enough smoothness and decay to justify the existence of all expressions described below.

Many questions in harmonic analysis involve first a decomposition of functions into some standard basis, often orthonormal or approximately orthonormal. There are of course many interesting decompositions to study, but we shall focus just on two closely related decompositions, which are arguably the most fundamental to the theory, and the model for countless generalizations: the *Fourier decomposition*, associated to the translation structure of $\mathbf{R}^d$ and the *spectral resolution* of the Laplacian, associated to the operator $\Delta := \sum_{j=1}^n \frac{\partial^2}{\partial x_j^2}$. The former is of course given by the *Fourier transform*, defined by

$$\hat f(\xi) := \int_{\mathbf{R}^d} f(x) e^{-2\pi i x\cdot\xi}\, dx,$$

which can be inverted by the *inverse Fourier transform*

$$g^\vee(x) := \int_{\mathbf{R}^d} g(\xi) e^{2\pi i x\cdot\xi}\, d\xi.$$

Using the Fourier transform, we can define *Fourier multiplier operators* $T_m$ for any function $m(\xi)$ (with some minimal measurability and growth assumptions on $m$) by the formula

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

or equivalently

$$T_m f(x) := \int_{\mathbf{R}^d} \hat f(\xi) m(\xi) e^{2\pi i x\cdot\xi}\, d\xi$$

Examples of Fourier multipliers include differential operators. For instance, if $m(\xi) = \xi_j$ (where $\xi = (\xi_1, \dots, \xi_n)$), then $T_m$ is just the differentiation operator $\frac{1}{2\pi i}\frac{\partial}{\partial x_j}$. More generally, if $m(\xi)$ is a polynomial, $m(\xi) = P(\xi_1, \dots, \xi_n)$, then $T_m$ is

[2]This emphasis on a priori estimates makes harmonic analysis a subtly different discipline from its close cousin, *real analysis*, which often prefers to deal with rough or otherwise pathological functions directly, occasionally to such an extent that tools from logic and model theory are required to resolve the truth, falsity, or undecidability of some of its assertions.

the corresponding polynomial of the commuting operators $\frac{1}{2\pi i}\frac{\partial}{\partial x_j}$, thus

$$T_m = P(\frac{1}{2\pi i}\frac{\partial}{\partial x_1}, \ldots, \frac{1}{2\pi i}\frac{\partial}{\partial x_n});$$

in particular, the Laplacian $\Delta$ mentioned earlier is a Fourier multiplier with symbol $m(\xi) := -4\pi|\xi|^2$. Thus Fourier multipliers are a generalization of constant coefficient differential operators. Another example of a Fourier multiplier are the *spectral multipliers* $F(\Delta)$ of the Laplacian for any function defined on $(-\infty, 0]$, define as the Fourier multipliers corresponding to the symbol $m(\xi) = F(-4\pi|\xi|^2)$; examples of these are the fractional differentiation operators $(-\Delta)^{s/2}$ for $s \in \mathbf{R}$, the *heat operators* $e^{t\Delta}$ for $t > 0$, the *Schrödinger evolution operators* $e^{it\Delta}$ for $t \in \mathbf{R}$, the *wave evolution operators* $\cos(t\sqrt{-\Delta})$ and $\frac{\sin(t\sqrt{-\Delta})}{\sqrt{-\Delta}}$ for $t \in \mathbf{R}$, the *resolvent operators* $(-\Delta - z)^{-1}$ for $z \in \mathbf{C}\backslash[0, \infty)$, and the closely related *Helmholtz operators* $p.v.(-\Delta - \lambda^2)^{-1}$ for $\lambda > 0$. One should also mention the *spectral projections* $1_{[\lambda,\lambda+a]}(\sqrt{-\Delta})$ for $\lambda, a > 0$ in particular the *disk multiplier* $1_{[0,R]}(\sqrt{-\Delta})$, or more generally the *Bochner-Riesz means* $(1 + \Delta/R^2)_+^\delta$, for $R > 0$ and $\delta \geq 0$. Since many important PDE involve the Laplacian, the above operators (and countless variations on these themes) tend to arise quite frequently in PDE. Clearly there are also many inter-relationships between these multipliers, and sometimes one can use properties of one such multiplier to help control another, by means of basic identities such as $T_{m_1 m_2} = T_{m_1} T_{m_2}$. For instance, the finite speed of propagation property for the wave equation leads to compact support properties of the wave operators, which can then be used to localize several of the other operators discussed above; similarly, the smoothing properties of the heat equation lead to corresponding smoothing properties of the heat operators, which can be used to understand the regularity properties of the other operators. These methods often also extend to more general contexts, for instance when the Euclidean Laplacian is replaced by some other differential operator; developing this theory more fully is still an ongoing and active project. For instance, some of the most basic mapping properties of square roots of non-self-adjoint perturbations of the Laplacian (the Kato square root problem) were only answered very recently.

Besides the spectral multipliers of the Laplacian, a number of other Fourier multipliers are also of interest. For instance, the *Riesz transform* $R := d(-\Delta)^{-1/2}$ arise in a number of contexts, notably in the Hodge splitting $1 = -RR^* - R^*R$ of a vector field into divergence-free and curl-free components, or in the theory of elliptic regularity (using control on an elliptic operator such as the Laplacian to deduce control on all other differential operators of the same order). In one dimension, the Riesz transform becomes the *Hilbert transform* $H = \frac{d}{dx}(-\Delta)^{-1/2}$, which is a Fourier multiplier with symbol $m(\xi) := i\mathrm{sgn}(\xi)$, or in physical space $Hf(x) = p.v. \int_R \frac{f(y)}{x-y}\, dy$. This transform plays a basic role in the theory of complex functions on the half-plane (for instance, if $f(x)$ extends analytically to the upper half-plane and has suitable decay at infinity, then the imaginary part of $f$ must be the Hilbert transform of the real part), and serves as a model or building block for other complex analysis operators associated to domains, for instance the *Cauchy transform*

$$Cf(z) := p.v. \frac{1}{2\pi i} \int_\gamma \frac{f(\zeta)}{z - \zeta}\, d\zeta$$

associated to a curve (or other one-dimensional set) $\gamma$. Harmonic analysis techniques (and in particular an advanced form of the Calderón-Zygmund theory mentioned below) have been used successfully to understand the mapping properties of this operator, and hence to shed light on analytic capacity, harmonic measure and other fundamental notions in complex anlaysis.

Fourier multipliers $T_m$ can also be naturally identified with convolution operators by the identity

$$T_m f(x) := \int_{\mathbf{R}^d} K(x-y) f(y)\, dy,$$

where the (possibly distributional) kernel $K$ is simply the inverse Fourier transform $K = m^\vee$ of $m$. These operators can also be thought of as the class of translation-invariant linear operators, acting on (say) test functions in $\mathbf{R}^n$. Such convolution operators fall into many classes, depending on the singular behavior of the kernel $K$, or equivalently on the asymptotic regularity and decay of the symbol $m$. For instance, suppose $m$ is a *homogeneous symbol of order* $k$ for some $k > -d$, in the sense that it obeys the *homogeneous symbol estimates*

$$|\nabla_\xi^\alpha m(\xi)| \leq C_\alpha |\xi|^{k-|\alpha|} \text{ for all } \xi \in \mathbf{R}^d \backslash \{0\}$$

for all multi-indices $\alpha = (\alpha_1, \ldots, \alpha_n)$, where $\nabla_\xi^\alpha := \prod_{j=1}^d \frac{\partial^{\alpha_j}}{\partial x_j^{\alpha_j}}$, $C_\alpha > 0$ is a constant depending on $\alpha$, and $|\alpha| := \alpha_1 + \ldots + \alpha_n$; roughly speaking, these estimates ensure that $m(\xi)$ "behaves like" $|\xi|^k$ with respect to differentiation. Then the corresponding kernel $K$ is a symbol of order $-d-k$ away from the origin (with the behavior near the origin needing to be interpreted in the sense of distributions when $k \geq 0$). For instance, the fractional integration operators $(-\Delta)^{-s/2}$ fall into this category, with $k = -s$. The case $k = 0$ is particularly special, because the kernel $K$ then decays like $|x|^{-d}$, which just barely fails to be integrable both near $x = 0$ and near $x = \infty$. As such, the associated convolution operators $T_m$ are referred to as *singular integral operators* or *Calderón-Zygmund operators*, which must then be interpreted in some suitable principal value sense. The Hilbert and Riesz transforms are prime examples of such integrals; the identity operator (with $m \equiv 1$) is another. An important and significant part of real-variable harmonic analysis is *Calderón-Zygmund theory*, which analyzes more general classes of singular integrals and quantifies the extent to which they behave like the identity operator; this is one of the most well-developed areas of the subject, and serves as a model for other extensions of the theory. As an example of the maturity of the theory, let us briefly mention the $T(1)$ *theorem* (and its generalization, the $T(b)$ *theorem*), which gives necessary and sufficient conditions for a singular integral operator to be bounded on any $L^p$, $1 < p < \infty$; such theorems have proven particularly useful in the study of analytic capacity of sets in the complex plane (since the Cauchy integral can be viewed as a singular integral operator).

The Fourier multipliers of symbol type discussed above were translation invariant; however they are part of a larger class of operators, known as *pseudo-differential operators*, which simultaneously generalize both Fourier multipliers as well as variable coefficient differential operators $\sum_{|\alpha| \leq k} a_\alpha(x) \nabla_x^\alpha$, and as such play an indispensable role in the modern theory of PDE. They can be defined in a number of ways; one such (the *Kohn-Nirenberg functional calculus*) is as follows. The Fourier multipliers defined above can be also written as

$$T_m f(x) = \int_{\mathbf{R}^d} e^{2\pi i x\cdot\xi} m(\xi)\hat{f}(\xi)\ d\xi,$$

whereas from the Fourier inversion formula we see that spatial multipliers $f(x) \mapsto a(x)f(x)$ can be written as

$$a(x)f(x) = \int_{\mathbf{R}^d} e^{2\pi i x\cdot\xi} a(x)\hat{f}(\xi)\ d\xi.$$

We can then generalize these two classes of operators by defining, for each *symbol* $a(x,\xi)$ on the *cotangent bundle* $T^*\mathbf{R}^d := \{(x,\xi) : x,\xi \in \mathbf{R}^d\}$, the *quantization* $Op(a)$ of that symbol, defined by

$$Op(a)f(x) := \int_{\mathbf{R}^n} e^{2\pi i x\cdot\xi} a(x,\xi)\hat{f}(\xi)\ d\xi;$$

thus if $a$ is purely a function of $\xi$ then this is a Fourier multiplier, while if $a$ is purely a function of space then this is a spatial multiplier. In order to perform any meaningful analysis on these operators, some smoothness and decay assumptions must be made on the symbol $a(x,\xi)$. There are many such classes of assumptions, and in practice one may need to tailor the assumptions to the application; but one of the more ubiquituous such assumptions in the literature is to require $a$ to be a *standard symbol of order* $k$, which means that $a$ obeys estimates of the form

$$|\nabla_\xi^\alpha \nabla_x^\beta a(x,\xi)| \leq C_{\alpha,\beta}(1+|\xi|)^{k-|\alpha|};$$

thus for instance a differential operator of order $k$ with uniformly smooth coefficients would also be the quantization of a standard symbol of order $k$, otherwise known as a *pseudodifferential operator of order* $k$. When $k=0$, the operators $Op(a)$ then turn out to be singular integrals of a type which is perfectly suited for the Calderón-Zygmund theory; for instance, this theory can show that such operators preserve the Lebesgue spaces $L^p(\mathbf{R}^n)$ (this is a generalization of the famous *Hörmander-Mikhlin multiplier theorem*), the Sobolev spaces $W^{s,p}(\mathbf{R}^n)$, and a number of other useful spaces, for all $1 < p < \infty$ (there are results of a more technical nature at the endpoints $p=1$ and $p=\infty$). As the term quantization suggests, the study of such operators is motivated in part by quantum mechanics; the operators $Op(a)$ can be viewed as an attempt to make rigorous the concept of an observable $a(X,D)$ of both the position operator $X : f(x) \mapsto xf(x)$ and the momentum operator $D : f(x) \mapsto \frac{1}{2\pi i}\nabla f(x)$ simultaneously, and in particular to find a map $a \mapsto Op(a)$ from symbols to operators which maps $x$ to $X$, $\xi$ to $D$, and is as close to being a homomorphism as possible. Of course, the well known failure of $X$ and $D$ to commute means that this scheme can run into difficulty, especially if one tries to localize both the spatial variable $x$ and the frequency variable $\xi$ to a region of space where $\delta x\cdot\delta\xi \ll 1$ (violating the famous *Heisenberg uncertainty principle*). However, the purpose of the symbol estimates is to ensure that the uncertainty principle is always respected, and in that case one has a good theory (in which the failure of the map $a \mapsto Op(a)$ to be a homomorphism only shows up in "lower order terms" that can be computed quite explicitly, and in fact quite geometrically, being intimately related to the symplectic geometry of classical physics). This leads to the study of *microlocal analysis*, which is an important and fascinating topic in its own right, but perhaps a little outside of the scope of this survey. We do remark however that the intuition of the uncertainty principle, and treating physical space and frequency space as part of the unified *phase space*, is certainly a very helpful viewpoint when approaching harmonic analysis problems.

Singular integral operators $T$ such as the pseudo-differential operators mentioned above have the characteristic of being *pseudolocal* - their kernels are smooth except near the diagonal, which (heuristically, at least) implies that the singularities of $Tf$ will only be located in the same regions of physical space as the original function $f$. (In fact one can phrase a more precise statement concerning the singularity set in phase space, otherwise known as the *wave front set*, but we will not do so here). As such they are intimately tied to the theory of those PDE which also have such pseudolocal characteristics, the primary example of such being elliptic PDE. However, there are other, more non-local, linear operators which are also of interest to harmonic analysis, some in part because of their links to other types of PDE, or to geometry or number theory, and some because they arise from the study of such fundamental questions in harmonic analysis as the convergence of Fourier series or Fourier integrals. One example of such a class of operators are the *oscillatory integral operators*, where in contrast to the singular integral operators, the interesting (and stll not fully understood) behaviour arises from phase oscillations in the kernel rather than singularities. (One can of course combine the two to form *singular oscillatory integral operators*, though in practice one often treats such operators by carefully decomposing them into a primarily singular part, and a primarily oscillatory part). Perhaps the most fundamental such oscillatory integral is the Fourier transform itself:

$$\mathcal{F}f(y) := \int_{\mathbf{R}^d} e^{-2\pi i x\cdot y} f(y)\ dy,$$

but one can consider more general oscillatory integrals, for instance the localized oscillatory integrals

$$T_\lambda f(y) := \int_{\mathbf{R}^d} e^{2\pi i\lambda\phi(x,y)} a(x,y) f(y)\ dy$$

from $\mathbf{R}^d$ to $\mathbf{R}^{d'}$, where $a : \mathbf{R}^d \times \mathbf{R}^{d'} \to \mathbf{C}$ is a suitable cutoff function, $\phi : \mathbf{R}^d \times \mathbf{R}^{d'} \to \mathbf{R}$ is a smooth phase function, and $\lambda$ is a large real parameter. Such oscillatory integrals, up to scaling and partitions of unity, can arise in a number of contexts, for instance in studying the Schrödinger evolution operators $e^{it\Delta}$ (in which $\phi(x,y) = |x-y|^2$) or in studying the *Fourier restriction problem* - to which sets can the Fourier transform of an $L^p$ function be meaningfully restricted (in such a case $\phi(x,y) = x \cdot \Phi(y)$ for some graphing function $y$). These sorts of integral operators arise frequently in *dispersive PDE* - evolution equations which move different frequencies in different directions, of which the Schrödinger equation $iu_t + \Delta u = 0$ is a prime example - and in a variety of questions related to the finer structure of the Fourier transform. They generalize the more classical subject concerning the *principle of stationary phase*, which seeks to control oscillatory integral expressions of the form $\int_{\mathbf{R}^d} e^{2\pi i\lambda\phi(x)} a(x)\ dx$, where $a$ is some explicit cutoff function and $\phi$ is a phase. In $L^2$, such operators can be studied by orthogonality methods (and perhaps a little bit of algebraic geometry to disentangle any degeneracies in the phase function $\phi$); the $L^p$ theory is considerably more difficult, and so far the most progress has arisen from decomposing the functions and operators involved into "wave packets", which carry both an oscillation and some spatial localization, and using oscillatory methods to handle the oscillation, and more geometric methods to control effect of the the spatially localization. Discrete analogues of these expressions, in which integrals are replaced by sums, are also of interest in number theory, although progress here appears to be substantially more difficult.

Another type of operator which arises in practice is a non-local singular integral operator, where the kernel is singular rather than oscillatory, but manages to propagate singularities by a non-zero distance. A typical such operator is a convolution with a measure $Tf := f * d\sigma$, where $d\sigma$ is some measure on some surface in $\mathbf{R}^d$ (e.g. the unit sphere; this operator also arises in the study of the wave equation); another example would be the Hilbert transform along a parabola in $\mathbf{R}^2$, $Tf(x_1, x_2) := p.v. \int_{-\infty}^{\infty} f(x_1 - t, x_2 - t^2) \frac{dt}{t}$. The analysis of such operators tends to rely either on the geometry of the singular set (e.g. if the measure $d\sigma$ is supported on a sphere, the question of how this sphere can intersect various translates of itself becomes relevant), or on the Fourier properties of these operators (e.g. the decay of the Fourier transform of $d\sigma$ may play a key role). Sometimes one applies various localization operators (both in physical and Fourier space) to ameliorate the singular nature of these expressions, or to replace these operators by tamer analogues.

Another class of interesting operators in a similar spirit to the above are the *Radon-like* transforms. One basic example of such are the $k$-plane transforms $X_k$ in $\mathbf{R}^d$ for $1 \leq k \leq d$, which map from (test) functions on a Euclidean space $\mathbf{R}^d$ to functions on the affine Grassmannian $Gr(d, k)$ of $k$-dimensional affine subspaces in $\mathbf{R}^d$ (not necessarily passing through the origin), defined by

$$X_k f(\pi) := \int_\pi f \text{ for all } \pi \in Gr(d, k),$$

where $\int_\pi$ is integration on the $k$-dimensional space $\pi$ with respect to induced Lebesgue measure. The case $k = d-1$ corresponds to the classical Radon transform, and the case $k = 1$ is the *x-ray transform*. These operators arises quite naturally[3] in scattering theory and in the asymptotic behavior of the wave equation, and the question of quantifying its invertibility (which can be resolved in part by harmonic analysis methods) is of importance in applications (e.g. magnetic resonance imaging). More generally, one can consider operators of the form

$$Tf(x) = \int_{R^m} f(\phi(x, z)) K(x, z) \; dz$$

from functions on an $n$-dimensional manifold $M^d$ to functions on an $d'$-dimensional manifold $M^{d'}$, where $\phi : M^{d'} \times \mathbf{R}^m \to M^d$ is some co-ordinate chart of some $m$-dimensional surface in $M^d$, with the family of such surfaces parameterized by $M^{d'}$, and $K$ is some localization function, possibly containing further singularities or oscillations. The treatment of such objects is still fairly difficult, even in model cases; as one might expect, the geometry of the surfaces in question (and in particular the various types of non-degeneracy and curvature conditions they enjoy) plays a major role, but one also needs to understand issues of oscillation (which can arise e.g. by expanding the surface measure on the $m$-dimensional surfaces via Fourier

[3] An interesting variant of the x-ray transform arises in analytic number theory, in which the role of lines is replaced by that of arithmetic progressions; this leads to *discrepancy theory* and to *sieve theory*, both of which are basic tools in analytic number theory.

expansion), for instance by working within the framework of the Fourier integral operators discussed below.

A more general class of operators which encompass Fourier multipliers, pseudo-differential operators, oscillatory integrals, and the non-local singular integral and Radon-like transforms operators discussed above are the *Fourier integral operators* (or FIOs for short), which roughly speaking are the quantum-mechanical analogue of the *canonical transformations* in classical mechanics (just as the pseudo-differential operators correspond to classical observables). There are many equivalent definitions; one such "local" definition is as follows. An operator $T$ is a FIO of order $k$ if it has the representation

$$Tf(x) := \int_{\mathbf{R}^n} e^{2\pi i \Phi(x,\xi)} a(x,\xi) \hat{f}(\xi)\ d\xi$$

where $a$ is a standard symbol of order $k$ as before (which we localze to be compactly supported in $x$), and $\Phi$ is a real-valued phase function which is homogeneous of degree 1 in $\xi$, is smooth on the support of $a$ (except possibly at $\xi = 0$), and is *non-degenerate* in the sense that the mixed Hessian $\nabla_\xi \nabla_x \Phi$ has non-zero determinant on the support of $a$. These operators (which include, for instance, the Schrödinger evolution operators $e^{it\Delta}$ and the wave evolution operators $\cos(t\sqrt{-\Delta})$, $\frac{\sin(t\sqrt{-\Delta})}{\sqrt{-\Delta}}$ as model examples) are fundamental in the theory of variable coefficient linear PDE, as one can rather easily construct FIOs which are *parametrices* (i.e. approximate solutions, modulo lower order or "smooth" errors) for such a PDE by requiring that the phase $\Phi$ and the amplitude $a$ solve certain very natural ODE in phase space (the *eikonal* and *Hamilton-Jacobi* equations respectively) and obey certain initial conditions related to the appropriate boundary conditions for the linear problem. One can also generalize these FIOs to map between Euclidean spaces of different dimension, or even between two manifolds of different dimension; indeed, the class of FIOs is highly robust and can easily cope with such actions as diffeomorphic change of variables (in contrast, for instance, to Fourier multipliers, which must maintain translation invariance throughout). At this level of generality it seems unreasonable at present to hope for a systematic and complete $L^p$ theory, on par with the highly successful Calderón-Zygmund theory, in part because such a theory must necessarily include a number of open conjectures (e.g. the restriction, local smoothing, and Bochner-Riesz conjectures, as well as the more geometric Kakeya conjecture) that remain quite far from resolution. Remarkably, however, there is a fairly satisfactory $L^2$ theory for large classes of the operators discussed above (for instance, a basic result due independently to Hörmander and Eskin asserts that FIOs of order 0 are bounded on $L^2$), and in a number cases (especially ones in which there is not much oscillation present) there is a reasonable $L^p$ theory also.

In many applications (notably in questions relating to the limiting behaviour of sequences of operators) it is not enough to study a single operator $T$, but instead to study a sequence $(T_n)_{n\in\mathbf{Z}}$ of such operators (or possibly a continuous family $(T_t)_{t>0}$. In such cases it is often necessary to understand the associated *maximal function*

$$T_* f(x) := \sup_{n\in\mathbf{Z}} |T_n f(x)|, \text{ or } T_* f(x) := \sup_{t>0} |T_t f(x)|,$$

or the closely related *square function*

$$Sf(x) := (\sum_{n \in \mathbf{Z}} |T_{n+1} f(x) - T_n f(x)|^2)^{1/2} \text{ or } Sf(x) := \int_0^\infty |t \frac{d}{dt} T_t f(x)|^2 \frac{dt}{t};$$

the latter square function can be seen to be comparable in strength to the former by making the change of variables $t = 2^n$ and approximating the resulting integral and derivative by a sum and difference respectively. Note that these operators tend to be sub-linear rather than linear (or alternatively, they can be viewed as linear operators taking values in a Banach space such as $l^\infty$ or $l^2$), but fortunately many of the techniques which are useful in the linear theory (e.g. the real interpolation method) continue to work in the sublinear setting. The relevance of maximal functions in convergence questions arises from the simple fact that in order for a sequence to converge, it must be bounded; and furthermore the *uniform* limit of convergent sequences remains convergent. Both of these situations require that one control expressions such as $T_* f$. The square function arises in a slightly different context, when the operator $T_n$ is transitioning through different scales as $n$ varies; one then expects the operators $T_{n+1} - T_n$ to be somewhat "orthogonal" to each other as $n$ varies, and in such circumstances one expects the summands in such telescoping sums as $\lim_{n \to +\infty} T_n f = T_0 f + \sum_{n=0}^\infty T_{n+1} f - T_n f$ to behave "independently" of each other, in the sense that the magnitude of the sum should be, "on average", comparable to the standard deviation $(\sum_{n=0}^\infty |T_{n+1} f - T_n f|^2)^{1/2}$. This heuristic, arising from fundamental facts of probability theory such as the Khintchine inequality, can be borne out in a number of important cases, notably that of the *Littlewood-Paley square function* which we discuss below.

Perhaps the most fundamental example of a maximal function is the *Hardy-Littlewood maximal function*

$$Mf(x) := \sup_{r>0} \frac{1}{|B(x,r)|} \int_{B(x,r)} |f(y)| \, dy$$

where $B(x,r)$ is the ball of radius $r$ centred at $x$, and $|B(x,r)|$ denotes the volume of $B(x,r)$. This maximal function is clearly related to the question of how a family of balls of varying radii and centres can overlap each other, and the basic result in that subject, the *Vitali covering lemma*, leads directly to the basic estimate for this maximal function, namely the *Hardy-Littlewood maximal inequality*, which asserts among other things that this operator $M$ is bounded on $L^p$ for all $1 < p \leq \infty$. This maximal function in turn plays a central role in Canderón-Zygmund theory, as it is well suited for controlling singular integrals; the point is that while a singular integral of the form

$$Tf(x) = p.v. \int_{\mathbf{R}^n} K(x,y) f(y) \, dy,$$

where $K(x,y)$ obeys singular integral estimates such as $|K(x,y)| \leq C/|x-y|^d$, cannot quite be estimated by the maximal function, each dyadic shell

$$T_n f(x) = \int_{2^n \leq |x| \leq 2^{n+1}} K(x,y) f(y) \, dy$$

of this function can indeed be pointwise dominated by some constant multiple of this maximal function, simply by placing absolute values everywhere and applying the above estimate for $K$. Thus the only issue arises when summing up the various scales $n$. If one is in possession of even a slight amount of cancellation which allows one to improve upon the above crude estimate, then one can hope to control $T$ by $M$ in the sense that one can assert that $T$ is small (at least in some average sense) whenever $M$ is small. This basic idea lies at the heart of Calderón-Zygmund theory and its generalizations.

In many applications, in which the relevant operators are not likely to be localized to small balls, the Hardy-Littlewood maximal operator must be replaced with other maximal operators, for instance ones in which the class of balls is replaced with a class of rectangles, lines, spheres, tubes, or other geometric object. We mention just one such example, the *Kakeya-Nikodym maximal function*

$$f_\delta^{**}(x) := \sup_{T \ni x} \frac{1}{|T|} \int_T |f(y)|\ dy,$$

where $0 < \delta < 1$ is a fixed parameter, and $T$ ranges over all $1 \times \delta$ tubes in $\mathbf{R}^n$ which contain $x$. While apparently quite similar to the Hardy-Littlewood maximal operator, this operator (which is needed to control for instance the operators arising from multi-dimensional Fourier summation) is far less well understood, and is related to a number of questions in geometric measure theory and combinatorics as well as to harmonic analysis.

One can also work with oscillatory versions of maximal functions, though as one might expect this makes such operators much more difficult to handle. A classic example is the *Carleson maximal function*

$$\mathcal{C}f(x) := \sup_{N \in \mathbf{R}} |\int_{-\infty}^N e^{2\pi i x \xi} \hat f(\xi)\ d\xi|$$

in one dimension; this operator is known to be bounded in $L^p$ for all $1 < p < \infty$, which in particular implies (and is almost equivalent to) the celebrated Carleson-Hunt theorem that the one-dimensional Fourier integrals of $L^p$ functions converge pointwise almost everywhere. We remark that the corresponding question for higher dimensions, namely the $L^2$ boundedness of the *maximal disk multiplier*

$$S_*^0 f(x) := \sup_{R>0} \int_{|\xi| \leq R} e^{2\pi i x \cdot \xi} \hat f(\xi)\ d\xi$$

is still open (and considered extremely difficult), even in two dimensions $d = 2$; it would imply the almost everywhere convergence of Fourier integrals of $L^2$ functions in higher dimensions. (To give some idea of the complexity of the problem, the corresponding convergence question for $L^p$ functions, $p \neq 2$, was answered in the negative in two and higher dimensions in a famous result of Fefferman, essentially using some known bad behavior of the Kakeya-Nikodym maximal function $f_\delta^{**}$ mentioned earlier.)

Closely related to the Hardy-Littlewood maximal function is the *Littlewood-Paley square function*. There is some flexibility as to how to define this function; one standard way is as follows. Let $\phi : \mathbf{R} \to \mathbf{R}^+$ be a smooth non-negative bump function which equals 1 on the interval $[-1, 1]$ and vanishes outside of $[-2, 2]$, and then for each integer $n$, let $P_n$ denote the spectral multiplier $P_n := \phi(\sqrt{-\Delta}/2^{n+1}) - \phi(\sqrt{-\Delta}/2^n)$, which localizes the frequency variable smootly to the annulus $2^n \leq$

$2\pi|\xi| \leq 2^{n+2}$. Formally, we have the telescoping series $f = \sum_n P_n f$; since each component $P_n f$ oscillates at a different frequency (comparable to $2^n$), we thus expect by the randomness heuristic alluded to earlier that we expect $f$ to be comparable in magnitude to that of the *Littlewood-Paley square function* $Sf$, defined as

$$Sf(x) := (\sum_n |P_n f(x)|^2)^{1/2}.$$

Indeed, the *Littlewood-Paley inequality* asserts that $f$ and $Sf$ have comparable $L^p$ norms for all $1 < p < \infty$; this basic estimate can in fact be obtained as a routine consequence of the more general machinery of Calderón-Zygmund theory (thinking of $S$ as a vector-valued singular integral). The usefulness of the Littlewood-Paley inequality lies in the fact that while the different Littlewood-Paley components in the original function $f = \sum_n P_n f$ can oscillate, and interfere with each other constructively or destructively, in the square function $Sf$ there is no possibility of cancellation between different frequency components, and thus the square function is often easier to estimate accurately. The Littlewood-Paley decomposition $f = \sum_n P_n f$ is also closely related to the heat kernel representation

$$f(x) = -\int_0^\infty \frac{d}{dt} e^{t\Delta} f(x)\ dt$$

of a function, as well as to similar decompositions such as the wavelet decomposition $f = \sum_{j,k \in \mathbf{Z}} \langle f, \psi_{j,k} \rangle \psi_{j,k}$ for various choices of wavelet $\psi_{j,k}$, which can be viewed as a discrete analogue of the above heat kernel representation. The Littlewood-Paley decomposition is particularly useful for analyzing function spaces, especially if those spaces involve "derivatives" or otherwise measure "regularity" (e.g. Sobolev spaces, Besov spaces, Triebel-Lizorkin spaces); roughly speaking, the reason is that the operators $P_n$ "diagonalize" differentiation operators such as $\nabla$ in the sense that $\nabla P_n$ is roughly "of the same strength" as $2^n P_n$ (since differentiation corresponds to multiplication by $|\xi|$ in frequency space, and $P_n$ localizes $|\xi|$ to be comparable to $2^n$). They also approximately diagonalize pseudo-differential operators $T$ (in the sense that $P_n T P_m$ becomes quite small when $n$ and $m$ are far apart). These basic facts about the Littlewood-Paley decomposition also carry over to the wavelet decomposition, which can be thought of as a variant of the Littlewood-Paley decomposition, followed up by a further decomposition in physical space; thus wavelets are a good tool for studying these classes of function spaces and operators. (For more general FIOs, the wavelet basis turns out to be insufficient; other representations such as wave packet bases or the FBI transform become more useful).

Most of the operators discussed above have been linear, or at least sub-linear. More recently, however, bilinear, multilinear, and even fully non-linear operators have also been studied, motivated by a number of nonlinear problems (typically from PDE) in which the dependence on some given data can be expanded as a multilinear expansion; conversely, problems which seemed inherently linear in nature have been successfully attacked by converting them into a more flexible bilinear or multilinear formulation. One obvious example of a multilinear operator is the pointwise product operator

$$T(f_1, \dots, f_n)(x) = f_1(x) \dots f_n(x)$$

for which the fundamental estimate is of course *Hölder's inequality*

$$\|T(f_1,\dots,f_n)\|_{L^p(\mathbf{R}^d)} \leq \|f_1\|_{L^{p_1}(\mathbf{R}^d)} \dots \|f_n\|_{L^{p_n}(\mathbf{R}^d)} \text{ whenever } 1/p = 1/p_1 + \dots + 1/p_n.$$

Things become more subtle however, if one wishes to understand the behaviour of this operator on other spaces, such as Sobolev spaces $W^{s,p}(\mathbf{R}^d)$; this is of importance for instance in nonlinear PDE. When the number of derivatives $s$ is a positive integer (e.g. $s = 1$), one can proceed by means of the *Leibnitz* rule

$$\nabla(fg) = (\nabla f)g + f(\nabla g)$$

which allows one to distribute derivatives, and then deduce Sobolev multiplication inequalities from their Lebesgue counterparts (and Sobolev embedding). One would also like to obtain similar rules for fractional derivatives, e.g.

$$(\sqrt{-\Delta})^{s/2}(fg) \approx ((\sqrt{-\Delta})^{s/2} f)g + f(\sqrt{-\Delta})^{s/2} g$$

but it is not immediately clear how to make such statements rigourous. One useful way of doing so is by introducing *multilinear Fourier multipliers*, which generalize the linear Fourier multipliers discussed earlier. The starting point for such multipliers is the Fourier inversion formula, which allows one to write the pointwise product operator as a multilinear Fourier integral:

$$f_1(x)\dots f_n(x) = \int_{\mathbf{R}^d} \dots \int_{\mathbf{R}^d} e^{2\pi i x(\xi_1+\dots+\xi_n)} \hat f_1(\xi_1)\dots \hat f_n(\xi_n)\ d\xi_n.$$

We generalize this by introducing a *multilinear symbol* $m(\xi_1,\dots,\xi_n)$, with the corresponding multilinear Fourier multiplier $T_m$ defined by

$$T_m(f_1,\dots f_n) = \int_{\mathbf{R}^d} \dots \int_{\mathbf{R}^d} e^{2\pi i x(\xi_1+\dots+\xi_n)} m(\xi_1,\dots,\xi_n) \hat f_1(\xi_1)\dots \hat f_n(\xi_n)\ d\xi_n;$$

note that the $n = 1$ case recovers the previous definition of a Fourier multiplier. For instance, the pointwise product operator corresponds to the case $m \equiv 1$, while the bilinear operator

$$T(f,g) := (\sqrt{-\Delta})^{s/2}(fg) - ((\sqrt{-\Delta})^{s/2} f)g - f(\sqrt{-\Delta})^{s/2} g$$

corresponds to the symbol

$$m(\xi_1,\xi_2) := (2\pi|\xi_1+\xi_2|)^{s/2} - (2\pi|\xi_1|)^{s/2} - (2\pi|\xi_2|)^{s/2}.$$

One can then try to justify the above "fractional Leibnitz rule" by observing some cancellation in this symbol; for instance when $|\xi_1|$ is substantially larger than $|\xi_2|$, then the two largest terms in this symbol nearly cancel each other out. (Thus the fractional Leibnitz rule is particularly accurate when one of the functions has a far higher frequency than the other). To carry out this program rigourously, one needs estimates that convert bounds on the symbol $m$ to bounds on the operator $T_m$. One such estimate is the *Coifman-Meyer multiplier theorem*, which asserts that if $m$ obeys symbol-type estimates such as

$$|\nabla_{\xi_1}^{\alpha_1} \dots \nabla_{\xi_n}^{\alpha_n} m(\xi_1,\dots,\xi_n)| \leq C_{\alpha_1,\dots,\alpha_n}(|\xi_1| + \dots + |\xi_n|)^{-|\alpha_1|-\dots-|\alpha_n|}$$

for all multi-indices $\alpha_1,\dots,\alpha_n$, then $T_m$ obeys the same estimates as the pointwise product, or more precisely

$$\|T(f_1,\dots,f_n)\|_{L^p(\mathbf{R}^d)} \leq \|f_1\|_{L^{p_1}(\mathbf{R}^d)} \dots \|f_n\|_{L^{p_n}(\mathbf{R}^d)} \text{ whenever } 1/p = 1/p_1 + \dots + 1/p_n, 1 < p_1,\dots,p_n < \infty.$$

We note that the Littlewood-Paley inequality mentioned earlier plays a pivotal role in proving this estimate, which has applications to controlling the non-linear terms of a PDE in various $L^p$ and Sobolev spaces.

Another important class of examples of multipliers covered by the Coifman-Meyer theory include the *paraproducts*, which roughly speaking are bilinear operators which contain only one frequency interaction component of the product operator $(f, g) \mapsto fg$; for instance, one might consider a "high-low" paraproduct which only interacts the high frequencies of $f$ with the low frequencies of $g$. Such an object can be defined more precisely by specifying a symbol $m(\xi_1, \xi_2)$ which is only non-zero in the region where $|\xi_1| \gg |\xi_2|$. These types of paraproducts have a number of uses, not only for multilinear analysis, but also in analyzing *nonlinear* operators such as $u \mapsto |u|^{p-1}u$ for some fixed $p > 1$, by means of the *paradifferential calculus* (and in particular *Bony's linearization formula*, which can be viewed as a variant of the chain rule which allows one to approximate frequency components of $F(u)$ by paraproducts of $u$ and $F'(u)$). In many applications, however, one needs to go beyond the Coifman-Meyer estimate, and deal with multipliers whose symbols obey more exotic estimates. For instance, in the context of nonlinear wave equations, one may wish to study bilinear expressions $Q(\phi, \psi)$ where $\phi(t, x)$ and $\psi(t, x)$ are approximate solutions to the wave equation $(\partial_t^2 - \Delta_x)\phi = (\partial_t^2 - \Delta_x)^2\psi = 0$; taking the spacetime Fourier transform $\tilde{\phi}(\tau, \xi) := \int_{\mathbf{R}} \int_{\mathbf{R}^d} e^{-2\pi i(t\tau + x\cdot\xi)}\phi(t, x)\ dxdt$ of $\phi$, we would then expect $\tilde{\phi}$ (and similarly $\tilde{\psi}$) to be supported near the light cone $|\tau|^2 - |\xi|^2 = 0$, which is the characteristic surface for the wave equation. This suggests the use of multilinear estimates whose symbol contains weights such as $(1 + ||\tau|^2 - |\xi|^2|)^\theta$ for some $\theta \in \mathbf{R}$. Such estimates were pioneered by Klainerman and Machedon (for nonlinear wave equations) and by Bourgain, Kenig, Ponce and Vega (for nonlinear dispersive equations) and now play a major role in the theory, especially for low regularity solutions.

A somewhat different type of bilinear operator is given by the *bilinear Hilbert transform*

$$H(f_1, f_2) := p.v. \int_{\mathbf{R}} f_1(x + t)f_2(x - t)\frac{dt}{t},$$

which is a bilinear multiplier with symbol $m(\xi_1, \xi_2) = \pi i \mathrm{sgn}(\xi_2 - \xi_1)$. This operator first appeared in the study of the Cauchy integral operator (it arises naturally if one Taylor expands that integral in terms of $\gamma$), and since has turned out to be related to a number of other objects, such as the Carleson maximal function and the nonlinear Fourier transform (also known as the scattering transform). It also has some intriguing similarities with the recurrence expressions which arise in ergodic theory and in the study of arithmetic progressions. The study of this operator, like that of the Carleson maximal operator, is rather difficult because the singularities of the operator are spread uniformly throughout frequency space, which makes standard techniques such as Calderón-Zygmund and Littlewood-Paley theory ineffective. However, it turns out that a phase-modulated version of these theories, combined with some geometrical combinatorics in phase space, can handle these operators; this was done for the Carleson operator by Carleson and Fefferman, and for the bilinear Hilbert transform by Lacey and Thiele. Some further progress has since been made in understanding these multilinear singular operators, but there is still much work to be done; for instance there are still no known $L^p$ estimates for the *trilinear Hilbert transform*

$$H(f_1, f_2, f_3) := \int_{\mathbf{R}} f_1(x + \alpha_1 t) f_2(x + \alpha_2 t) f_3(x + \alpha_3 t) \frac{dt}{t},$$

where $\alpha_1, \alpha_2, \alpha_3$ are distinct non-zero real numbers. Somewhat surprisingly, there are hints that this question may hinge on arithmetic and number-theoretic properties of these parameters, and in particular whether the $\alpha_j$ are rationally commensurate. This phenomenon of the arithmetic, combinatorial, and number-theoretic structure of the real numbers playing a decisive role in the deeper questions of this field has also emerged in a number of other contexts, notably in dealing with the Kakeya-Nikodym operator mentioned earlier; again, this phenomenon is very far from being fully understood.

Finally, let us briefly mention that there has been some work in understanding more nonlinear operators. A simple example of a non-linear operator (but one of importance to semilinear PDE) are power maps such as $u \mapsto |u|^{p-1}u$, where $p > 1$ is some fixed exponent (not necessarily integer). These nonlinear maps have been understood to some extent by the *paradifferential calculus*, which seeks to approximate nonlinear expressions by paraproduct expressions; a typical formula in this regard is *Bony's linearization formula* $F(u) \approx T_m(u, F'(u))$, where $F$ is a fairly smooth function of one variable (such as $F(x) = |x|^{p-1}x$) and $T_m$ is a paraproduct which interacts the high frequencies of $u$ against the low frequencies of $F'(u)$; this formula can be viewed as a sophisticated form of the chain rule. There are however even more nonlinear maps which are of interest but only partially understood at present, such as the map from the boundary of a domain to some quantity related to a PDE on that domain (e.g. harmonic measure, fundamental solutions, Cauchy integral, analytic capacity), or similarly the map from a potential $V$ to another quantity related to a PDE with that potential $V$ (notably the *scattering transform* or *nonlinear Fourier transform*, which maps $V$ to the scattering data of an operator such as $-\Delta + V$). Other sources of nonlinear behavior can arise when studying not scalar functions, but functions taking values in non-commutative groups (e.g. matrix-valued functions), or sections of non-trivial bundles, or various tensor fields; another important source comes from studying various inverse problems (e.g. relating a matrix to its resolvents, or recovering a potential from its scattering data). In many cases these non-linear operations have been studied by approximating them by better understood linear or multilinear expressions, such as via the paradifferential calculus; it may well be however that in the future we may see more fully nonlinear techniques in harmonic analysis emerging to handle these types of operators.

## 3. Techniques

Having briefly surveyed some of the operators of interest in harmonic analysis, we now discuss in general terms the types of questions we ask of these operators, and then describe some of the techniques one uses to answer these questions, although we emphasize that our discussion here is only a brief tour of the theory, and does not claim to be complete or exhaustive in any sense. In some circumstances there are some general theorems (e.g. the Hörmander-Mikhlin multiplier theorem, the Coifman-Meyer multiplier theorem, the $T(1)$ and $T(b)$ theorems, the Cotlar-Knapp-Stein lemma, the Riesz-Thorin-Stein and Marcinkeiwicz interpolation theorems, Schur's test, the Christ-Kiselev lemma) which are applicable, and there are a number of basic inequalities and estimates (e.g. the Hölder, Young, Chebyshev, and Cauchy-Schwarz inequalities, Plancherel's identity, the John-Nirenberg, Sobolev, Hardy-Littlewood and Littlewood-Paley inequalities, the Carleson embedding theorem, even the humble triangle inequality and Fubini's theorem) that are ubiquitous throughout the theory. However, it is fair to say that the larger portion of harmonic analysis technique does not fit neatly into formal theorems and universal inequalities; instead, they are organized around a number of simple but powerful heuristics and principles, backed up by some well-understood model examples of these principles in action. In practice one usually must modify the application of these principles in a number of technical ways from the model examples in order to adapt it to the problem at hand; this seems a feature of the field (or of the wider discipline of analysis in general), that there are countless permutations and variations on the objects under consideration and so there is often no hope of (or desire for) a definitive theorem that encompasses all of them (although it does seem possible to discover robust *principles* which have very wide applicability). Even the general theorems mentioned above sometimes have to be tinkered with in a specific application in which one of the required hypotheses doesn't quite apply, or the conclusion is not quite the desired one. Also, as in many fields in mathematics, it is not always so obvious *a priori* which questions will have interesting (or satisfying) answers. For instance, the question of completely classifying all the Fourier multipliers bounded on a fixed $L^p$ space seems hopeless by current technology (even in one dimension), although if one restricts these multipliers to a special class (particularly one motivated by some external application) then interesting and deep progress can be made. Indeed, it seems that the most effective way to make progress in this field is to focus on specific model problems first, and to slowly extend back to increasingly general scenarios once enough insight has been gained on the model problems, rather than try to handle a very general class of problems directly; one reason for this is that any conjecture which was too sweeping has typically proven to have some interesting pathological counterexample, which then illustrates the need for additional structural assumptions on the problem which were not previously apparent. Indeed, it might be argued that such ambituous conjectures, and the counterexamples which dashed them, have been at least as important to the advancement of harmonic analysis as the more well-known positive results, for instance by eliminating unfruitful avenues of research and thus concentrating resources on the more promising ones. Nevertheless we will continue to focus on the techniques used to solve problems affirmatively, given that it is much more difficult to describe in general terms one would seek to formulate conjectures, or discover counterexamples to conjectures.

A basic question concerning a linear or sublinear operator $T$ is whether it is bounded from some Banach space $X$ to another $Y$, or more precisely whether there exists a constant $C > 0$ such that

$$\|Tf\|_Y \leq C\|f\|_X$$

for all $f \in X$; once one has such an estimate in hand, other properties of $T$ (or in some cases, of $X$ and $Y$) can often be deduced relatively easily. Actually, in practice it suffices to prove this for a dense subset of $X$ (typically one can use test functions or Schwartz functions for this role), since one can then often extend the estimate to all of $X$ by some limiting argument (or in some cases simply by fiat, extending $T$ from the densely defined set to the whole space in the unique continuous manner). The most common examples of Banach spaces used here are the $L^p$ spaces (particularly for $p = 1, 2, \infty$), but there are a large variety of other spaces used in practice. In PDE applications the Sobolev spaces $W^{s,p}$ (particularly the energy space $W^{1,2}$) are used frequently; other spaces include Besov spaces $B^{s,p}_q$ (or the slight variant $\Lambda^{s,p}_q$), Triebel-Lizorkin spaces $F^{s,p}_q$, Morrey spaces $M^p_q$, Lorentz spaces $L^{p,q}$ (particularly $L^{1,\infty}$), Hardy spaces $\mathcal{H}^p$ (especially $\mathcal{H}^1$), Orlicz spaces $\Phi(L)$ (especially $L \log L$), Hölder spaces $C^{k,\alpha}$, and the space BMO of bounded mean oscillation; in many applications one also needs to consider weighted versions of these spaces, either with a simple power weight such as $(1 + |x|)^s$ or with more general weights (notably the *Muckenhaupt* classes $A_p$ of weights). Many other refinements and variations on these spaces are also considered in the literature; as an example, one can relax the requirement that one work with Banach spaces, instead working for instance with quasi-normed spaces such as $L^p$ for $0 < p < 1$. The reason for such a diversity of spaces is that each space quantifies the various features of a function (regularity, decay, boundedness, oscillation, distribution) to a different extent; the addition of the weights also gives more precise control on the localization of the operator $T$, quantifying the extent to which the size of $Tf$ locally is influenced by $f$. In some cases the more precise estimates available in a more sophisticated space are needed in order to conclude the desired estimate in a simpler space. Fortunately, this zoo of function spaces can be organized and treated in a reasonably unified manner, thanks to a number of basic tools from Calderón-Zygmund theory, notably the Littlewood-Paley inequality; many of these spaces can also be characterized rather efficiently by wavelet bases. Also, one can often deduce estimates for one set of spaces directly from one or more corresponding estimates for another set of spaces; sometimes this occurs because one of the norms is simply dominated by another, but more commonly one has to *interpolate* between two estimates to produce a third. A very typical example of such an interpolation technique is given by the *Riesz-Thorin interpolation theorem*, which asserts that if a linear operator $T$ maps $L^{p_0}$ to $L^{q_0}$ and $L^{p_1}$ to $L^{q_1}$ for some $1 \leq p_0, p_1, q_0, q_1 \leq \infty$, then it will also map $L^{p_\theta}$ to $L^{q_\theta}$ for all $0 \leq \theta \leq 1$, where $1/p_\theta = (1-\theta)/p_0 + \theta/p_1$ and $1/q_\theta = (1-\theta)/q_0 + \theta/q_1$. An important and powerful generalization of this theorem is the *Stein interpolation theorem*, which allows $T$ to depend analytically on a complex interpolation parameter. This can allow one to deduce estimates for an operator $T$ in some intractable-looking spaces $X, Y$ from estimates from modifications of that operator in more friendly spaces, with considerable flexibility in choosing how to modify the operator and which spaces to work in. There is a much more general theory of interpolation available (as well as similar function space techniques such as *factorization*, which take advantage of various symmetries of $T$, $X$, and $Y$) which we do not have the space to describe more fully here. Another simple but important trick is that of *duality*; in order to show that $T$ maps $X$ to $Y$, it usually suffices to show that the adjoint $T^*$ maps $Y^*$ to $X^*$, or that the bilinear form $\langle Tf, g\rangle$ is bounded on $X \otimes Y^*$. (There can be some minor technicalities when

$T$ or $T^*$ is only densely defined, or if the spaces are not reflexive, but these can usually be overcome, for instance by first working with truncated versions of $T$, $X$, and $Y$ and then taking limits at the end).

Note that we do not, in general, specify exactly what the constant $C$ is; merely knowing that such a constant exists is usually enough for many applications, although there are some notable exceptions (for instance, if one wants to invert $1-T$ by Neumann series, one may need the operator norm of $T$ from $X$ to $X$ to be strictly less than 1). However if $T$ (or $X$ or $Y$) depends on some parameters, then the growth rate of $C$ with respect to these parameters is sometimes important. Indeed, one common trick in this area is to apply enough truncations and smoothing operators that it becomes obvious that $T$ is bounded with *some* finite constant $C$ which *a priori* could depend on the truncation and smoothing parameters; one then manipulates this constant with the aim of showing that it ultimately does not depend on these parameters, at which point one can usually take limits to recover bounds for the original operator. Because we do not eventually specify the constant $C$, it is often quite acceptable to use arguments which lose factors of $O(1)$ or so in the estimates. For instance, one might split $T$ (or $f$, or $Tf$ jointly) into a finite number of components, estimate each component separately (often using a completely different argument for each component), and then use a crude estimate such as the triangle inequality to piece these multiple terms together; the loss from the triangle inequality is typically only at most $O(1)$ so this type of decomposition can usually be done "for free" (unless there is some extremely strong cancellation between the components that one needs to exploit). In practice, this often means that if the operator has a number of independent features (e.g. if its local and global behaviour are quite distinct, or it treats low-frequency functions differently from high-frequency functions) then one can often localize attention to a single such feature by means of a suitable decomposition (e.g. into local and global pieces, or into low and high frequency pieces). Even if this multiplies the number of expressions one needs to consider, this is often a simplification of the problem. One can view this decomposition strategy as a more general and flexible version of the *real interpolation method*, which is an example of an interpolation technique as discussed above.

More advanced decompositions are of course available. One common decomposition is the *Littlewood-Paley decomposition* $f = \sum_n P_n f$ mentioned earlier, which localizes the function $f$ to components of a single dyadic frequency range; one can similarly decompose the operator $T$ as $T = \sum_{n,m} P_m T P_n$. Such decompositions are especially useful for expressions involving differential or pseudo-differential operators, which of course arise quite commonly in PDE applications. The Fourier inversion formula can also be viewed as a decomposition, as can decompositions into other bases such as local Fourier bases, wavelets, and wave packet bases, or use of continuous decompositions such as heat kernel representations or the FBI transform; these bases become especially useful if they give an efficient representation of the operator $T$ (e.g. as a diagonal matrix, a nearly-diagonal matrix, or otherwise sparse or factorizable matrix). For instance, Calderón-Zygmund operators are nearly diagonalized by wavelet bases, and Fourier multipliers are exactly diagonalized by the Fourier basis. In Calderón-Zygmund theory there are two primary decompositions. The first is the *Calderón-Zygmund decomposition*, which basically decomposes an $L^p$ function $f$ into a "good" part $g$ which remains bounded, and a "bad" part $\sum_I b_I$ which can be large but is localized to a controlled number of balls or cubes $I$, and which can also be assumed to be "high frequency" in the sense that it obeys certain moment conditions. Related to this is the *atomic decomposition*, notably of Hardy spaces such as $\mathcal{H}^1$, which decompose a function as a linear combination of simpler objects, called *atoms*, with controlled support, size, and moment conditions; they can be viewed as somewhat rougher versions of wavelets. These very useful decompositions have the feature of being *adaptive*, and hence *nonlinear*; they typically proceed by executing some sort of stopping time algorithm or greedy algorithm on the function in question. One feature of this theory is that one then needs to understand the geometric combinatorics of the balls or cubes $I$ which one has selected by these algorithms (which is where the Hardy-Littlewood maximal function plays a fundamental role). More recently, analogues of these decompositions have begun emerging in phase space rather than physical space, in particular the technique of *tree decomposition* in phase space has been a crucial factor in obtaining results for such difficult operators as the Carleson maximal function and the bilinear Hilbert transform; in this case, the geometric combinatorics one needs to control is that of *tiles* in phase space - rectangles in both space and frequency which respect the Heisenberg uncertainty principle $\delta x \cdot \delta \xi \gtrsim 1$. In the study of oscillatory integrals in higher dimensions, the use of *wave packet decompositions* - decomposing both $f$ and $Tf$ into *wave packets*, i.e. bump functions with a certain explicit oscillation localized to a simple region of physical space such as a rectangle or tube - has played a decisive role, and in these cases the underlying geometry of physical space becomes that of the overlap of these rectangles and tubes (at which point maximal functions such as the Kakeya-Nikodym maximal function play a major role). Part of the "art" in harmonic analysis is recognizing which decompositions are suited for a problem, and applying them appropriately; an inappropriate decomposition may end up giving far worse estimates than if one had not decomposed at all. Fortunately, a number of general heuristics (in particular the uncertainty principle, which roughly speaking asserts that it is safe to partition phase space as long as the Heisenberg relation $\delta x \cdot \delta \xi \gtrsim 1$ is respected) and examples can often be relied upon to suggest what decompositions are suitable for the problem at hand. Often after applying the right decomposition, the task becomes much momre geometric and combinatorial in nature, the focus now being on controlling the overlap or interaction between the various geometric objects (balls, tubes, tiles, rectangles, curves, etc.) associated with the decompositions used; often geometric concepts such as curvature and transversality then play a key role in achieving this control. One feature of these sorts of geometric combinatorics is that they tend to be "fuzzy" in nature; since we ultimately do not usually care about the final constant $C$ in our estimates, we are willing to replace these geometric objects with other objects of comparable size, thus for instance one often does not care to distinguish between balls and cubes (although a very eccentric ellipsoid or tube would still be considered different from a ball or cube if the eccentricity could be unbounded). Because of this fuzziness, sometimes the geometric objects can be modeled by a more discrete set of objects (e.g. dyadic cubes), at which point the problem becomes almost completely combinatorial in nature.

Another useful type of decomposition are averaging type representations, in which one considers the original expression $Tf$ as an average of simpler operators dependent on one or more parameters, as those parameters vary over some fixed set. One early example of this idea is the *method of rotations*, in which a multidimensional singular integral was expressed as an average of rotations of one-dimensional operators. Another useful class of decompositions in this spirit is to express a continuous operator as an average of translated or dilated *dyadic model operators*. This can be viewed as placing a randomly chosen dyadic grid on the ambient Euclidean space $\mathbf{R}^d$ and the decomposing expressions such as $Tf$ relative to that grid. The use of randomly chosen grids or similar objects to aid decomposition has a long history in combinatorics, and has also found some important applications to harmonic analysis; for instance, these random grid methods can allow one to perform certain aspects of Calderón-Zygmund theory even when the underlying measure is so irregular as not to obey any sort of doubling condition.

Some decompositions are more in the nature of *approximations*; for instance, one takes an expression such as $Tf$, and estimates it by some main term $M$, plus some error $Tf - M$ which is presumably "lower order" and thus controllable by more crude estimates. For instance, if $f$ is oscillating on some ball or cube $I$, one might approximate $f$ on this ball by its average $\frac{1}{|I|}\int_I f$, which does not oscillate, plus an error which has average zero (and is thus likely to enjoy cancellation, especially when an operator $T$ with some regularity in its kernel is applied); this simple trick underpins a large portion of Calderón-Zygmund theory, for instance. Other instances of approximation techniques are based around some sort of expansion (Taylor expansion, Fourier expansion, Neumann series, stationary phase expansions, etc.), designating the dominant terms in this expansion as the "main terms" and estimating all the remaining terms as errors; in some cases one can estimate these errors by a small multiple of the final constant $C$ in the desired estimate, leading to a bound such as $C \leq O(1) + \frac{1}{2}C$ which (as long as $C$ is *a priori* known to be finite) gives the boundedness of $C$. This type of *bootstrap technique* - starting with some very weak *a priori* control on a constant $C$ and "bootstrapping" it to a much better bound - is particularly useful in nonlinear or perturbative problems. Another example of such an approximation arises in manipulating operators which obey some sort of approximate functional calculus; for instance one might approximate the composition $ST$ of two operators by the reverse composition $TS$, plus the commutator $[S, T]$ which is often lower order and thus easier to estimate. As with the other types of decompositions, knowing which approximations are accurate and which ones are unreliable for a given situation is part of the art of the field, and is often justified by rather fuzzy (but remarkably useful!) heuristics, such as the fractional Leibnitz rule mentioned earlier.

An alternate form of approximation is to *dominate* the norm of $Tf$ by (some constant multiple of) the norm of some simpler expression. For instance, when estimating a sum such as $\sum_n f_n$, where the $f_n$ are all oscillating functions, one may hope to exploit some "randomness" or "independence" in the oscillation to dominate this expression, at least in norm, by the standard deviation or *square function* $(\sum_n |f_n|^2)^{1/2}$; this heuristic can be made rigourous for instance in the context of Littlewood-Paley theory. More generally, one often seeks to control an oscillating function by some non-oscillatory (or at least less oscillatory) proxy for that function, such as a maximal function or square function. Another useful tool in the context of dominating one expression by another is the *Christ-Kiselev lemma*, which roughly speaking asserts that if one operator $T_1$ is formed from another $T_2$ by restricting the support of the kernel, then in certain circumstances one can dominate the operator norm of $T_1$ by that of $T_2$.

When one or more of the spaces involved is a Hilbert space such as $L^2$, then *orthogonality* or *almost orthogonality* techniques can be used. A very useful example of such a technique is the $TT^*$ *method*: if $T : H \to X$ is an operator from a Hilbert space $H$ to a Banach space $X$, with an adjoint $T^* : X^* \to H$, then the composition $TT^* : X^* \to X$ has (formally, at least) operator norm equal to the square of that of $T$ or $T^*$, and thus to bound $T$ it suffices to bound $TT^*$; a similar statement also applies for $T^*T$. The point is that $TT^*$ can be much better behaved than $T$ or $T^*$ individually, especially if there is some orthogonality in the kernel of $T$ (since the kernel of $TT^*$ is essentially formed by taking inner products of rows of the kernel of $T$). For instance, if $T$ is the Fourier transform or Hilbert transform (which contain oscillation or singularities respectively in their kernel), then $TT^*$ is the identity operator (which has no oscillation and an extremely simple singularity). The same method also allows one to control sums such as $\| \sum_n c_n \phi_n \|_H$ in a Hilbert space, where the $c_n$ are constants lying in some sequence space (e.g. $l^2$) and $\phi_n$ are "almost orthogonal", in the sense that the matrix of inner products $\langle \phi_n, \phi_m \rangle$ is very close to being diagonal; for instance, if we have the bound

$$\sup_n \sum_m |\langle \phi_n, \phi_m \rangle| \leq A$$

then one can easily show the approximate Bessel inequality

$$\| \sum_n c_n \phi_n \|_H \leq A^{1/2} (\sum_n |c_n|^2)^{1/2}.$$

Remarkably, a similar orthogonality principle holds when the vectors $\phi_n$ are replaced by *operators*; more precisely, the *Cotlar-Knapp-Stein lemma* asserts that if $T_n : H \to H$ are a (finite) sequence of (bounded) operators obeying the operator norm bounds

$$\sup_n \sum_m \|T_n^* T_m\|_{H \to H}^{1/2} \leq A; \quad \sup_n \sum_m \|T_n T_m^*\|_{H \to H}^{1/2} \leq A$$

then

$$\| \sum_n T_n \|_{H \to H} \leq A.$$

These types of orthogonality methods can be heuristically summed up as follows: if an expression such as $T$ or $Tf$ can be split into "almost orthogonal" components whose interaction is fairly weak, then the task of estimating the whole expression can often be reduced to estimating individual components, with relatively little loss of efficiency in the constants (in contrast to if one used cruder tools such as the triangle inequality, which typically cause a loss proportional to the number of terms in the decomposition).

Orthogonality principles can be very powerful, but unfortunately they are largely restricted to Hilbert spaces such as $L^2$, and largely explains why the understanding of oscillatory expressions in such spaces is significantly more advanced than that in other spaces[4] such as $L^p$, although multilinear analogues of orthogonality techniques can be pursued in $L^4, L^6, \ldots$. For instance, one can convert a linear problem, e.g. that of estimating the $L^p$ norm of $Tf$, into a bilinear problem, such as estimating the $L^{p/2}$ norm of $TfTf$. This trick may seem trivial, but in many cases the bilinear problem is more tractable than the linear one, either because the new space $L^{p/2}$ is easier to work with than the old one $L^p$ (this is especially true when $p = 4$), or because there is more freedom available in the bilinear setting, for instance the two copies of $f$ and the two copies of $T$ can be decomposed separately; this has turned out to be particularly fruitful in oscillatory integral problems such as the Bochner-Riesz summation problem or the Fourier restriction problem. This is an example of a *lifting method* - passing from a lower-dimensional problem to a higher-dimensional one which is more "free" or "decoupled", performing some manipulations which are only possible in this higher dimensional setting, and then eventually *descending* back to the original problem. While somewhat unintuitive, these techniques can be surprisingly powerful, for instance they play a major role in the analysis of singular Radon-like operators propagated along vector fields. The lifting of a problem in physical space to one in phase space can also be viewed as a form of lifting technique (although here the lifting is usually in a heuristic sense only, rather than a rigorous one).

Last, but not least, we should mention a very different type of argument which has been developed for proving these types of estimates, going by such names as *induction on scales* or the *Bellman function method*. Roughly speaking, the idea is to truncate the problem so that there are only a finite number of scales present, and then induct on the number of such scales, keeping care to hold the constants $C$ under control and ultimately be bounded independently of how many scales are "in play". Now the estimate $\|Tf\|_Y \leq C\|f\|_X$ is being used as both the conclusion and hypothesis of the induction argument, and a potentially unbounded number of iterations of this inductive argument are needed to reach the an arbitrary number of scales. Hence one must often take far more care with the growth constants $C$ than in other types of arguments (although error terms of lower order can often still be handled quite crudely). Indeed, in order to close the induction, it is often necessary to modify the induction hypothesis somewhat, for instance by proving a modified estimate such as $\|Tf\|_Y \leq B(\alpha_1, \ldots, \alpha_k)$ where the $\alpha_j$ are a finite number of parameters depending on $f$ (for instance, one of the $\alpha_j$ might be the $X$ norm of $f$, while another might be some sort of average value of $f$), and $B(\alpha_1, \ldots, \alpha_k)$ is some explicit function (the *Bellman function*) of these parameters which is comparable to (but not exactly equal to) the $X$ norm of $f$; sometimes one also needs to modify the left-hand side $\|Tf\|_Y$ in a similar manner). Because we only modify the right-hand side or left-hand side up to a constant, proving this estimate is equivalent to proving the original estimate; however the advantage of performing a modification like this is that one may be able to enforce some sort of "convexity" on the Bellman function $B$ which makes it possible to close the induction with no loss of constants

[4]To give some indication of the difficulty, we present one simple but still open problem in the field. From Bessel's inequality it is easy to see that $\sum_n c_n e^{2\pi i n^2 x} \in L^2(\mathbf{R}/\mathbf{Z})$ whenever $c_n$ is square summable. It is conjectured but known that these functions are not only in $L^2$, but are in fact in $L^p$ for all $p < 4$, but no result for any $p > 2$ is known.

whatsoever; this convexity might not be present in the original formulation of the problem. In principle, the problem can now be reduced to the elementary calculus task of constructing an explicit Bellman function of a finite number of variables which obeys certain convexity inequalities; however, the actual discovery of such a function can still be a non-trivial task, even if the verification of the desired properties once the function is found is usually straightforward. There are cruder versions of this strategy available in which one is prepared to lose a small factor as the number of scales increase, in which case one does not have to be nearly as careful with selecting a Bellman function $B$. Somewhat in a similar spirit to these approaches are the *variational methods*, in which one tries to optimize $\|Tf\|_Y$ while holding $\|f\|_X$ fixed, or vice versa; ideally this type of approach not only gives the bound, but also the sharpest possible value of the constant $C$ and also knowledge of what functions $f$ extremize (or approximately extremize) this inequality. These methods have proven successful with operators that are highly geometric (e.g. in finding the best constants for the Sobolev inequalities, which are related to the isoperimetric inequality) but have not as yet been developed into a tool suitable for handling general classes of operators such as most of the classes discussed above.

Department of Mathematics, UCLA, Los Angeles CA 90095-1555

*E-mail address*: tao@math.ucla.edu