HARMONIC ANALYSIS

TERENCE TAO

Analysis in general tends to revolve around the study of general classes of *functions* (often real-valued or complex-valued) and *operators* (which take one or more functions as input, and return some other function as output). Harmonic analysis[1] focuses in particular on the *quantitative* properties of such functions, and how these quantitative properties change when apply various (often quite explicit) operators. A good example of a quantitative property is for a function $f(x)$ being uniformly bounded in magnitude by an explicit upper bound M, or perhaps being square integrable with some bound A, thus $\int |f(x)|^2 \, dx \leq A$. A typical question in harmonic analysis might then be the following: if a function $f : \mathbf{R}^n \to \mathbf{R}$ is square integrable, and its gradient ∇f exists and is also square integrable, does this imply that f is uniformly bounded? (The answer is yes when $n = 1$, no when $n > 2$, and just barely no when $n = 2$; this is a special case of the *Sobolev embedding theorem*, which is of fundamental importance in the analysis of PDE.) If so, what are the precise bounds one can obtain?

Real and complex functions, such as a real-valued function $f(x)$ of one real variable $x \in \mathbf{R}$, are of course very familiar in mathematics, starting back in high school. In many cases one deals primarily with *special functions* - polynomials, exponentials, trigonometric functions, and other very explicit and concrete functions. Such functions typically have a very rich algebraic and geometric structure, and there are many techniques from those fields of mathematics that can be used to give exact solutions to many questions concerning these functions.

In contrast, analysis is more concerned with *general* classes of functions - functions which may have some qualitative property such as measurability, boundedness, continuity, differentiability, smoothness, analyticity, integrability, decay at infinity, etc., but which cannot be usefully expressed as a special function, and thus has little or no algebraic or geometric structure. These types of generic functions occur quite frequently for instance in the study of ordinary and partial differential equations, since the solutions to such equations often cannot be given in an explicit algebraic form, but are nevertheless known to obey various qualitative properties such as differentiability. In other cases, the functions can be very explicit and structured, but for one reason or another it is difficult to exploit this structure in a purely algebraic manner, and one must rely (at least in part) on more analytical tools

[1]Strictly speaking, this sentence describes the field of *real-variable harmonic analysis*. There is another field of *abstract harmonic analysis*, which is primarily concerned with how real or complex-valued functions (often on very general domains) can be studied using symmetries such as translations or rotations (for instance via the Fourier transform and its relatives); this field is of course related to real-variable harmonic analysis, but is perhaps closer in spirit to representation theory and functional analysis, and will not be discussed here.

instead. A typical example is the Airy function $\mathrm{Ai}(x) := \int_{-\infty}^{\infty} e^{i(x\xi+\xi^3)}\, d\xi$, which is given as an explicit integral, but in order to understand such basic questions as whether $\mathrm{Ai}(x)$ is always a convergent integral, and whether this integral goes to zero as $x \to \pm\infty$, it is easiest to proceed using the tools of harmonic analysis. In this case, one can use the *principle of stationary phase* to answer both these questions affirmatively, although there is the perhaps surprising fact that the Airy function decays almost exponentially fast as $x \to +\infty$, but only polynomially fast as $x \to -\infty$.

Harmonic analysis, as a sub-field of analysis, is particularly interested in the study of quantitative bounds on these functions. For instance, instead of merely assuming that a function f is bounded, one could ask *how* bounded it is - in particular, what is the best bound $M \geq 0$ available such that $|f(x)| \leq M$ for all (or almost all) $x \in \mathbf{R}$; this number is known as the *sup norm* or L^∞ *norm* of f and is denoted $\|f\|_{L^\infty}$. Or instead of assuming that f is absolutely integrable, one can quantify this by introducing the L^1 *norm* $\|f\|_{L^1} := \int |f(x)|\, dx$; more generally one can quantify p^{th}-power integrability for $0 < p < \infty$ via the L^p *norm* $\|f\|_{L^p} := (\int |f(x)|^p\, dx)^{1/p}$. Similarly, most of the other qualitative properties mentioned above can be quantified by a variety of *norms*, which assign a non-negative number (or $+\infty$) to any given function and which provide some quantitative measure of one characteristic of that function. Besides being of importance in pure harmonic analysis, quantitative estimates involving these norms are also useful in applied mathematics, for instance in performing an error analysis of some numerical algorithm.

Functions tend to have infinitely many degrees of freedom, and it is thus unsurprising that the number of norms one can place on a function are similarly infinite; there are many ways one can quantify how "large" a function is. These norms can often differ quite dramatically from each other. For instance, it is possible for a function f to have large L^∞ norm but small L^1 norm (think of a very spiky function which is large on a very small set, but zero elsewhere), or conversely to have small L^∞ norm but large L^1 norm (think of a very broad function which is very small but spread out over a very large set). Similarly if one compares the L^2 norm with the L^1 or L^∞ norms. However, it turns out that the L^2 norm lies in some sense "between" the L^1 and L^∞ norms, in the sense that if one controls *both* the L^1 and L^∞ norms, then one also automatically controls the L^2 norm. Intuitively, the point is that L^∞ control eliminates all the spiky functions, and L^1 control eliminates most of the broad functions, and the remaining functions end up being well behaved in the intermediate norm L^2. More quantitatively, we have the inequality

$$\|f\|_{L^2} \leq \|f\|_{L^1}^{1/2} \|f\|_{L^\infty}^{1/2}$$

which is a simple consequence of the algebraic fact that if $|f(x)| \leq M$, then $|f(x)|^2 \leq M|f(x)|$. This is a special case of *Hölder's inequality*, which is one of the fundamental inequalities in harmonic analysis; the idea that control of two "extreme" norms automatically implies further control on "intermediate" norms can be generalized tremendously and leads to the very powerful and convenient methods of *interpolation theory*, which is another basic tool in harmonic analysis.

The study of a single function and all its norms eventually gets somewhat tiresome, though. As in many other fields of mathematics, the subject gets a lot more interesting when one not only considers these objects (functions) in isolation, but also introduces *maps* from one object to another; these maps, which take one (or more) functions as input and returns another as output, are usually referred to as *operators* or *transforms*. Operators may seem like fairly complicated mathematical objects - after all, their inputs and outputs are functions, which in turn have inputs and outputs which are usually numbers - however they encode many natural operations one performs on these functions, such as *differentiation*

$$f(x) \mapsto \frac{d}{dx} f(x)$$

and its well-known (partial) inverse, *integration*

$$f(x) \mapsto \int_{-\infty}^{x} f(y)\ dy.$$

A less intuitive, but particularly important, example is the *Fourier transform*

$$f(x) \mapsto \hat{f}(x) := \int_{-\infty}^{\infty} e^{-2\pi i x y} f(y)\ dy. \tag{1}$$

It is also of interest to consider operators which take two or more inputs; two particularly common examples are *pointwise product*

$$(f(x), g(x)) \mapsto f(x)g(x),$$

and *convolution*

$$(f(x), g(x)) \mapsto f * g(x) := \int_{-\infty}^{\infty} f(y)g(x-y)\ dy.$$

There are of course many other operators of interest in harmonic analysis. Historically, harmonic analysis was first concerned with the operations that were connected to Fourier analysis, real analysis, and complex analysis; nowadays, however, the methods of harmonic analysis have been brought to bear on a much broader set of operators. These techniques have been particularly fruitful in understanding the solutions of various linear and non-linear partial differential equations (with the solution being viewed as an operator applied to the initial data), as well as in analytic and combinatorial number theory, when one is faced with understanding the oscillation present in various expressions such as exponential sums. Harmonic analysis has also been applied to analyze operators which arise in geometric measure theory, probability theory, ergodic theory, numerical analysis, and differential geometry.

A primary concern of harmonic analysis is in obtaining both qualitative and quantitative information about how these sorts of operators act on generic functions. A typical example of a quantitative estimate is the inequality $\|f*g\|_{L^\infty} \leq \|f\|_{L^2}\|g\|_{L^2}$ for all $f, g \in L^2$, which is a special case of Young's inequality and is easily proven using the Cauchy-Schwarz inequality; as a consequence of this we have the qualitative conclusion that the convolution of two L^2 functions is necessarily continuous (this is basically because the continuous functions form a closed subspace of L^∞, and because L^2 functions can be approximated to arbitrary accuracy by smooth, compactly supported functions). We give some further examples of qualitative and quantitative analysis of operators in the next section.

1. Example: Fourier summation

To illustrate the interplay between quantitative and qualitative results, we sketch some of the basic theory of summation of Fourier series, which historically was one of the main motivations for studying harmonic analysis in the first place.

In this section, the function f under consideration will always be assumed be periodic with period 2π, thus $f(x+2\pi) = f(x)$ for all x; for instance, f could be a trigonometric polynomial such as $f(x) = 3 + \sin(x) - 2\cos(3x)$. If f is also a continuous function (or at least an absolutely integrable one), then we can define the Fourier coefficients $\hat f(n)$ for all integers n by the formula

$$\hat f(n) := \frac{1}{2\pi} \int_0^{2\pi} f(x) e^{-inx}\ dx.$$

The theory of Fourier series then suggests one has the identity

$$f(x) = \sum_{n=-\infty}^{\infty} \hat f(n) e^{inx}$$

but the rigourous justification of this identity requires some effort. If f is a trigonometric polynomial (i.e. a finite linear combination of functions of the form $\sin(nx)$ and $\cos(nx)$) then all but finitely many of the coefficients $\hat f(n)$ are zero, and the identity is easily verified; however the problem becomes more subtle when f is not a trigonometric polynomial. It is then natural to introduce the *Dirichlet summation operators* S_N for $N = 0, 1, 2, 3, \ldots$ by

$$S_N f(x) := \sum_{n=-N}^{N} \hat f(n) e^{inx}.$$

One can then ask whether $S_N f$ necessarily converges to f as $N \to \infty$. The answer to this question turns out to be surprisingly complicated, depending on how one defines "convergence", and on what assumptions one places on the function f. For instance, one can construct examples of continuous f for which $S_N f$ fails to converge uniformly to f, or even to converge pointwise; however, $S_N f$ will necessarily converge to f in the L^p topology for any $0 < p < \infty$, and will also converge pointwise almost everywhere (i.e. the set where pointwise convergence fails will have measure zero). If instead one only assumes f to be absolutely integrable, then it is possible for the partial sums $S_N f$ to diverge pointwise everywhere, as well as being divergent in the L^p topology for any $0 < p \leq \infty$. All of these results ultimately rely on very quantitative results in harmonic analysis, and in particular on various L^p type estimates on the Dirichlet sum $S_N f(x)$, as well as the closely related maximal operator $\sup_{N>0} |S_N f(x)|$.

As these results are a little tricky to prove, let us first discuss a simpler result, in which the Dirichlet summation operators S_N are replaced by the *Fejér summation operators* F_N, defined for $N = 1, 2, \ldots$ by the formula

$$F_N := \frac{1}{N}(S_0 + \ldots + S_{N-1}).$$

One can verify the identity

$$F_N f(x) = \int_{-\pi}^{\pi} \frac{\sin^2(Ny/2)}{N \sin^2(y/2)} f(x-y)\ dy.$$

Also, it is easy to show that $F_N f$ converges uniformly to f whenever f is a trigonometric polynomial, since this will imply that $S_N f = f$ for all but finitely many values of N. To extend this result from trigonometric polynomials to a larger class of functions, such as continuous functions, let us make the quantitative estimate

$$\|F_N f\|_{L^\infty} \leq \|f\|_{L^\infty}$$

for all continuous periodic functions f and all $N \geq 1$. Indeed, from the triangle inequality (and the positivity of $\frac{\sin^2(ny/2)}{n \sin^2(y/2)}$) we have

$$|F_N f(x)| \leq \int_{-\pi}^{\pi} \frac{\sin^2(ny/2)}{n \sin^2(y/2)} \|f\|_{L^\infty}\ dy = F_N 1(x) \|f\|_{L^\infty}.$$

But $F_N 1 = 1$, and the claim follows. Using this quantitative estimate, one can now deduce that $F_N f$ converges uniformly to f for any continuous f. To see this, first observe by the Weierstrass approximation theorem that given any continuous f and any $\varepsilon > 0$, there exists a trigonometric polynomial g such that $\|f - g\|_{L^\infty} \leq \varepsilon$. Applying the above estimate (and the linearity of F_N) we also have $\|F_N f - F_N g\|_{L^\infty} \leq \varepsilon$ for all N. Finally, since g is a trigonometric polynomial we have $\|g - F_N g\|_{L^\infty} \leq \varepsilon$ for all sufficiently large N. By the triangle inequality we conclude that $\|f - F_N f\|_{L^\infty} \leq 3\varepsilon$ for all sufficiently large N, and the claim follows.

A similar argument (using Minkowski's integral inequality instead of the triangle inequality) shows that $\|F_N f\|_{L^p} \leq \|f\|_{L^p}$ for all $1 \leq p \leq \infty$, $f \in L^p$ and $N \geq 1$. As a consequence, one can modify the above argument to show that $F_N f$ converges to f in the L^p topology for every $f \in L^p$. A slightly more difficult result (relying on a basic result in harmonic analysis known as the *Hardy-Littlewood maximal inequality*) asserts that for every $1 < p \leq \infty$, there exists a constant C_p such that one has the maximal inequality $\| \sup_N |F_N f| \|_{L^p} \leq C_p \|f\|_{L^p}$ for all $f \in L^p$; as a consequence, one can show that $F_N f$ converges to f *pointwise almost everywhere* for every $f \in L^p$ and $1 < p \leq \infty$. A slight modification of this argument also allows one to treat the endpoint case when f is merely assumed to be absolutely integrable; see the discussion on the Hardy-Littlewood maximal inequality in the next section.

Now we return briefly to Dirichlet summation. Using moderately sophisticated techniques in harmonic analysis (such as Calderón-Zygmund theory) one can show that when $1 < p < \infty$, the Dirichlet operators S_N are bounded in L^p uniformly in N, or in other words there exists a constant $0 < C_p < \infty$ such that $\|S_N f\|_{L^p} \leq C_p \|f\|_{L^p}$ for all $f \in L^p$ and all N. As a consequence, one can show that $S_N f$ converges to f in the L^p topology for all $f \in L^p$ and $1 < p < \infty$. However, the quantitative estimate on S_N fails at the endpoints $p = 1$ and $p = \infty$, and from this one can also show that the convergence result also fails at these endpoints (either by explicitly constructing a counterexample, or by using general results such as the uniform boundedness principle). The question of almost everywhere pointwise convergence is significantly harder. It is known that one has an estimate of the form $\| \sup_N |S_N f| \|_{L^p} \leq C_p \|f\|_{L^p}$ for $1 < p < \infty$; this result (the Carleson-Hunt theorem) implies in particular that the Dirichlet sums of an L^p function converge almost everywhere when $1 < p \leq \infty$. On the other hand, this estimate fails at the endpoint $p = 1$, and in fact there is an example of Kolmogorov of an absolutely integrable function whose Dirichet sums are everywhere divergent. These results require quite a lot of harmonic analysis theory, in particular using many decompositions of both the spatial variable and the frequency variable, keeping the Heisenberg uncertainty principle in mind, and then reassembling these pieces carefully and exploiting various manifestations of orthogonality.

To summarize, quantitative estimates such as L^p estimates on various operators provide an important route to establishing qualitative results, such as convergence of certain series or sequences. In fact there are a number of principles (notably the uniform boundedness principle and Stein's maximal principle) which assert that in certain circumstances this is the *only* route, in the sense that a quantitative estimate must exist in order for the qualitative result to be true.

2. Some general themes in harmonic analysis: decomposition, oscillation, geometry

One feature of harmonic analysis methods is that they tend to be *local* rather than *global*; for instance, it is quite common for a function f to be analyzed by applying cutoff functions in either the spatial or frequency variables to decompose f into a number of localized pieces. These pieces would then be estimated separately and then recombined later. One reason for this "divide and conquer" strategy is that a generic function f tends to have many different features (e.g. f could be very "spiky", "discontinuous", or "high frequency" in some places, and "smooth" or "low frequency" in others), and it would be difficult to treat all of these features at once. A well chosen decomposition of the function f can isolate these features from each other, so that each component only has one salient feature that could cause difficulty. In reassembling the estimates from the individual components, one can use crude tools such as the triangle inequality, or more refined tools, for instance those relying on some sort of orthogonality, or perhaps a clever algorithm that groups the components into manageable clusters. The main drawback of the decomposition method (other than aesthetic concerns) is that it tends to give bounds that are not quite optimal; however in many cases one is content with estimates which differ from the best possible one by a multiplicative constant.

To give a simple example of the method of decomposition, let us consider the Fourier transform $\hat{f}(\xi)$ of a function $f : \mathbf{R} \to \mathbf{C}$, defined (for suitably nice functions f) by the formula

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

From the triangle inequality it is clear that if f lies in L^1, then $\hat{f}$ lies in L^∞. The Plancherel theorem implies, among other things, that if f lies in L^2, then $\hat{f}$ also lies in L^2. The question is then what happens if f lies in an intermediate L^p space for some $1 < p < 2$. Since L^p is not contained in either L^1 or in L^2, one cannot use either of the above two results directly. However, by decomposing f into two pieces, one supported on where f is large (e.g. where $|f(x)| \geq \lambda$ for some suitable threshold λ) and one where f is small (e.g. $|f(x)| < \lambda$), then one can apply the triangle inequality to the first piece (which will be in L^1, since $|f(x)| \leq |f(x)|^p/\lambda^{p-1}$ here) and the Plancherel theorem to the second piece (where $|f(x)|^2 \leq \lambda^{2-p}|f(x)|^p$) to obtain a good estimate. Indeed, by utilizing this strategy for all λ and then combining all these estimates together, one can obtain the *Hausdorff-Young inequality*, which asserts that for every $1 < p < 2$ there exists a constant C_p such that $\|\hat{f}\|_{L^{p'}} \leq C_p\|f\|_{L^p}$ for all $f \in L^p$, where $p' := p/(p-1)$ is the dual exponent to p. This particular decomposition method is an example of the method of *real interpolation*. It does not give the best possible value of C_p, which turns out to be $p^{1/2p}/p'^{1/2p'}$ and is computed by more delicate methods.

Another basic theme in harmonic analysis is the attempt to quantify the elusive phenomenon of *oscillation*. Intuitively, if an expression oscillates wildly in phase, then its average value should be relatively small in magnitude. For instance, if a 2π-periodic function f is smooth, then its Fourier coefficients $\hat{f}(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)e^{-inx}$ will be rapidly decreasing as $n \to \pm\infty$ (in other words $\lim_{n\to\pm\infty} |n|^k|\hat{f}(n)| = 0$ for any fixed k). This assertion is easily proven by repeated integration by parts. Generalizations of this phenomenon include the *principle of stationary phase*, which among other things allows one to obtain precise control on the Airy function $\mathrm{Ai}(x)$ discussed earlier, as well as the *Heisenberg uncertainty principle*, which relates the decay and smoothness of a function to the decay and smoothness of its Fourier transform.

A somewhat different manifestation of oscillation lies in the principle that if a sequence of functions oscillate in different ways, then their sum should be smaller than what the triangle inequality would predict. For instance, the Plancherel theorem in Fourier analysis implies, among other things, that a trigonometric polynomial $\sum_{n=-N}^{N} c_n e^{inx}$ has an L^2 norm of

$$(\frac{1}{2\pi}\int_0^{2\pi} |\sum_{n=-N}^{N} c_n e^{inx}|^2)^{1/2} = (\sum_{n=-N}^{N} |c_n|^2)^{1/2},$$

which is smaller than the upper bound of $\sum_{n=-N}^{N} |c_n|$ that can be obtained from the triangle inequality. This identity can be viewed as a special case of *Pythagoras' theorem*, together with the observation that the harmonics e^{inx} are all *orthogonal* to each other with respect to the inner product $\langle f, g\rangle := \frac{1}{2\pi}\int_0^{2\pi} f(x)\overline{g(x)}\, dx$. This concept of orthogonality has been generalized in a number of ways, for instance to a more general and robust concept of *almost orthogonality*, which roughly speaking means that a collection of functions have inner products which are small rather than vanishing completely.

Many arguments in harmonic analysis will, at some point, involve a combinatorial statement about certain types of geometric objects such as cubes, balls, or boxes. For instance, one useful such statement is the *Vitali covering lemma*, which asserts that given any collection $B_1, \ldots, B_k$ of balls in Euclidean space $\mathbf{R}^n$, there exists a subcollection of balls $B_{i_1}, \ldots, B_{i_m}$ which are disjoint, and furthermore contain a significant fraction of the volume covered by the original balls, in the sense that

$$\mathrm{vol}(\bigcup_{j=1}^{m} B_{i_j}) \geq 5^{-n}\mathrm{vol}(\bigcup_{j=1}^{k} B_j).$$

(The constant 5^{-n} can be improved, but this will not concern us here.) This result is proven by a standard greedy algorithm argument, where at each stage of the algorithm one selects the largest ball amongst the B_j which is disjoint from all the balls already selected. One consequence of this lemma is the *Hardy-Littlewood maximal inequality*

$$\mathrm{vol}\{x \in \mathbf{R}^n : \sup_{r>0} \frac{1}{\mathrm{vol}(B(x,r))} \int_{B(x,r)} |f(y)|\, dy > \lambda\} \leq 5^n \frac{\|f\|_{L^1}}{\lambda}$$

for any $\lambda > 0$ and $f \in L^1(\mathbf{R}^n)$, where $B(x,r)$ denotes the ball of radius r centred at x; this is proven by observing that the set appearing on the left-hand side can be covered by balls $B(x,r)$ on which the integral of $|f|$ is at least $\lambda \mathrm{vol}(B(x,r))$, and then applying the Vitali covering lemma. This quantitative inequality then implies as a qualitative consequence the *Lebesgue differentiation theorem*, which asserts that for all absolutely integrable f on $\mathbf{R}^n$, we have

$$\lim_{r\to 0} \frac{1}{\mathrm{vol}(B(x,r))} \int_{B(x,r)} f(y)\, dy = f(x)$$

for almost every $x \in \mathbf{R}^n$. This example demonstrates the importance of the underlying geometry (in this case, the combinatorics of metric balls) in harmonic analysis.

3. Acknowledgements

The author thanks Songsong Lu for corrections and comments.

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

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