Advanced Real Analysis

Harvard University — Math 212a
Course Notes, C. McMullen

Contents

1 Introduction

Aims and scope. The course aims to provide a sequel to a typical course in Lebesgue integration and Banach spaces. The sequel focuses on the following broad areas of analysis:

1. Locally convex topological vector spaces, especially Fréchet spaces;
2. Distributions;
3. The Fourier transform;
4. Banach algebras; and
5. The spectral theorem for self–adjoint and normal operators.

Many additional topics fall inside these broad areas, such as: linear partial differential equations, Sobolev spaces, elliptic regularity, Wiener's proof of the prime number theorem, ergodic theory, quantum mechanics, and probability theory.

The motivation for the first three topics is, in large part, to deal systematically with C^∞ functions in the same way one deals with, say, L^2 functions, and in particular to provide an adequate duality theory via distributions. This requires a broader arena for functional analysis than just Banach spaces.

Another central theme is the study of harmonic analysis on the Lie groups $\mathbb{R}$ and $\mathbb{R}^n$, going beyond the case of Fourier series on $\mathbb{Z}$ and S^1 and preparing for the study of representations of noncompact Lie groups.

Our main reference will be: Rudin, *Functional Analysis.* Other useful references include:

Berberian, *Lectures on Functional Analysis and Operator Theory*;
Reed and Simon, *Functional Analysis*; and
Riesz and Nagy, *Functional Analysis.*

Here is a glimpse of some of the perspectives that will emerge.

1. *Convolution.* The Fourier transforms $\widehat{f}(\xi)$ gives the 'diagonal entries' of the operator, convolution with f.

   Similarly for any other translation-invariant operator, such as d/dx (which goes over to multiplication by $i\xi$).

2. *Prime numbers and ideals in $L^1(\mathbb{R})$.* The space L^1 forms a Banach algebra under convolution, and Wiener's theorem characterizes its ideals in terms of points where the Fourier transform vanishes. When the Fourier transform is nowhere vanishing, the ideal is as large as possible. It turns out that $\zeta(1+is)$ can be regarded as the Fourier transform of a suitable measure on the prime numbers, and the fact that it has no zeros underlies the proof of the prime number theorem.

3. *Functional calculus.* Once you have your Hilbert space spread out before you — $H = \int H_t \, dm(t)$ — 'like a patient etherized upon the table' — it is very easy to discuss operators, even unbounded ones, since they are just functions on $(\mathbb{R}, m)$.

4. *Quantum theory.* The *logic* of quantum mechanics is that the state of a system is represented by a unit vector $v \in H$, and a proposition about

the system corresponds to a closed subspace $A \subset H$. The *negation* of A is the *complementary* subspace, $A^\perp$. Then the probability of A (or not A) being true is given by $\|\pi(v)\|^2$, where π is the projection of v to A (or $A^\perp$).

Example. In traditional probability, we might have a distribution on a measure space (X, m) given by $|f(x)|^2$ where $f \in L^2(X, m)$. Then an event is specified by a measurable subset $A \subset X$, which determines a closed subspace $L^2(A) \subset L^2(X)$. The projection is given by $\pi_A = \chi_A$ acting by multiplication. The negation of A corresponds to the set $A' = X - A$ with $\pi_{A'} = I - \pi_A$ and with perpendicular subspace $L^2(A')$. Finally the probability of A occurring is given by $\int_A |f|^2 = \|\chi_A f\|^2$.

In classical probability theory, the projections *commute*; that is, $\chi_A \chi_B = \chi_B \chi_A$. In the quantum theory, they need not.

Finally, the *dynamics* of the quantum theory is given by an evolution of the states that preserve norms: $v_t = U_t v$, where U_t is a semigroup of unitary operators.

2 Quick review of classical real analysis

We will assume some background in Lebesgue measure and operator theory which we briefly summarize here. See e.g. Royden, *Real Analysis*.

Littlewood's 3 principles. These describe properties of measurable sets, functions and sequences of such functions. They say:

1. A set of finite measure is nearly a finite union of intervals;
2. A measurable function is nearly a continuous function; and
3. A pointwise convergent sequence of measurable functions is nearly uniformly convergent.

Let us briefly recall that the Borel sets in $\mathbb{R}$ (or any topological space) are the smallest σ-algebra containing the open sets; that a set E has measure zero if we have $E \subset \bigcup_1^\infty [a_i, b_i]$ with $\sum |b_i - a_i|$ as small as we like; and a set E is measurable if it differs from a Borel set B by a set N of measure zero ($E = B \triangle N$).

Baire category. A useful principle for analysis on complete metric spaces X is the theorem of Baire: if U_n is a sequence of dense open sets, then

$G = \bigcup_1^\infty U_n$ is also dense. In a sense these dense G_δ sets are analogous to sets of full measure. However they can be very different: the set of Liouville numbers is a dense G_δ of measure zero.

Remark: there is no function continuous at the rationals and discontinuous at the irrationals. In fact the set of points of continuity of f can be written as $E = \bigcup_1^\infty U_n$, where

$$U_n = \{x \,:\, \operatorname{diam} f(U) < 1/n \text{ for some open set } U \text{ containing } x\},$$

so it is a G_δ; the rationals are not.

Measurable functions and L^p spaces. A function is measurable if $f^{-1}(U)$ is measurable for all open sets U; equivalently, for all Borel sets U.

The main remarkable property of measurable functions is that they are closed under pointwise limits. Because of this, the spaces such as $L^p(\mathbb{R})$ are *complete.* A classical motivation of for developing the Lebesgue theory was to give a meaning to the statement:

$$f(x) = \sum a_n e^{inx}$$

whenever $\sum |a_n|^2 < \infty$. The modern solution is that f is an element of $L^2(S^1)$ and the convergence is in norm. (It is a much harder theorem, due to Carleson (1966), that one has pointwise convergence a.e.)

The idea of a distribution. One the main topics of this course will be a very different answer to the question, 'what is an L^2 function?' The answer will be: an L^2 function is a continuous map $T : C_c^\infty(\mathbb{R}^n) \to \mathbb{R}$, such that

$$|T(f)| \le C \int |f|^2$$

for every compactly supported smooth function f. Note that we do not try to describe T as an object pointwise at all.

Dominated convergence. A typical issue in Lebesgue integration is to determine when $f_n \to g$ implies that $\int f_n \to \int g$. There are two ways this can fail: mass can escape to infinity horizontally, and vertically. These are both prevented if $|f_n| \le h$ and $\int h < \infty$; this is the *dominated convergence theorem.*

Banach spaces. The central features of Banach spaces are their completeness, convexity, and duality. Completeness, via the Baire category theorem, underlies the 3 main principles of functional analysis:

1. The open mapping theorem: if $f : X \to Y$ is a surjective, continuous linear map between Banach spaces, then f is open. In particular, if f is bijective, then f^{-1} is continuous.

   In particular, if f is a bijection, then its inverse is continuous.

2. The closed graph theorem: if the graph of f is closed in $X \times Y$, then f is continuous.

3. The uniform boundedness principle: if $\Phi \subset X^*$ is a set of linear functions and $|\Phi(x)|$ is bounded for all $x \in X$, then $\|\Phi\|$ is bounded in X^*.

Let us remark on the proofs of the first two results. Let $B \subset X$ be the unit ball. If f is surjective then $Y = \bigcup nf(B)$, and has $\overline{f(B)}$ has nonempty interior by Baire category. In fact one can show that $\overline{f(B)}$ contains $B(0, r)$ for some $r > 0$. This means that for any $y \in Y$ we can find an $x \in X$, $\|x\| \leq M\|y\|$, with $f(x)$ as close to y as we like; where $M = 1/r$. For example if $\|y\| \leq 1$ we can find x_1 such that $\|f(x_1) - y\| \leq 1/2$; then x_2 such that $\|f(x_2) + f(x_1) - y\| \leq 1/4$, etc, with $\|x_i\| \leq M/2^i$. Then $x = \sum x_i$ has norm at most $2M$ and $f(y) = y$. In other words the image of the ball of radius $2M$ contains the unit ball in Y.

For the closed graph theorem, the trick is to define a new norm on X by $\|x\|' = \|x\| + \|f(x)\|$; show it is complete; and then apply the open mapping theorem to deduce a bound in the reverse direction, $\|f(x)\| \leq \|x\|' \leq M\|x\|$.

Duality. To produce many elements of X^* in general one uses:

Theorem 2.1 (Hahn–Banach) *Let $\phi_S : S \to \mathbb{R}$ be a continuous linear map defined on a subspace $S \subset X$. Then ϕ_S extends to map $\phi \in X^*$ with $\|\phi\| = \|\phi_S\|$.*

Corollary 2.2 *Every Banach space X embeds isometrically into X^{**}.*

Proof. We can find ϕ such that $\|\phi\| = 1$ and $\phi(x) = \|x\|$, whenever $x \neq 0$. ■

In particular, X^* separates the points of X.

Duality for L^p spaces. In the classical Banach spaces the dual spaces are concretely visible: L^p and L^q are dual when $1/p + 1/q = 1$; in particular, the Hilbert space L^2 is self–dual; *except* $(L^\infty)^*$ is not L^1, in general.

To see this last statement, one can use the Hahn–Banach theorem to extend a point evaluation from $C[0,1] \subset L^\infty[0,1]$ to the whole space. Another proof will be given below: L^1 is not the dual of *anything*.

Continuous functions. Let $C(K)$ denote the Banach algebra of continuous functions on a compact metric space K. We then have:

The *Stone–Weierstrass theorem* (over $\mathbb{R}$): A subalgebra $A \subset C(K)$ is dense if and only if it separates points and contains the constant functions.

The *Arzela–Ascoli theorem*: A subset $F \subset C(K)$ has compact closure if and only if it is bounded and equicontinuous.

The *Riesz representation theorem*: The dual of $C(K)$ is exactly the space $M(K)$ of Borel measures on K.

Weak and weak* topology. Recall that X has a natural *weak topology*, which makes all linear functionals $\phi \in X^*$ continuous but is otherwise as weak as possible. In this topology, $x_\alpha \to y$ iff $\phi(x_\alpha) \to \phi(y)$ for all $\phi \in X^*$. (We have written convergence in terms of *nets* since the weak topology need not be metrizable.)

The dual space has a *weak* topology*, in which $\phi_\alpha \to \psi$ iff $\phi_\alpha(x) \to \psi(x)$ for all $x \in X$. This is the same as the topology of pointwise convergence on X. Using compactness of $X^{[0,1]}$, it is easy to demonstrate:

Theorem 2.3 (Alaoglu's theorem) *The unit ball in X^* is compact in the weak* topology.*

A typical application: the space of probability measures on a compact set K is compact in the weak* topology.

Convexity. The weak and weak* topologies both have bases consisting of convex sets. Thus they make X and X^* into *locally convex* topological vector spaces.

Let $K \subset X$ be a convex set, and let $[a,b]$ denote the interval joining any two points in X. A point $p \in K$ is an *extreme* point if whenever $p \in [a,b] \subset K$, we have $p = a = b$.

Theorem 2.4 (Krein–Milman) *Let K be a compact convex set in a locally convex topological vector space X. Then K is the closed convex hull of its extreme points.*

Corollary 2.5 $L^1(\mathbb{R})$ *is not a dual space.*

Proof. The unit ball in L^1 has no extreme points. ∎

Here is a related statement due to Milman:

Theorem 2.6 (Milman) *Let K be a compact convex set in a locally convex topological vector space X. Suppose K is the closed convex hull of a compact set L. Then every extreme point of K lies in L.*

Proof. Let $p \in K$ be an extreme point. Using compactness, cover L by finitely many small compact sets $L_1, \ldots, L_n$, with closed convex hulls $K_1, \ldots, K_n$. Then K_i is close to L_i, and K is just the convex hull of $K_1 \cup \cdots \cup K_n$. Then p must belong to one of the K_i, else it is a nontrivial linear combination of points in K. Since K_i is close to L_i, we can pass to a limit and get that $p \in L$.

Application: recovering K from $C(K)$. It is well–known that if K is a compact Hausdorff space, then the Banach algebra $C(K)$ determines K. In fact K is homeomorphic to the space of multiplicative linear functions $\phi : C(K) \to \mathbb{R}$ in the weak* topology, where *multiplicative* means $\phi(xy) = \phi(x)\phi(y)$. Every such functional is given by point evaluation.

It is also true that the structure of $C(K)$ as a Banach space alone determines K. The reason is that the set of extreme points of the unit ball in $M(K)$ is homeomorphic, in the weak* topology, to two copies of K. These copies come from the measures $\pm\delta_p$ as p varies over K.

3 Applications of classical functional analysis

To familiarize ourselves with more of the results quickly summarized above, let us see them in action.

Symmetric operators. Here is a classical result that gives 'continuity' for free.

Theorem 3.1 *Let $T : H \to H$ be a linear operator on a Hilbert space such that $\langle Tx, y\rangle = \langle x, Ty\rangle$ for all $x, y \in H$. Then T is continuous.*

Proof. We will show the graph of T is closed. Suppose $x_n \to x$ and $y_n = Tx_n \to y$. Then for all z, we have

$$\langle Tx_n, z\rangle = \langle x_n, Tz\rangle \to \langle x, Tz\rangle = \langle Tx, z\rangle,$$

and also $\langle Tx_n, z\rangle \to \langle y, z\rangle$. Since this is true for all z, we have $y = Tx$ and so the graph of T is closed. ∎

Grothendieck's theorem. Here is another elegant application of the open mapping theorem.

Theorem 3.2 *Let $S \subset L^2[0,1]$ be a closed subspace consisting entirely of bounded functions. Then S is finite–dimensional.*

Proof. We certainly have $\|f\|_2 \le \|f\|_\infty$, so S is complete in the L^∞ norm. (I.e. a uniformly convergent sequence in S also converges in L^2, so its limit belongs to S.) By the open mapping theorem we have the reverse inequality: there exists an M such that

$$\|f\|_\infty \le M\|f\|_2$$

for all $f \in S$.

Now let $f_1, \ldots, f_n$ be an orthonormal set in S. Then for any $p \in [0,1]$, we have

$$\| \sum f_i(p) f_i\|_2^2 = \sum |f_i(p)|^2,$$

by orthonormality; and hence

$$M^2 \sum |f_i(p)|^2 \ge \| \sum f_i(p) f_i\|_\infty^2 \ge \left(\sum |f_i(p)|^2\right)^2,$$

which gives

$$M^2 \ge \sum |f_i(p)|^2;$$

integrating over $[0,1]$ we get $M^2 \ge n$. ■

Remark. One can find an *infinite dimensional* closed subspace S of $L^2(S^1)$ such that S is contained in $L^4(S^1)$. The basic idea is to consider the closed span S of space of $(z^i \ : \ i \in I)$, where a given integer n can be written as $i+j$, $i,j \in I$, in at most one way. (For example, take I to be the powers of 10.) Then for $f(z) = \sum_I a_i z^i$, we find

$$f(z)^2 = \sum_{i,j \in I} a_i a_j z^{i+j},$$

and hence

$$\|f\|_4^4 = \int_{S^1} |f|^4 = \left(\sum |a_i|^2\right)^2 = \|f\|_2^2.$$

This shows $S \subset L^4(S^1)$.

Haar measure. The following fundamental result that applies not just to Lie groups like SU_n but also groups like $\mathbb{Z}_p$, $\mathrm{GL}_n(\mathbb{Z}_p)$, $\prod_1^\infty S^1$, etc.

Theorem 3.3 *Let G be a compact topological group. Then there exists a unique right–invariant probability measure on G; it is also left–invariant.*

For the proof we will use a result that is useful in its own right:

Theorem 3.4 (Kakutani) *Let G be a group operating by isometries on a Banach space X, leaving invariant a compact convex set K. Then G has a fixed point in K.*

Proof. Using the Axiom of Choice, one can construct a *minimal* G–invariant compact convex set $L \subset K$. Suppose L contains a nontrivial interval $[x, y]$ with midpoint z. Let M be the closed convex hull of Gz; it is contained in L, so it must equal L. By Krein–Milman, L has an extreme point z'. Since the closure of Gz is compact, we have $z' = \lim g_n z$ by Milman's theorem. We can then, by passing to a subsequence, also form $[x', y'] = \lim g_n([x, y])$. Then z' is the midpoint of this limit, contradicting the fact that it is an extreme point. ■

Lemma 3.5 *In a Banach space, the convex hull of a compact set is compact.*

Proof. Let H be the convex hull of the compact set K. It suffices to show H is totally bounded. Given $r > 0$ choose finitely many balls $B(k_i, r)$ that cover K, and let F be the convex hull of their centers k_i. Then F is (finite dimensional and) compact, and hence F is contained in $F' + B(0, r)$ for some finite set F'. Then we have

$$H \subset F + B(0, r) \subset F' + B(0, 2r),$$

so H is covered by finitely many balls of radius $2r$. Therefore it is totally bounded. ■

Remark: What happens more generally? Here is an example of a compact set $L \subset X$ in a LCTVS such that the closure of hull(L) is not compact. Let $L \subset M[0, 1]$ be the set of δ-masses δ_p, $p \in [0, 1]$, and let X be the linear span of L in the weak* topology. Then L is compact (it is homeomorphic to $[0, 1]$) but its convex hull contains the measures $(1/n) \sum \delta_{k/n} \to dx$ which have no limit in X. (We would get compactness of X were a Fréchet space, see [Ru, Theorem 3.20].)

Proof of Theorem 3.3. The key to the proof is to observe that for any $f \in C(G)$, the convex hull K of $G \cdot f$ is compact, by equicontinuity and the Lemma above. Thus $\mathrm{hull}(G \cdot f)$ contains a G–invariant function c, which is of course a constant. To construct Haar measure we wish to simple set $M(f) = c$, and show that M is a well–defined, positive, continuous linear functional on $C(G)$, and by construction its value only depends on the orbit of G. Moreover its value is the only possible candidate for the average of G under a G–invariant measure.

To check that $M(f)$ is well–defined, we use two facts. First, c can be approximated by the image of f under an averaging operator of the form

$$T_\mu(f) = \sum_1^n a_i f(g_i x),$$

where $0 \leq a_i$ and $\sum a_i = 1$. This sum can be regarded as a convolution of f with the probability measure $\mu = \sum a_i \delta_{g_i}$. Such an operator satisfies $\|T_\mu\| \leq 1$ and $T_\mu(a) = a$ for any constant function a. Second, using right multiplication we can form a constant $c' \approx S_\nu(f)$. Since right and left multiplication commute, we conclude that $c' = c$ and hence both c and c' exist and agree. Thus $M(f)$ is well–defined and both right and left invariant.

It is also clear that $M(f)$ is a positive operator, $|M(f)| \leq \|f\|_\infty$, $M(\lambda f) = \lambda M(f)$, and $M(c) = c$ for any constant c. To see M is linear, observe that $M(T_\mu f) = M(f)$ since $T_\mu(f) \in \mathrm{hull}(G \cdot f)$. Given f and g, choose μ and ν such that $T_\mu(f) = M(f) + \epsilon_1$ and $S_\nu(g) = M(g) + \epsilon_2$. Then

$$\begin{aligned} T_\mu S_\nu(f+g) &= S_\nu(T_\mu f) + T_\mu S_\nu(g) = M(f) + S_\nu(\epsilon_1) + M(g) + T_\mu(\epsilon_2) \\ &= M(f) + M(g) + \epsilon_3, \end{aligned}$$

where $\|\epsilon_3\| \leq \|\epsilon_1| + \|\epsilon_2\|$. Thus

$$M(f+g) = M(T_\mu S_n u(f+g)) = M(f) + M(g) + O(\|\epsilon_3\|),$$

which gives linearity as desired. ∎

Remark. One could also approach the existence of Haar measure by looking directly at the action of G on $M(G)$, but it would be hard to prove uniqueness.

Bishop's proof of the Stone–Weierstrass theorem. The usual proof of this result uses the fact that $|x|$ can be approximated by polynomials, and then some reasoning about lattices of functions. Here is a different proof just using extreme points.

Theorem 3.6 *Let $A \subset C(K)$ be a subalgebra of the real–valued continuous functions on a compact Hausdorff space K, such that A contains the constant functions and separates points. Then $\overline{A} = C(K)$.*

Proof. Let $P \subset C(K)^* = M(X)$ be the closed, convex set of signed measures μ with totally variation $\|\mu\| = \int_K |\mu| \leq 1$ and $\mu \in A^\perp$. Note that P is compact in the weak* topology, so it is the closed convex hull of its extreme points. If $P = (0)$ then $\overline{A} = C(K)$. Otherwise, P has an extreme point μ with $\|\mu\| = 1$.

Since A separates points, and contains the constants, there is a function $f \in A$ that is not constant on the support of μ. We can assume $\|f\| < 1$.

Let $\alpha = (1+f)\mu/2$ and let $\beta = (1-f)\mu/2$. *Since A is an algebra*, both of these measures are in $A^\perp$. Then $\mu = \alpha + \beta$ and, since $1 + f > 0$ and $1 - f > 0$, we have

$$\|\alpha\| + \|\beta\| = \frac{1}{2}\left(\int_K (1+f)|\mu| + \int_K (1-f)|\mu|\right) = \|\mu\|.$$

Then μ is a nontrivial linear combination of the measures $\alpha/\|\alpha\|$ and $\beta/\|\beta\|$ in P, contradicting the assumption that μ is an extreme point. ∎

4 Locally convex topological vector spaces

The key features of Banach spaces are metric completeness, convexity (of the unit ball), and duality (the existence of many continuous linear functionals). To broaden the discussion we introduce topological vector spaces, with the added feature that there exists a convex base for the topology. These are useful for handling spaces of continuous, holomorphic and smooth functions on a domain $\Omega \subset \mathbb{R}^n$; the spaces $C(\Omega)$, $C^\infty(\Omega)$, as well as $C^\infty_c(\Omega)$, do not have natural Banach space topologies, nevertheless there are good notions of convergence and completeness.

For a very detailed reference, see Bourbaki, *Topological Vector Spaces.*

Topological vector spaces. Let X be a vector space over $\mathbb{C}$ or $\mathbb{R}$. Note that the notion of *convexity* just uses linear algebra. Similarly we say a set $S \subset X$ is *balanced* if $\alpha S \subset S$ whenever $|\alpha| \leq 1$.

Let τ be a topology on X. We say (X, τ) is a *topological vector space* if its algebraic operations are continuous, *and* the points of X are closed. In this case τ is determined by a *local base*, meaning a collection of open sets $\{U_i\}$ such that for any neighborhood V of 0, we have $0 \in U_i \subset V$ for some i.

Proposition 4.1 *A topological vector space is Hausdorff.*

Proof. Given $x \neq y$ there is an open neighborhood U of the origin such that $x + U$ does not contain y. By continuity of addition, there is an open neighborhood V of 0 such that $V \pm V \subset U$. Then $x + V$ and $y + V$ are disjoint open neighborhoods of x and y. ∎

By similar reasoning, using scalar multiplication instead of addition, one can show:

Proposition 4.2 *A topological space has a balanced local base.*

A set U is *absorbing* if $\bigcup_{t>0} tU = X$.

Proposition 4.3 *Every neighborhood of the origin is absorbing.*

Proof. If $y \notin \bigcup tU$, then there are $t_n \to 0$ such that $t_n y \notin U$, a contradiction. ∎

Boundedness. This is a topological notion. We say $E \subset X$ is *bounded* if for any neighborhood U of 0, we have $E \subset tU$ for all $t \gg 0$.

For example, any nontrivial subspace S of X is *unbounded*. Indeed, if $y \in S$ is nonzero, then there is a neighborhood U of 0 with $y \notin U$; then $ty \in S - tU$, so no dilate of U can include all of S.

Local compactness. We will almost always be interested in the infinite–dimensional case; in passing we remark:

Proposition 4.4 *If X is locally compact, then X is finite–dimensional.*

Local convexity. We say X is *locally convex* if it has a local basis of convex sets. In these spaces much of classical functional analysis goes through; in particular, we have useful versions of the Hahn–Banach theorem.

Banach spaces from unit balls. Banach spaces are locally convex and locally bounded. Moreover, if we are given a *bounded, balanced, convex, absorbing set B* in a vector space X, we can define a norm by

$$\|x\|_B = \inf\{t \geq 0 \,:\, x \in tB\}$$

and get a topology on X. In this way a normed space is determined by its unit ball.

Unbounded topologies. Note however that there many examples where every neighborhood of 0 *contains a infinite–dimensional subspace*. In this case, X is *not* locally bounded. (If $S \subset U$ is such a subspace, and $y \neq 0$ lies in S, then there exists an open set V containing 0 but not y; and then no dilate tV of V can ever contain U.)

Failure of convexity: $L^{1/2}[0,1]$. Note that $X = L^p[0,1]$, $0 < p < 1$, is complete, metrizable, but not locally convex. In fact, for p in this range we have $(a+b)^p \leq a^p + b^p$, when $a, b > 0$; so if we define

$$d(0,f) = \int_0^1 |f|^p,$$

then $d(0, f+g) \leq d(0,f) + d(0,g)$ and we find X has a complete, translation invariant metric.

However, that convex hull of any open neighborhood of the origin in X is the entire space. This can be seen by writing $f \in L^p[0,1]$ as $\sum_1^n f_i$ where $\int |f_i|^p = (1/n) \int |f|^p$. Then $f = (1/n) \sum n f_i$ is a convex combination of functions with $\|nf_i\| = O(n^{1-1/p}) \to 0$.

As a corollary, X^* is trivial. For if $\phi \in X^*$, then $\phi^{-1}(-1,1)$ provides a convex open neighborhood U of the origin. Then $U = X$ and hence $\phi = 0$.

Note: $X = \ell^p(\mathbb{N})$ is also not locally convex, but X^* is nontrivial (e.g. $(a_n) \mapsto a_1$ is continuous).

Completeness. Let X be a topological vector space. We say $x_n \in X$ is a Cauchy sequence if for every open neighborhood U of 0, we have $x_n - x_m \in U$ for all $n, m \gg 0$; and X is *complete* if every Cauchy sequence has a limit.

Proposition 4.5 *Cauchy sequences are bounded.*

Proof. Let (x_n) be a Cauchy sequence, let U be a neighborhood of 0, and choose a balanced neighborhood V of 0 such that $V + V \subset U$. Then there is an N such that $x_i - x_j \in V$ for all $i, j \geq N$, and there is a $t > 1$ such that $\{x_1, \ldots, x_N\} \in tV$. Then the whole Cauchy sequence is contained in $tV + V \subset t(V+V) = tU$, so it is bounded. ∎

Many of the results for Banach spaces that rely on the Baire category theorem also apply to complete, locally convex topological vector spaces.

Spaces of functions. To illustrate several phenomena and constructions in the locally convex setting, we now examine various spaces of functions on a domain $\Omega \subset \mathbb{R}^n$. These spaces are:

1. The space $C(\Omega)$ of all continuous functions $f : \Omega \to \mathbb{R}$.
2. The space $C^\infty(\Omega)$ of all smooth functions $f : \Omega \to \mathbb{R}$.
3. The space $C_0(\Omega)$ of functions tending to zero at infinity. (This means $f(x_n) \to 0$ if x_n eventually leaves every compact subset of Ω.)
4. The space $C_c(\Omega)$ of compactly supported continuous functions.
5. The space $C_c(K)$ of functions supported on a particular compact set $K \subset \Omega$. (The larger domain Ω is implicit.)

The spaces $C_0(\Omega)$ and $C_c(K)$ are naturally Banach spaces in the sup-norm. We will see that the space $C(\Omega)$ is naturally a *Fréchet space* and an inverse limit, while the space $C_c(\Omega)$ carries a natural *inductive* topology and is a direct limit.

Fréchet spaces and seminorms. It is very convenient if the topology on X is induced by a complete, translation–invariant metric. If in addition, X is locally convex, it is called a *Fréchet space*.

These spaces are often described using seminorms. A seminorm $p : X \to \mathbb{R}_+$ satisfies all the usual properties of a norm, except that $p(x) = 0$ need not imply that $x = 0$. Thus the unit ball B for a seminorm need not be bounded.

A sequence of seminorms p_n on a vector space V determines a metrizable topology, where $x_i \to 0$ if and only if $p_n(x_i) \to 0$ for every i. The associated translation invariant metric can be given by

$$d(0, x) = \sum_1^\infty 2^{-n} \min(1, p_n(x)).$$

The corresponding a topology on $C(\Omega)$ is obtained by requiring that $x_i \to x$ if and only if $p_n(x - x_i) \to 0$ for each n. A local base at the origin is given by

$$U_N = \{x \; : \; p_n(x) < 1/N, n = 1, 2, \ldots, N\}.$$

Proposition 4.6 *A set $B \subset X$ is bounded if and only if there exists a sequence M_n such that $\sup_B p_n(x) \le M_n$ for all n.*

Proof. This is straightforward, using the basis U_N. If B is bounded, then there are t_N such that $B \subset t_N U_N$ for each N. Thus $p_N(x) \le t_N/N$ for all $x \in B$. The converse is similar. ■

The space $C(\Omega)$. An *exhaustion* of an open set $\Omega \subset \mathbb{R}^n$ is a sequence of compact sets $K_1 \subset K_2 \subset \cdots$ such that $\bigcup K_i = \Omega$. Each compact set determines a seminorm on $C(\Omega)$ by:

$$p_i(f) = \sup_{K_i} |f(x)|.$$

Note that there are nonzero f with $p_i(f) = 0$. Then $f_i \to f$ if and only if f_i converges to f uniformly on compact sets. Since a locally uniform limit of continuous functions is continuous, we find:

Theorem 4.7 *The space $C(\Omega)$ is a Fréchet space. That it, it is complete in the topology of uniform convergence on compact sets.*

We note that every bounded set in $C(\Omega)$ is contained in one of the following form:

$$B(M) = \{f \,:\, |f(x)| \le M(x)\},$$

where $M \in C(\Omega)$ is a positive function.

Holomorphic functions. If Ω is contained in $\mathbb{C}^n$, then the space of holomorphic functions $H(\Omega)$ is a *closed subspace* of $C(\Omega)$, and hence a Fréchet space in its own right. As an example on the unit disk in $\mathbb{C}$, we have $f_n(z) = z^n \to 0$ in $H(\Delta)$, even though $\|z^n\|_\infty = 1$ for all n.

The space $C^\infty(\Omega)$. Here we must take derivatives $D^\alpha f$ into account, so we use the seminorms

$$p_n(f) = \sup_{x \in K_n, |\alpha| \le n} |D^\alpha f(x)|.$$

Then $C^\infty(\Omega)$ is also a Fréchet space. It has another remarkable property, not shared by $C(\Omega)$: by the Arzela–Ascoli theorem we have:

Proposition 4.8 *In $C^\infty(\Omega)$, closed, bounded sets are compact.*

However this space is *not* locally bounded, so it is not locally compact.

Inverse limits. We can also regard $C(\Omega)$ as an inverse limit,

$$C(\Omega) = \lim_{\longleftarrow} C(K_n).$$

The general definition is as follows. Let $X_1 \leftarrow X_2 \leftarrow X_3 \cdots$ be a sequence of continuous map between locally convex topological vector spaces. Usually we assume that X_i maps to a dense subset of X_{i-1}. We then give

$$X = \lim_{\longleftarrow} X_i$$

the *strongest* topology such that a linear map

$$T : Y \to X$$

is continuous if and only if the maps $Y \to X \to X_i$ are continuous for every i. A local basis for this topology is given by

$$U(V, i) = \{x \; : \; x \text{ maps into } V \text{ under projection to } X_i\},$$

where i ranges over all indices and V ranges over all neighborhoods of 0 in X_i. This is also the *weakest* topology such that all the maps $X \to X_i$ are continuous.

The space $C_c(K)$. This space denotes the continuous functions on Ω which are supported on a given compact set K. Equipped with the sup-norm, it is a Banach space. Note that all such functions vanish on ∂K. The notation is slightly misleading since Ω is implicit; however f is determined by $f|K$, since f vanishes elsewhere.

The space $C_c(\Omega)$. The space of compactly supported functions has a natural topology making it *complete.* (The sup norm is defined on this space, but does not make it complete; the metric completion is $C_0(\Omega)$.)

First observe that $X = C_c(\Omega)$ is naturally an ascending union of Banach spaces X_n; we have:

$$C_c(\Omega) = \bigcup C_c(K_n) = \bigcup X_n.$$

We give $\bigcup X_i$ the *inductive* topology: a local base consists of those *convex* sets U such that $U \cap X_i$ is open for every i.

Here is a local basis for $C_c(\Omega)$. Given a continuous function $\epsilon(x) > 0$ on Ω, let

$$U(\epsilon) = \{f \ : \ |f(x)| < \epsilon(x) \ \forall x\}.$$

Each such set is a convex, open neighborhood of $f = 0$, and they form a local basis for the topology. Cf. [Tr, p.18].

Equivalently, the topology on $C_c(\Omega)$ is defined by the uncountable family of norms

$$\|f\|_\epsilon = \sup_\Omega |f(x)|/\epsilon(x).$$

Proposition 4.9 *Every bounded set $B \subset C_c(\Omega)$ is contained in $C_c(K_n)$ for some n.*

Proof. Suppose B is not. Then there exist $x_n \to \infty$ and $f_n \in B$ such that $f_n(x_n) \neq 0$. Let $U = \{f \ : \ |f(x_n)| \leq |f_n(x_n)|/10^n$. Then U is open, but $f_n \notin nU$, so B is not bounded. ■

Corollary 4.10 *The space $C_c(\Omega)$ is complete in the inductive topology.*

Proof. Let f_n be a Cauchy sequence in $C_c(\Omega)$. By the result above, all f_n are supported in a single convex set K. Since they form a Cauchy sequence, they converge uniformly on K, so their limit is in $C_c(\Omega)$. ■

Direct limits. The space $C_c(\mathbb{R})$ is an example of a *direct limit* of locally convex topological vector spaces.

In general, if $X_1 \subset X_2 \subset \cdots$ is an ascending sequence of locally convex topological vector spaces, with continuous inclusions, then the *direct limit topology* on

$$X = \lim_{\longrightarrow} X_i = \bigcup X_i$$

define by taking as a local base at those *convex sets* U such that $U \cap X_i$ is a convex, balanced open set for each i.

A local base for this *inductive* or *direct limit* topology is given by open sets of the form $U = U_1 + U_2 + \cdots$, where $U_i \subset X_i$ is an open convex neighborhood of the origin.

The direct limit topology is characterized as the weakest topology such that a linear map to a locally convex topological vector space,

$$T : X \to Y,$$

is continuous iff $T|X_i$ is continuous for each i. It is also the *strongest* locally convex topology such that the map $X_i \to X$ are continuous for each i.

Duality. The *dual* X^* for a topological vector space is the space of all continuous linear functions $\phi : X \to \mathbb{R}$. Here is a useful property of such ϕ:

Proposition 4.11 *If $B \subset X$ is bounded, then $\phi(B) \subset \mathbb{R}$ is also bounded.*

Proof. Let $U = \phi^{-1}(-1, 1)$. Then we have $B \subset tU$ for some $t > 0$, so we have $\phi(B) \subset t(-1, 1) = (-t, t)$. ∎

There are at least two natural topologies on X^*: the *weak* (or weak*) topology, and the *strong* topology.

The strong topology is meant to generalize the norm topology on a Banach space. It is given by the seminorms

$$p_B(\phi) = \sup_B |\phi(x)|,$$

where B ranges over all *bounded* subsets of X. When X is a Banach space, it is enough to take B to be the unit ball.

The weak* topology, as usual, is the weakest topology such that $\phi \mapsto \phi(x)$ is continuous for all $x \in X$.

It is a general fact that for locally convex spaces, we have

$$(\varprojlim X_i)^* = \varinjlim X_i^* \quad \text{and} \quad (\varinjlim X_i)^* = \varprojlim X_i^*$$

as vector spaces.

Remark: The real vector space of countable dimension. Let $X = C_c(\mathbb{N})$. It can be thought of as the space of sequences (a_n) with $a_i = 0$ for all $i \gg 0$.

As a vector space, X has countable basis. It *cannot* arise as the *algebraic dual* of another real vector space, i.e. $X \neq \mathrm{Hom}(V, \mathbb{R})$. The dimension of such a dual space is either finite or uncountable. However, it does arise as the topological dual of $C(\mathbb{N})$, the space of all (continuous) functions on $\mathbb{N}$ with the topology of uniform convergence on finite (compact) sets.

Remark: The dimension of Banach spaces. It is known that any two infinite–dimensional Banach spaces with $|X_1| = |X_2|$ are isomorphic as abstract vector spaces. For example, any two separable Banach spaces are isomorphic as vector spaces.

To give an idea of the proof, first note that there exists a collection $\mathcal{S}$ of infinite subsets of $\mathbb{N}$ such that $|A \cap B| < \infty$ for any $A \neq B$ in $\mathcal{S}$, and $|\mathcal{S}| = |\mathbb{R}|$. (One construction: identity $\mathbb{N}$ with $\mathbb{Z}^2$ and for each $s \in \mathbb{R}$ let A_s be the set of integer lattice points in $\mathbb{R}^2$ within distance one of the line $y = sx$.)

Now consider the case $X = \ell^1(\mathbb{N})$. For each $A \in \mathcal{S}$ let x_A the sequence given by $a_n = 2^{-n}$ if $n \in A$ and $= 0$ if $n \notin A$. It is then easy to show these vectors $x_A \in \ell^1(\mathbb{N})$ are linearly independent, since for any $A_1, \ldots, A_k \in S$ we have $A_1 - (A_2 \cup \cdots \cup A_n) \neq \emptyset$. Thus $\dim_{\mathbb{R}} X \geq |\mathbb{R}|$ and the reverse inequality is obvious.

A similar argument shows $\dim_{\mathbb{R}} X \geq |\mathbb{R}|$ for any Banach space X of infinite dimension.

The bouquet of circles and the Hawaiian earring. One can compare the Hawaiian earring space $X = \bigcup S^1(1/n)$ and the CW complex $Y = \bigvee_1^\infty S^1$. The latter is given the topology where a set is open if its intersection with each finite union of circles is open. These spaces are *not* homeomorphic (although there is a natural continuous bijection $Y \to X$). Indeed, the space Y is not metrizable, and any compact subset of Y lies in a finite union of circles. This is useful in topology: $\pi_1(Y, *)$ is a direct limit of free groups, while $\pi_1(X, *)$ is an inverse limit of free groups. The first is countable, and the second is not.

5 Distributions

In this section we introduce distributions on $\mathbb{R}^n$ and their basic properties.

Motivation. Like category theory, the theory of distributions makes precise some mathematical ideas that were used informally for some time before their general theory was developed.

Distributions bring to the theory of functions the same advantages that negative numbers bring to arithmetic. Just as a calculation with a positive outcome may involve negative numbers at an intermediate stage, a theorem about C^k functions may involve $C^{-\ell}$ functions at an intermediate stage too.

In Bourbaki, a measure is *defined* as an element of the dual of $C(K)$. The main idea of a *distribution* is to permit more exotic generalized functions – beyond measures – by thinking of them as lying in the dual of the space of *smooth* functions rather than just continuous functions. These distributions are 'observable' only by integration against compactly supported smooth functions $\phi \in C_c^\infty(\mathbb{R}) = \mathcal{D}(\mathbb{R})$. So a continuous function $f(x)$ on $\mathbb{R}$ is identified with the *functional*

$$\Lambda(\phi) = \int_{\mathbb{R}} f\phi$$

for all test functions ϕ. Integrating by parts, it is then natural to identify the nth derivative $(D^n f)$ with

$$(D^n\Lambda)(\phi) = \int f^{(n)}\phi = (-1)^n \int f\phi^{(n)}.$$

The final expression makes sense for all smooth ϕ, and we now have a way to differentiate continuous functions as often as we like.

Test functions. Let us now formalize the space of test functions by describing it as a topological vector space.

Let $\Omega \subset \mathbb{R}^n$ be an open set. We let

$$\mathcal{D}(\Omega) = C_c^\infty(\Omega)$$

denote the vector space of all compactly support, infinitely–differentiable functions ϕ. For each compact set $K \subset \Omega$ we let

$$\mathcal{D}_K = C_c^\infty(K) \subset \mathcal{D}(\Omega).$$

and the union of these gives the full space of test functions.

We recall that the seminorms

$$\|\phi\|_N = \sup\{|D^\alpha\phi| \;:\; |\alpha| \le N\}$$

turn each subspace $\mathcal{D}_K$ into a Fréchet space in which bounded sets are compact, by the Arzela–Ascoli theorem.

The space $\mathcal{D}(\Omega)$ is given the *direct limit topology.* If we pick an exhaustion K_n of Ω, then we can write

$$\mathcal{D}(\Omega) = \lim_{n\to\infty} \lim_{\infty\leftarrow i} C_c^i(K_n).$$

As was the case for $C(\Omega)$, any bounded subset of $\mathcal{D}(\Omega)$ lies in $\mathcal{D}_K$ for some K. With this topology, $\mathcal{D}(\Omega)$ is a locally convex topological vector space.

Practical facts. Here are some of the main facts – easily verified – that one uses when working with $\mathcal{D}(\Omega)$.

1. Any bounded subset of $\mathcal{D}(\Omega)$ is contained in $\mathcal{D}_K$ for some K. There, it is contained in a bounded set of the form

$$B = \{\phi : \|\phi\|_{C^i} \leq M_i\}$$

for some sequence $M_i \to \infty$.

2. It follows that any closed, bounded set is compact.

3. Since Cauchy sequences are bounded, $\mathcal{D}(\Omega)$ is complete, as is each $\mathcal{D}_K$.

4. We have $\phi_n \to \phi$ if and only if all ϕ_n are supported in a single compact set K, and $D^\alpha \phi_n \to D^\alpha \phi$ uniformly, for all α.

5. The elements of a local base for $\mathcal{D}(\Omega)$ are given by a pair of positive continuous functions $A(x)$, $\epsilon(x)$ on Ω, via

$$U(A, \epsilon) = \{\phi : \forall x, |\alpha| \leq A(x) \implies |D^\alpha \phi(x)| \leq \epsilon(x)\}.$$

In other words, an open set controls the size of finitely many derivatives of ϕ on each compact subset K of Ω, and the required bounds can tend to zero and the number of derivatives can tend to infinity as K grows. See [Tr, p.18].

6. The map $D^\alpha : \mathcal{D}(\Omega) \to \mathcal{D}(\Omega)$ is continuous for each α.

7. Similarly, the map $\phi \mapsto f\phi$ is continuous for all $f \in C^\infty(\Omega)$. (Note that f need not have compact support.)

8. Putting these together, we find that linear differential operators given by finite sums of the form

$$L(\phi) = \sum f_\alpha(x) D^\alpha \phi$$

give continuous maps of $\mathcal{D}(\Omega)$ to itself.

Distributions. The space of *distribution* on Ω is defined by:

$$\mathcal{D}'(\Omega) = \mathcal{D}(\Omega)^*.$$

Alternatively, one writes:

$$C^{-\infty}(\Omega) = C_c^\infty(\Omega)^*.$$

(The $\mathcal{D}$ and $\mathcal{D}'$ notation goes back to Laurent Schwartz, who was awarded the Fields Medal for his work on distributions.) A distribution is typically denoted by Λ.

Examples.

1. The delta function δ_x at $x \in \Omega$ is the distribution given by $\Lambda_x(f) = f(x)$.

2. The function
$$\Lambda(f) = (D^\alpha)(x)$$
is also a distribution; it can be thought of, up to sign, as a derivative of δ_x.

3. In general, we define
$$(D^\alpha \Lambda)(\phi) = (-1)^{|\alpha|}\,\Lambda(D^\alpha \phi).$$

4. Integration with respect to Lebesgue measure gives a distribution
$$\Lambda(\phi) = \int_{\mathbb{R}^n} \phi(x)\,dx.$$

5. The continuous functions $f \in C(\Omega)$ can be identified with distributions via
$$\Lambda_f(\phi) = \int_{\mathbb{R}^n} f(x)\phi(x)\,dx.$$

6. Similarly, any locally finite signed measure μ on Ω gives a distribution. In particular, $L^1(\Omega)$ is contained in $\mathcal{D}'(\Omega)$.

Electrostatics and point charges. In $\mathbb{R}^n$ the operator $\Delta f = \sum D_i^2 f$ relates the electrical potential f into the underlying charge density, $\Delta f = \rho$. The idea of a point charge can be made rigorous via distributions.

For example, in $\mathbb{R}^2$ we have
$$\Delta(\log r) = 2\pi\delta_0.$$

The constant 2π is just the area of the unit circle; it can be justified by considering the flux of $\nabla \log r$ through S^1. Similarly, on $\mathbb{R}^n$ we have
$$\Delta(r^{2-n}) = -\operatorname{vol}(S^{n-1})\delta_0.$$

(In both case, ∇f is a unit vector field pointed outward on the unit sphere.) We will later discuss how this and other *fundamental solutions* to particular differential equations can be used to solve $Lu = f$.

The order of a distribution. Let us recall that

$$\mathcal{D}_K = \lim_{\longleftarrow} C^i_c(K),$$

and thus

$$\mathcal{D}^*_K = \lim_{\longrightarrow} C^i_c(K)^*.$$

In other words, for every compact set K, there exists an N such that

$$|\Lambda(\phi)| \leq \|\phi\|_{C^N}$$

whenever $\phi \in \mathcal{D}_K$. This N is called the *order* of Λ. In fact, for distribution Λ we can find a sequence of signed measures μ_α such that

$$\Lambda(\phi) = \sum_{|\alpha|} \int D^\alpha \phi \, \mu_\alpha$$

for every $\phi \in \mathcal{D}_K$. To see this, use the fact that Λ extends to a continuous linear functional on $C^N_c(K)$, and the latter maps, via $\phi \mapsto (D^\alpha \phi)$, to a closed subspace of a finite product of copies of $C(K)$.

If the same N works for all K, we say Λ has order N. Otherwise, Λ has infinite order.

Limits of distributions. As for any dual space, there are two natural topologies on the space of distributions: the strong and the weak* topologies, discussed in §4. Let us focus on the weak* topology for now. In this topology, $\Lambda_i \to \Lambda$ if and only if

$$\Lambda_i(\phi) \to \Lambda(\phi)$$

for all test functions ϕ. (Although the topology on $\mathcal{D}'(\Omega)$ is not metrizable, sequentially closed sets are closed, so we will concentrate on sequences.)

Examples.

1. Given a compactly support smooth function ψ on $\mathbb{R}^n$, with $\int \psi = 1$, let $\psi_r(x) = r^{-n}\psi(x/r)$. Then as $r \to \infty$, $\psi_r \to \delta_0$ as a distribution.

2. Note that for $n = 1$, $\psi'_r(x) \to D\delta_0$ as a distribution on $\mathbb{R}$. We also have $2n(\delta_{1/n} - \delta{-1/n}) \to \delta'_0$.

3. Let $f_n(x) = \sin(nx)$ on $\mathbb{R}$. Then as $n \to \infty$, $f_n \to 0$ as a distribution.
4. Let $f_n(x) = n^2 \sin(n\pi x)$ on the interval $[-1/n, 1/n]$ (and 0 elsewhere). Then for any test function ϕ, we have

$$\begin{aligned} \int f_n \phi &= \int f_n(\phi(0) + x\phi'(0) + O(x^2)) \sim \phi'(0) \int x f_n \\ &\sim \phi'(0) \frac{2}{\pi}. \end{aligned}$$

Thus f_n converges, as a distribution, to a multiple of $D\delta_0$.

Convergence in norm on compact sets. The weak* topology on distributions is actually stronger than it might appear at first sight. Here are two basic facts about the weak* topology on $\mathcal{D}'(\Omega)$.

1. If Λ_i is a sequence of distributions, and

$$\Lambda\phi = \lim \Lambda_i \phi$$

exists for all test functions ϕ, then Λ is a distribution.

2. If $\Lambda_i \to \Lambda$ in $\mathcal{D}'(\Omega)$, then for each compact set K there exists an integer k such that all these distributions extend to $C^k(K)$ and

$$\|\Lambda_i - \Lambda\|_{C^k(K)^*} \to 0$$

as $i \to \infty$. In other words, weak* convergence implies convergence in norm on each compact set.

Another way to say this is that the weak* and strong topologies agree on sequences.

Uniform boundedness. These assertions rely on a version of the uniform boundedness principle for Fréchet spaces, and on compactness of the natural map $C_c^{k+1}(K) \to C_c^k(K)$.

For the first assertion, let us begin by simply assuming that for each ϕ, $\sup |\Lambda_i(\phi)| < \infty$. Let F_M be the subset of $C_c^\infty(K)$ where all Λ_i are bounded by M. Then F_M is closed, and $\bigcup F_M = C_c^\infty(K)$. Since the latter is a Fréchet space, we can apply the Baire category theorem to conclude that F_M – and hence F_1 – contains a neighborhood U of zero. Referring to the seminorms defining $C_c^\infty(K)$, we see that U contains an open set defined by the condition

$$\|\phi\|_{C^k} < r$$

for some k and some $r > 0$. Thus every Λ_i satisfies

$$|\Lambda_i \phi| \le (1/r)\|\phi\|_{C^k}.$$

Now if $\Lambda_i\phi \to \Lambda\phi$ for each ϕ, the same bound holds for Λ, and hence Λ is a distribution. In fact, we have shown that all Λ_i lie in the same space $C^k_c(K)^*$ and they are uniformly bounded there.

For the second assertion, we recall an important fact about continuous functions on a compact set: if $f_n \to f$ pointwise, and the family of functions f is equicontinuous, then $f_n \to f$ uniformly. This is the basis of the Arzela–Ascoli theorem.

To complete the proof, observe that the unit ball B in C^{k+1} has compact closure in C^k. We have just seen that the Λ_i are uniformly bounded as operators on C^k, so they are equicontinuous when restricted to C^{k+1}. Then by compactness, they converge uniformly on B, which means they converge in the norm topology on $(C^{k+1})^*$.

Summing up:

1. Weak* convergence of Λ_i gives pointwise convergence;
2. The uniform boundedness principle implies equicontinuity; and then
3. Compactness gives convergence in norm.

Reference: [Tr, p.22].

Corollary 5.1 *The multiplication map*

$$C^\infty(\Omega) \times C^{-\infty}(\Omega) \to C^{-\infty}(\Omega)$$

is continuous for sequences: that is, if $f_i \to f$ and $\Lambda_i \to \Lambda$, then $f_i\Lambda_i \to f\Lambda$.

Proof. Suppose $\phi \in C^\infty_c(\Omega)$. Then $f_i\phi \to f\phi$ in every C^k; since the Λ_i are *uniformly* bounded, we also have

$$\Lambda_i(f_i\phi) \to \Lambda(f\phi).$$

But this says exactly that $f_i\Lambda_i \to f\Lambda$. ∎

Remark. The weak* and strong topologies on $\mathcal{D}'(\Omega)$ are definitely different. For example, every weak* neighborhood of the origin contains a finite codimension subspace, and this is not true in the strong topology.

A good picture to keep in mind is:

$$\mathcal{D}'(\Omega) = \lim_{\infty \leftarrow n} \lim_{i \to \infty} C^i_c(K_n)^*.$$

The sheaf of distributions; currents. Given an open set $U \subset V$, we have $C^\infty_c(U) \subset C^\infty_c(V)$ (extend by zero) and hence a restriction map $C^{-\infty}(V) \to C^{-\infty}(U)$.

Using these restriction maps, the distributions become a *sheaf.* (The sheaf axioms follow using a partition of unity.) We can then define the sheaf of distributions $\mathcal{D}'_M$ on a *smooth manifold* M^n using charts.

The cohomology of the sheaf of distributions is trivial. One can also define distributional *forms*; these *deRham currents* provide a beautiful way of incorporating submanifolds and other objects into the cohomology of a manifold.

Supports. With the notion of restriction at hand, we can define the support of a distribution. Namely $\operatorname{supp} \Lambda = F$ if F is the smallest closed set such that $\Lambda|(\Omega - F) = 0$.

If Λ has compact support, then Λ has finite order and

$$\Lambda(\phi) = \Lambda(f\phi)$$

for any smooth function f with $f = 1$ on a neighborhood of $\operatorname{supp} \Lambda$. Moreover, there exist k and M such that

$$|\Lambda(\phi)| \le M \|\phi\|_{C^k}$$

for all $\phi \in C^\infty_c(\Omega)$.

Skyscrapers.

Theorem 5.2 *If* $\operatorname{supp} \Lambda = p$*, then*

$$\Lambda = \sum_{|\alpha| \le N} c_\alpha D^\alpha \delta_p.$$

Proof. Assume $p = 0$ and let $\psi_r = \psi(x/r)$, where $\psi(x)$ is a bump function, with $\operatorname{supp} \psi_r \subset B(0, r)$ and $\psi_r = 1$ on a neighborhood of 0. Then $|D^\alpha \psi_r| = O(r^{-|\alpha|})$.

Suppose Λ has order N and ϕ vanishes to order N at 0, meaning $D^\alpha \phi = 0$ for $|\alpha| \le N$. Then for $x \in B(0,r)$ we have:

$$\phi(x) = O(r^{N+1}),$$

and more generally

$$D^\alpha \phi(x) = O(r^{N+1-|\alpha|}).$$

Thus $\psi_r \phi = O(r^{N+1})$ everywhere and

$$D^\alpha(\psi_r \phi) = O(r^{N+1-|\alpha|}).$$

Thus $\psi_r \phi \to 0$ in $C^N(\mathbb{R}^n)$. But then

$$\Lambda\phi = \Lambda(\psi_r \phi) \to 0$$

as $r \to 0$, and thus $\Lambda\phi = 0$. Thus Λ only depends on the N-jet of ϕ at 0. ∎

Theorem 5.3 *A positive distribution is a measure.*

Proof. The statement is a local one. Localize the distribution so it defines a functional on $C^\infty(K)$, and use positivity to extend it to $C(K)$. ∎

Theorem 5.4 *Any compactly supported distribution is a finite sum of derivatives of signed measures:*

$$\Lambda = \sum_{|\alpha| \le N} D^\alpha \mu_\alpha.$$

Proof. Suppose Λ has order k. Map $f \in C_c^\infty(K)$ into a product of copies of $C(K)$ by sending f to $(D^\alpha f \,:\, |\alpha| \le k)$. Then Λ is continuous in the *sup norm* on the image. By the Hahn-Banach theorem, Λ extends to a linear functional on the whole product, which in turn is given by a list of measures μ_α. This shows $\Lambda = \sum D^\alpha \mu_\alpha$ where the μ_α are measures. ∎

Theorem 5.5 *Any compactly supported distribution has the form*

$$\Lambda = \sum_{|\alpha| \leq N} D^\alpha f_\alpha,$$

with f_α continuous.

Proof. It remains only to show that the measures are all derivatives of functions. First consider the case of $\mathbb{R}$: then setting $F(x) = \mu[-\infty, x)$ we find $DF = \mu$, because

$$\int \phi(x)\, d\mu(x) = -\int \phi'(x) F(x)\, dx.$$

Now F is a bounded function, so doing it one more time we see μ is the second derivative of a (Lipschitz) continuous function.

Similarly on $\mathbb{R}^n$, if we set $F(x) = \mu\{(y : y_i < x_i)\}$ then we get

$$\int \phi(x)\, d\mu(x) = (-1)^n \int (D^1 \cdots D^n \phi)(x) F(x)\, dx.$$

Replacing μ with $F(x)\, dx$ and repeating, we have expressed μ as a higher–order derivative of a Lipschitz function. ■

Convolutions. We now turn to a fundamental technique in harmonic analysis, the use of convolutions. This operation natural links distributions and the Fourier transform.

The Banach algebra $L^1(\mathbb{R}^n)$. We define, for $f, g \in L^1(\mathbb{R}^n)$,

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

This is, by Fubini's theorem, another L^1 function, satisfying

$$\|f * g\|_1 \leq \|f\|_1 \cdot \|g\|_1.$$

Thus it makes $L^1(\mathbb{R}^n)$ into a commutative Banach algebra (without identity). In particular, we have

$$f * g = f * g \quad \text{and} \quad (f * g) * h = f * (g * h).$$

Group rings. To put this construction in context, G be a multiplicative group (not necessarily commutative). To make G into an *algebra* over $\mathbb{R}$ we consider the *group ring* $\mathbb{R}[G]$. Its elements are formal finite sums $f = \sum a_g \cdot g$; they can be thought of as maps $f : G \to \mathbb{R}$ with finite support. The *product* of two such elements is defined using the distributive law and the product in G:

$$\left(\sum a_g \cdot g\right)\left(\sum b_g \cdot g\right) = \sum_{g,h} (a_g b_h) \cdot gh = \sum_g \left(\sum_h a_h b_{h^{-1}g}\right) \cdot g.$$

We recognize the term in parentheses as the convolution of two functions on G with finite support.

Thus $(L^1(\mathbb{R}), *)$ generalizes the group ring to the continuous setting.

A principal motivation for the group ring is that linear representations of G over $\mathbb{R}$ correspond to modules over $\mathbb{R}[G]$. Similarly, a continuous representation of a Lie group (such as $\mathbb{R}^n$) on a Banach space gives rise to a module over the ring $L^1(G)$ with convolution.

Independent random variables. A second motivation for convolution comes from probability theory: namely if X and Y are independent random variables with distribution functions f and g (meaning $P(a < X < b) = \int_a^b f(x)\,dx$, and similarly for g), then the distribution function of $X + Y$ is $f * g$.

The central limit theorem is thus related to iterated convolution, $f * f * f * \cdots * f$.

Convolution and L^∞. We can also consider the convolution $(f * g)(y)$ whenever the defining integral makes sense for a.e. y. For example, this is the case when $f \in L^1$ and $g \in L^\infty$, and have:

Theorem 5.6 *If $f \in L^1$ and $g \in L^\infty$ then $(f * g)(y)$ is continuous, and*

$$\|f * g\|_\infty \le \|f\|_1 \|g\|_\infty.$$

Proof. The inequality above is immediate. Since $C_c(\mathbb{R}^n)$ is dense in L^1, $f * g$ can be uniformly approximated by $h * g$ with h compactly supported and continuous. But then $h * g$ is continuous, and thus $f * g$ is a uniform limit of continuous functions, hence continuous. ∎

Note that in L^1, $f * g$ need *not* be continuous; for example, if f is not in L^2, and $g(x) = f(-x)$, then $(f * g)(0) = \int |f|^2 = \infty$.

***A* − *A*.** We can use convolutions to prove the classical result:

Proposition 5.7 *If $A \subset [0,1]$ has positive measure, then $A - A$ contains an open interval.*

Proof. Let $f(x) = \chi_A(x)$ and let $g(x) = f(-x)$. Then $f, g \in L^1 \cap L^\infty$ so $f * g(y)$ is continuous. Moreover, $f * g(0) = m(A) > 0$. Thus $(f * g)(y) > 0$ on some interval $(-\alpha, \alpha)$. But $(f * g)(y) > 0$ implies $y \in A - A$, by the definition of convolution. ∎

Approximate identities. The algebra $L^1(\mathbb{R}^n)$ has no identity element. (Morally, the identity element should be the delta function at 0.) However $L^1(\mathbb{R}^n)$ almost contains an identity element. Let $\psi \in C_c^\infty(\mathbb{R}^n)$ satisfy $\int \psi = 1$, and as usual set $\psi_r = r^{-n}\psi(x/r)$. We then have

$$\|\psi_r * g - g\|_1 \to 0$$

as $r \to 0$. This is easily verified first for the case where $c \in C_c(\mathbb{R}^n)$, and then observing that such functions are dense in $L^1(\mathbb{R}^n)$.

Example. We can usually regard $f * g$ as a limit of convex combinations of translates of g. We can even think of

$$F : x \mapsto f(x)g(y - x)$$

as a map from $\mathbb{R}$ to a Banach space B containing g, and the consider the convolution as the integral of a B–valued function with $\int \|F(x)\|_B \, dx < \infty$.

Theorem 5.8 *Suppose $g \in B$, a Banach space with translations acting continuously and isometrically. Then $f * g \in B$ as well, for all $f \in L^1(\mathbb{R}^n)$; and*

$$\|f * g\|_B \leq \|f\|_B \|g\|_1.$$

This means $f * g$ tends to inherit the good properties of both f and g. Under similar hypotheses, we have

$$\|\psi_r * g - g\|_B \to 0$$

as $r \to 0$.

Mollification. If $\phi \in C_c^\infty(\mathbb{R}^n)$, then $f * \phi \in C^\infty(\mathbb{R}^n)$ and all its derivatives are uniformly bounded. Moreover we have:

$$D^\alpha(f * \phi) = f * D^\alpha \phi.$$

This combined with the existence of approximate identities shows 'constructively' that smooth functions are dense in L^1 and many other Banach spaces B.

Warning on L^∞. An example where these ideas do not quite work comes about when $B = L^\infty(\mathbb{R})$. Then when g is discontinuous, it cannot be the uniform limit of its continuous smoothings $\psi_r * g$. The problem here is that translation by $x \in \mathbb{R}^n$ does not act continuously on $L^\infty(\mathbb{R})$; that is, we do not have $\|T_x(g) - g\|_\infty \to 0$ as $x \to 0$, as can be seen by taking g to be a step function.

Special values and supports. In general we let $\check{f}(x) = f(-x)$. We note that the value of a convolution $f * g$ at zero (or any given point) gives the usual pairing between f and g, except we must replace f with $\check{f}$ or g with $\check{g}$:

$$(f * g)(0) = \int f(x) g(-x)\, dx = \langle f, \check{g} \rangle = \langle \check{f}, g \rangle.$$

Also as can be expected from the example of the group algebra, the support of $f * g$ is controlled by that of f and g:

$$\operatorname{supp}(f * g) \subset \overline{\operatorname{supp}(f) + \operatorname{supp}(g)}.$$

Convolutions, test functions and distributions. By the remarks above, if ϕ and ψ are test functions, then so is $\phi * \psi$, and indeed convolution gives a continuous linear map

$$* : C_c(\mathbb{R}^n) \times C_c(\mathbb{R}^n) \to C_c(\mathbb{R}^n).$$

Let us now define the convolution of a distribution $\Lambda \in C^{-\infty}(\mathbb{R}^n)$ with a test function ϕ. Reasoning formally, it is given by

$$(\Lambda * \phi)(y) = \Lambda_x(\phi(y - x)).$$

It follows from basic properties of distributions that $\Lambda * \phi \in C^\infty(\mathbb{R}^n)$, and that

$$D^\alpha(\Lambda * \phi) = \Lambda * D^\alpha \phi.$$

The delta function now plays the role of an actual identity for convolution: we have

$$\delta_0 * \phi = \phi$$

for all test functions ϕ (and, when suitably defined, for all L^1 functions). Similarly we have

$$(\delta_p * \phi)(y) = \phi(y - p),$$

and

$$(D^\alpha \delta_0) * \phi = D^\alpha \phi.$$

Theorem 5.9 *For any two test functions ϕ and ψ, we have*

$$(\Lambda * \phi) * \psi = \Lambda * (\phi * \psi).$$

Proof. It suffices to evaluate both sides at $y = 0$. Approximating the integral by a finite sum, and using continuity properties of Λ, we find:

$$\begin{aligned}(\Lambda * (\phi * \psi))(0) &= \Lambda_x((\phi * \psi)(-x)) = \Lambda_x(\int \psi(y)\phi(-x-y)\,dy) \\ &= \Lambda_x(\int \psi(-y)\phi(y-x)\,dy) = \int \psi(-y)\Lambda_x\phi(y-x)\,dy \\ &= \int \psi(-y)(\Lambda * \phi)(y)\,dy = ((\Lambda * \phi) * \psi)(0).\end{aligned}$$

■

Corollary 5.10 *The smooth functions $C^\infty(\mathbb{R}^n)$ are dense in the space of distributions $C^{-\infty}(\mathbb{R}^n)$.*

Proof. Let $\psi_r \to \delta_0$ be an approximate identity. We claim that

$$\Lambda * \psi_r \to \Lambda$$

as a distribution. Since the convolution makes Λ into a smooth function, this will complete the proof. To verify convergence, let ϕ be a test function. Then

$$(\Lambda * \psi_r)(\phi) = ((\Lambda * \psi_r) * \check{\phi})(0) = (\Lambda * (\psi_r * \check{\phi}))(0) = \Lambda(\check{\psi}_r * \phi) \to \Lambda(\phi),$$

as required. ■

Remark. This argument also shows that compactly supported distributions are approximated by test functions.

One can also argue directly that compactly supported distributions are dense, and then approximate $D^\alpha f$ where f is continuous by $D^\alpha \phi$. Using the behavior of supports, one can similarly show:

Corollary 5.11 *The space of test functions $\mathcal{D}(\Omega)$ is dense in the space of distributions $\mathcal{D}'(\Omega)$.*

Translation invariant operators. We say $L(f)$ is translation invariant if it commutes with the operators T_p for all p, where $T_p(f)(x) = f(x+p)$.

It is easy to see formally that

$$T_p(f * g) = f * (T_p g).$$

(In fact, differentiation with respect to p gives $D(f * g) = f * (Dg)$.)

Theorem 5.12 *Any continuous translation invariant operator*

$$L : C_c^\infty(\mathbb{R}^n) \to C_c^\infty(\mathbb{R}^n)$$

is given by $Lf = \Lambda * f$, for some compactly supported distribution Λ.

Proof. Set $\Lambda(f) = (L\check{f})(0)$. ∎

Convolution of distributions. Using the last result, when Λ_1 and Λ_2 are compactly supported, we can define $\Lambda_1 * \Lambda_2$ to be the unique distribution such that

$$(\Lambda_1 * \Lambda_2) * \phi = \Lambda_1 * (\Lambda_2 * \phi)$$

for any test function ϕ. Note that $\Lambda_2 * \phi$ is a compactly supported smooth function, so this convolution makes sense.

For example, $\delta_p * \delta_q = \delta_{p+q}$. One can then show that convolution of distributions of this type is commutative and associative, etc.

One can also form $\Lambda_1 * \Lambda_2$ when only *one* of these is compactly supported, but more care should be taken. For general distributions the convolution may not be defined, or it may be defined and commutative can fail, etc.

6 The Fourier transform

The Fourier transform is the case $G = \mathbb{R}^n$ of Pontryagin duality between $L^2(G)$ and $L^2(\widehat{G})$, where $\widehat{G}$ is the dual of a locally compact commutative group G. Fourier series arise in the case $G = S^1$, $\widehat{G} = \mathbb{Z}$. Many beautiful topics are related to the Fourier transform, such as the central limit theorem, the uncertainty principle, Poisson summation, theta functions and automorphic forms. It brings to real analysis the kind of magic that Cauchy's integral formula brings to complex analysis.

An essential point will be that only functions with some degree of regularity have Fourier transforms. For example, there are continuous functions, such as $e^{|x|}$, for which there is no reasonable definition of the Fourier transform. However, we will find there is a broad class of *tempered distributions* that do have Fourier transforms, and which include all compactly supported distributions. These can be further filtered by the *Sobolev spaces* H^s, which will play a leading role in the study of partial differential equations to come.

Duality. In general one defines the dual $\widehat{G}$ when G is abelian by the group of characters

$$\widehat{G} = \mathrm{Hom}(G, S^1),$$

where $S^1 \subset \mathbb{C}^*$. The characters $\chi : G \to S^1$ are required to be continuous.

This group describes the 1–dimensional unitary representations of G. For $G = \mathbb{R}^n$ we also have $\widehat{G} = \mathbb{R}^n$, and $\chi_t \in \widehat{G}$ is given by

$$\chi_t(x) = \exp(ixt).$$

Here $xt = \sum x_i t_i$. Thus from an intrinsic or basis–free perspective, $\widehat{G}$ is the dual of the vector space $G = \mathbb{R}^n$. One can regard $G \times \widehat{G}$ as the cotangent bundle of $\mathbb{R}^n$.

Complex inner product. Since characters are complex valued, we must work with complex valued functions throughout. The inner product on $L^2(\mathbb{R}^n, dm)$ is given by the Hermitian form:

$$\langle f, g\rangle = \int f(x)\overline{g}(x)\, dm.$$

where $dm = dx/(2\pi)^{n/2}$. This normalization of the measure makes the inversion formula symmetric while also keeping the behavior of the Fourier transform with respect to differentiation very simple. Throughout the following discussion, we use this measure to define the L^p norms and convolution on $\mathbb{R}^n$.

The Fourier transform on $L^1(\mathbb{R}^n)$. The Fourier transform of f, like convolution itself, is initially defined for $f \in L^1(\mathbb{R}^n)$. It is the function $\widehat{f}(t)$ given by

$$\widehat{f}(t) = \int_{\mathbb{R}^n} f(x) \exp(-ixt)\, dm.$$

Theorem 6.1 *The Fourier transform defines a continuous linear map*

$$\mathcal{F} : L^1(\mathbb{R}^n) \to C_0(\mathbb{R}^n),$$

satisfying $\|\widehat{f}\|_\infty \le \|\widehat{f}\|_1$.

Proof. The norm inequality is immediate, and it is easy to see that $\widehat{f}$ is continuous when f is a compactly supported continuous function; since such f are dense in L^1, we have $\widehat{f} \in C(\mathbb{R}^n)$. To see that $\widehat{f}(t)$ tends to zero as $|t| \to \infty$, approximate f by a step function (a sum of indicator functions of rectangles). ∎

Characters. Like Fourier series on the circle, the idea of the Fourier transform is that $\widehat{f}(t)$ detects the different *frequency components* of the function $f(x)$, i.e. it measure the degree to which $f(x)$ is synchronized with the periodic functions $\chi_t(x) = \exp(itx)$.

One motivation is to turn the *algebra* $A = (L^1(\mathbb{R}^n), *)$ into an algebra of *continuous functions* with pointwise multiplication. In general, the point evaluations are given *multiplicative linear functionals* on A, i.e. elements of A^* satisfying

$$\phi(f * g) = \phi(f)\phi(g).$$

Let us note that any character

$$\chi : \mathbb{R}^n \to S^1 \subset \mathbb{C}^*$$

defines such a multiplicative linear functional on A, by

$$\phi(f) = \langle f(x), \chi(x) \rangle = \int_{\mathbb{R}^n} f(x)\chi(x)\, dx.$$

Indeed, setting $z = y - x$, we have:

$$\phi(f*g) = \int f(x)g(y-x)\chi(y)\,dx\,dy = \int f(x)g(z)\chi(x+z)\,dx\,dz = \phi(f)\phi(g).$$

The same holds true if we replace dx by dm (since this changes the definition of convolution). Since the functions $\chi_t(x)$ are characters, we have shown

Theorem 6.2 *The Fourier transform gives an algebra homomorphism from* $(L^1(\mathbb{R}^n), *)$ *to* $(C_0(\mathbb{R}^n), \cdot)$.

Frequency versus position. A second motivation for the Fourier transform is to decompose $f \in L^2(\mathbb{R})$ (or $\mathbb{R}^n$) into eigenfunctions for differentiation.

The (unbounded) operator $T = (1/i)D$ is self–adjoint, and $T(\chi_t) = t\chi_t$. It should be compared to the operator $S(f) = xf(x)$, which is also self–adjoint, and which satisfies $S(\delta_p) = p\delta_p$. In neither case are the eigenvectors elements of $L^2(\mathbb{R}^2)$.

Rather, a general element of $L^2(\mathbb{R})$ should be thought of as a continuous linear combination of these eigenfunctions. For the operator S, this means $f = \delta_0 * f$. For the operator T, this means we will have an inversion formula:

$$f(x) = \int \widehat{f}(t)\exp(itx)\,dm(t).$$

This formula presents f as a continuous linear combination of functions with constant frequency.

Overview. Here is a plan for our investigation of the Fourier transform.

1. We introduce the space of smooth, rapidly decreasing *Schwartz functions* $\mathcal{S}(\mathbb{R}^n)$, and showing the Fourier transform sends this space into itself.

2. We show the Fourier transform reverses local and global properties, and behaves well with respect to differentiation, translation, convolution, and multiplication by polynomials.

3. We study the Fourier transforms of Gaussian distributions, and show they satisfy the desired inversion formula.

4. We prove the inversion formula $f(x) = \int \widehat{f}(t)\exp(ixt)\,dm(t)$ for Schwartz functions.

5. We show $\langle f, g\rangle = \langle \widehat{f}, \widehat{g}\rangle$ and thus the Fourier transform extends to an isometry $\mathcal{F} : L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$. This is the most important and basic result in the theory.

6. We give another proof that $\mathcal{F}$ is an isometry by relating the Fourier transform to Fourier series.

Schwartz functions. We now define a class of smooth functions containing $C_c^\infty(\mathbb{R}^n)$, contained in $L^1(\mathbb{R}^n)$, and closed under $\mathcal{F}$.

Note that for a *compactly* support function, $\widehat{f}(t)$ makes sense for *complex* values of t, and that $\widehat{f}(t)$ is real analytic. Thus we cannot hope for $\widehat{f}$ to be compactly supported.

Anticipating the behavior of $\widehat{f}$ with respect to differentiation and multiplication, we define a class of functions $\mathcal{S}(\mathbb{R}^n) \subset L^1(\mathbb{R}^n)$ that is as large as possible, while being closed under both differentiation and multiplication by polynomials.

Let us say a bounded function $f(x)$ is *rapidly decreasing* if as $|x| \to \infty$, we have

$$|f(x)| = O(|x|^{-N})$$

for all $N > 0$. That is, f decays faster than any polynomial.

We say $f \in C^\infty(\mathbb{R}^n)$ is a *Schwartz function* if $D^\alpha f$ is rapidly decreasing for all α. The space of all such functions will be denoted by $\mathcal{S}(\mathbb{R}^n)$. It is a Fréchet space with respect to the seminorms:

$$p_N(f) = \sup_{\mathbb{R}^n} \sup_{|\alpha| \le N} (1 + |x|^2)^{N/2} |D^\alpha f|.$$

Here are some of its basic properties.

1. If f is a Schwartz function, so is $P(x)f(x)$ and $P(D)f(x)$, for any polynomial P.

2. We also have $f \in L^1(\mathbb{R}^n)$; indeed any power of f is also integrable.

3. In particular, the Fourier transform $\mathcal{F}(f) = \widehat{f}$ is defined for all Schwartz functions.

Basic properties. Here are some basic properties of the Fourier transform on $\mathcal{S}(\mathbb{R}^n)$. The first few also hold for all $f \in L^1(\mathbb{R}^n)$.

1. $\widehat{f}(0) = \int f\,dm$. (Note the reversal of pointwise and global properties).

2. $\widehat{f * g}(t) = \widehat{f}(t)\widehat{g}(t)$.

3. $\widehat{f(ax)} = a^{-n}\widehat{f}(t/a)$. (Function homogeneous of degrees 0 and $-n$ are interchanged.)

4. $f\widehat{(x+a)}(t) = e^{iat}\widehat{f}(t)$.

5. $e^{\widehat{iax}f(x)}(t) = \widehat{f}(t-a)$.

6. $\widehat{f'(x)} = (it)\widehat{f}(t)$. (IBP. Infinitesimal form of translation.)

7. $\widehat{P(D)f} = P(it)\widehat{f}(t)$.

8. $\widehat{xf(x)} = i\widehat{f}'(t)$. (Infinitesimal form of multiplication by e^{iax}.)

9. $\widehat{P(x)f} = P(iD)\widehat{f}(t)$.

10. $f\widehat{(A^{-1}x)} = \det(A)\widehat{f}(A^*t)$. (Transform space is the cotangent bundle.)

N.B. Here we use the traditional D^α, not Rudin's special D_α. All these properties can be proved by change of variable, integration by parts and interchange of differentiation and integration. These operations are justified by the rapid because of f.

Since the Schwartz functions lie in L^1, their Fourier transforms are in $C_0(\mathbb{R}^n)$; in particular, they are bounded. When combined with the properties above, this shows:

Theorem 6.3 *The Fourier transform defines a continuous map $\mathcal{F} : \mathcal{S}(\mathbb{R}^n) \to \mathcal{S}(\mathbb{R}^n)$.*

As an example, note that if $P(t) = 1 + t^2$ then $P(-iD) = 1 - \Delta$. Thus to show $\sup(1 + t^2)|\widehat{f}(t)|$ is bounded, it suffices to show that $(1 - \Delta)f$ is in L^1. This follows by rapid decay of the second derivatives of f.

Borderline integrability. It is an essential fact that

$$\int_{\mathbb{R}^N} = \frac{1}{(1+|x|^2)^{s/2}} < \infty \iff s > n.$$

Thus $p_N(\widehat{f})$ is controlled by $p_{N+n/2+1}(f)$, since we need the extra polynomial decay to bound the L^1 norm.

The inversion formula. Our next goal is to show:

Theorem 6.4 *For any Schwartz function $f(x)$, we have*

$$f(x) = \int \widehat{f}(t) \exp(ixt)\, dm(t).$$

Since we have changed $\exp(-ixt)$ to $\exp(ixt)$, this formula is equivalent to:

$$\mathcal{F}^2(f) = f(-x),$$

which gives:

Corollary 6.5 *The map $\mathcal{F} : \mathcal{S}(\mathbb{R}^n) \to \mathcal{S}(\mathbb{R}^n)$ is a bijection, with a continuous inverse.*

The normal distribution. To prove the inversion formula, we will first prove it holds for a particular (Gaussian) function, and then observe that the subspace where it holds is closed under translation, scaling and limits, so it is the whole space.

Can we find a function in $\mathcal{S}(\mathbb{R})$ which is its own Fourier transform? The answer is yes. To do so, first observe that the Fourier transform interchanges the maps $Df = f'(x)$ and $Tf = xf(x)$. So if

$$\frac{df}{dx} + xf = 0,$$

then $it\widehat{f}(t) + i\widehat{f}'(t) = 0$. Since f and $\widehat{f}$ satisfy the same first order differential equation, they should at least be proportional. (If they don't come out equal, we can always renormalize volume measure so they do.)

We are thus lead to consider the Gaussian distribution $f(x) = \exp(-x^2/2)$ on $\mathbb{R}$ (and its generalization to $\mathbb{R}^n$). Since f is in Schwartz class, so is $\widehat{f}$, and thus we indeed have $f = \alpha\widehat{f}$ for some α by the reasoning above.

To check that $\widehat{f} = f$, we need only compare their values at zero. Of course $f(0) = 1$. To compute $\widehat{f}(0))$, observe that on $\mathbb{R}$ we have the usual calculation

$$\left(\int \exp(-x^2/2)\, dx\right)^2 = \int_0^\infty \exp(-r^2/2) 2\pi r\, dr = 2\pi,$$

and thus $\int f(x)\,dx = \sqrt{2\pi}$; taking into account the normalization of measure we get $\widehat{f}(0) = 1$. Thus $f = \widehat{f}$.

On $\mathbb{R}^n$ we use the fact $x^2 = \sum x_i^2$ plus Fubini's theorem to deduce that

$$\int_{\mathbb{R}^n} \exp(-x^2/2)\,dx = (2\pi)^{n/2},$$

and again the normalizing factor completes the proof.

The L^2 norm of a Gaussian. For applications to quantum mechanics, it is usual to normalize functions $f \in L^2(\mathbb{R}^n)$ so $\|f\|_2 = 1$. To this end we note that

$$\|f\|_2 = 1 \quad \text{if } f(x) = 2^{1/4}\exp(-x^2/2),$$

where we use the measure dm on $\mathbb{R}$ to compute the L^2 norm.

Gaussian and the central limit theorem. The central limit theorem describes the behavior of $S_n = (1/\sqrt{n})\sum X_i$, where X_i are independent random variables with $E(X_i) = 0$ and $E(X_i^2) = \sigma^2$ finite. The statement is that the limiting distribution of S_n is given by a Gaussian with variance σ^2.

When the distribution of X_i is governed by a function $f(x)\,dx$, this result can be approached using the fact that the distribution function of a sum of independent random variables is the convolution of their distributions; and that if X is distributed according to f, then aX is distributed according to $af(x/a)$, a function whose Fourier transform is $\widehat{f}(ax)$. Thus we have

$$\hat{S_n} = \widehat{f}(x/\sqrt{n})^n.$$

Since $E(X_i) = \int x f(x)\,dx = C\widehat{f}'(0) = 0$, we have $\widehat{f}(x) = 1 - bx^2 + O(x^3)$, which gives

$$\hat{(}F_n) = (1 - bx^2/n)^n \to \exp(-bx^2),$$

where F_n is the distribution function of S_n. Since the Fourier transform of a Gaussian is a Gaussian, this gives a formal proof that F_n converges to a Gaussian as well.

Proof of the inversion formula . Since $f(x) = f(-x)$ for a Gaussian, we have just verified the inversion formula for this function. We now extend the proof to all Schwartz functions.

Proof 1. By the functorial features of the Fourier transform, $\mathcal{F}^2$ behaves as indicated on Gaussians, their rescalings and their translates. (Note: if $\widehat{f} = f$, then

$$\mathcal{F}^2(f(x+a)) = \mathcal{F}(e^{iax}f(x)) = f(x-a).$$

By taking Gaussians ψ_n with standard deviation tending to zero, we have $f = \lim f * \psi_n$ in $\mathcal{S}$ and thus every $f \in \mathcal{S}$ is in the closed linear span of the Gaussians. By continuity, the formula $\mathcal{F}^2(f) = f(-x)$ holds throughout $\mathcal{S}(\mathbb{R}^n)$. ∎

Proof 2. We will use the L^2-structure on $\mathcal{S}$. First note that:

$$(f, \mathcal{F}(g)) = \int f(x)g(y)\exp(-ixy)\, dm(x)\, dm(y) = (\mathcal{F}(f), g);$$

that is, $\mathcal{F}$ is symmetric (note that we do not take complex conjugation). Also $\mathcal{F}$ conjugates translation by a to multiplication by e^{ia}, and multiplication by e^{ia} to translation by $-a$, so $\mathcal{F}^2$ intertwines translation, and we need only show that the final equality holds below:

$$(\mathcal{F}^2 f)(0) = \int \widehat{f}(x)\, dm(x) = (\mathcal{F}(f), 1) = f(0).$$

To this end, consider a sequence ψ_n of Gaussians normalized with $\int \psi_n\, dm = 1$ and with standard deviation tending to zero; then $\psi_n \to \delta$ as distributions, and in fact $\psi_n \to \delta$ in $\mathcal{S}^*$.

Then $\phi_n = \mathcal{F}(\psi_n)$ has $\phi_n(0) = 1$ and standard deviation tending to infinity; thus $\phi_n \to 1$. Therefore:

$$(\mathcal{F}(f), 1) = \lim(\mathcal{F}(f), \phi_n) = \lim(f, \mathcal{F}(\phi_n)) = (f, \delta) = f(0).$$

∎

The Plancheral formula. We next study the interaction between the Fourier transform and inner products. Let f and g be Schwartz functions. We then have, by the inversion formula:

$$\langle f, g\rangle = \int \widehat{f}(t)\exp(ixt)\overline{g}(x)\, dm(x)\, dm(t) = \langle \widehat{f}, \widehat{g}\rangle.$$

In particular, this shows that

$$\|\widehat{f}\|_2 = \|f\|_2.$$

Since Schwartz functions are dense, we then have the following result of signal importance:

Theorem 6.6 *The Fourier transform extends to an isometry $\mathcal{F} : L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$, satisfying $\mathcal{F}^2(f) = f(-x)$.*

The real line as a limit of circles. Here is a second proof, that also explain where the π in the normalizing factor comes from, without the use of Gaussians.

Suppose f is smooth on $\mathbb{R}$ and, for convenience, that $\operatorname{supp} f \subset [-M/2, M/2]$. Then we can consider f as a function on $\mathbb{R}$ and also on $S^1 = \mathbb{R}/M\mathbb{Z}$. In $L^1(S^1, dx)$, the functions

$$e_n(x) = M^{-1/2} \exp(2\pi i n x/M).$$

form an orthonormal basis. Write $f = \sum a_n e_n$, and define (ignore 2π's)

$$\widehat{f}(t) = \int f(x) \exp(-ixt)\, dx.$$

Let $h = 2\pi/M$; we then have

$$a_n = \langle f_n, e_n \rangle = M^{-1/2} \widehat{f}(nh) = (2\pi)^{-1/2} \sqrt{h} \widehat{f}(nh).$$

Recalling that $\widehat{f}(t)$ is continuous, we find that

$$\|f\|_2^2 = \sum_{-\infty}^{\infty} |a_n|^2 = (2\pi)^{-1} \sum h \cdot |\widehat{f}(nh)|^2 \to (2\pi)^{-1} \int |\widehat{f}(t)|^2\, dt$$

as $M \to \infty$. Changing dx and dt to $dm(x)$ and $dm(t)$ gives the Plancheral theorem. ∎

Additional properties of the Fourier transform. Summing up the results above, we have:

1. $\mathcal{F}^2(f) = f(-x)$.
2. $f(x) = \int e^{ixt} \widehat{f}(t)\, dm(t)$.
3. $\|\widehat{f}\|_2 = \|f\|_2$.
4. $\int \widehat{f}(x) g(x)\, dm(x) = \int f(x) \widehat{g}(x)\, dm(x)$.
5. For $g(x) = e^{-x^2/2}$, we have $\widehat{g(x)} = g(t)$ and $\int g(x)\, dm(x) = 1$.

6. If $f(x)$ is real, then $\widehat{f}(-t) = \overline{\widehat{f}(t)}$.

Poisson summation. Some remarkable identities emerge from the following relation between f and $\widehat{f}$.

Theorem 6.7 *For $f \in \mathcal{S}(\mathbb{R}^n)$ we have*

$$\sum_{\mathbb{Z}} f(n) = \sqrt{2\pi} \sum_{\mathbb{Z}} \widehat{f}(2\pi n). \tag{6.1}$$

Remark. A more common and equivalent formulation is

$$\sum f(n) = \sum \widehat{f}(n)$$

where we define

$$\widehat{f}(t) = \int f(x) \exp(-2\pi i t x)\, dx.$$

Proof. Let us prove the second version. We have a natural map $\pi : \mathbb{R} \to S^1 = \mathbb{R}/\mathbb{Z}$, which formally satisfies

$$\langle \pi^* f, g \rangle = \langle f, \pi_* g \rangle.$$

Let $F = \pi_*(f)$ and let $e_n(x) = \exp(2\pi i n x)$ on S^1. Then e_n is an orthonormal basis for $L^2(S^1)$, so we have

$$F(x) = \sum a_n e_n(x)$$

where

$$a_n = \langle F, e_n \rangle = \langle \pi_* f, e_n \rangle = \langle f, e_n \rangle = \widehat{f}(n).$$

Evaluating F at $x = 0$, we find

$$\sum f(n) = F(0) = \sum a_n e_n(0) = \sum \widehat{f}(n).$$

∎

Note: Poisson summation holds under weaker hypotheses as well.

Jacobi's theta function. Poisson summation provides an entry to the interplay between Fourier series and automorphic forms. A first step in this direction is provided by:

Theorem 6.8 *For $y > 0$,*

$$\theta(y) = \sum_{\mathbb{Z}} \exp(-\pi n^2 y)$$

satisfies the remarkable identity:

$$\theta(1/y) = \sqrt{y}\,\theta(y).$$

Let us first discuss the assertion heuristically. Letting $r = \exp(-\pi y) < 1$, we are studying a sort of super geometric series, $\sum r^{n^2}$. For $y \gg 1$, r is close to zero and the constant term dominates: we have $\theta(y) = 1 + O(r)$. For y near 0, r is close to one and the terms appearing in the sum are almost exactly one until n is large enough that $n^2 y$ is comparable to 1. Thus we have $\theta(y) \approx \sqrt{1/y}$ when y is small, and this is consistent with the identity above.

Proof of Theorem 6.8. Setting

$$f(x) = \exp(-\pi x^2 y) = \exp(-(x\sqrt{2\pi y})^2/2),$$

we have

$$\widehat{f}(t) = \frac{1}{\sqrt{2\pi y}} \exp\left(-\frac{1}{2}\left(\frac{t}{\sqrt{2\pi y}}\right)^2\right).$$

Then the Jacobi formula follows from Poisson summation, since

$$\theta(y) = \sum f(n) = \sqrt{2\pi} \sum \widehat{f}(2\pi n) = y^{-1/2} \sum \exp(-\pi n^2/y).$$

∎

From θ above we obtain the *Jacobi theta function*

$$\vartheta(z) = \sum_{\mathbb{Z}} \exp(\pi i n^2 z),$$

satisfying $\vartheta(iy) = \theta(y)$. Clearly $\vartheta(z)$ is invariant under $z \mapsto z + 2$; it also transforms reasonably under $z \mapsto -1/z$, namely:

$$\vartheta(-1/z) = (-iz)^{1/2}\vartheta(z).$$

(Here the square-root is chosen so that if $z = iy$, then $(-iz)^{1/2} = \sqrt{y} > 0$.) Thus $\vartheta(z)\, dz^{1/4}$ is an *automorphic form* for a congruence subgroup of $SL_2\mathbb{Z}$.

Letting $q = \exp(2\pi i z)$, we can also write

$$\vartheta(z) = \sum q^{n^2/2} = \sum_{\Lambda} q^{\langle\lambda,\lambda\rangle/2},$$

where Λ is the lattice of integers $\mathbb{Z}$. Notice then that

$$\vartheta(z)^k = \sum a_m(k) q^{m/2}$$

where $a_m(k)$ is the number of ways to express m as a sum of squares of k integers. Equivalently $a_m(k)$ counts the number of integral points on a sphere in $\mathbb{R}^k$ of radius $\sqrt{m}$ about the origin.

Compare [Ser, Chapter VII.6]. Note that the above is consistent with the factor $(iz)^{n/2}$ that appears on p.109, since Serre only treats the case where $8|n$.

Quantum mechanics. Let $f \in H = L^2(\mathbb{R})$ be a *state* in quantum mechanics, i.e. a vector in Hilbert space with norm one. Real-valued observables correspond to *self-adjoint operators* $A : H \to H$; the expected value of A is $\langle Af, f\rangle$.

Two of the most important observables are:

$$\begin{aligned} \text{position:} &\quad \Longleftrightarrow \quad Q(f) = xf(x), \quad \text{and} \\ \text{momentum:} &\quad \Longleftrightarrow \quad P(f) = -i\hbar\frac{df}{dx}. \end{aligned}$$

Here $\hbar = h/2\pi$ where $h = 6.626... \times 10^{-34} m^2 kg/s$ is Planck's constant, defined so that the energy of a photon with (wave) frequency ν is $E = h\nu$. If we work in units where $\hbar = 1$, then we have $\widehat{P(f)} = t\widehat{f}(t)$, so Q looks just like P in Fourier transform coordinates. In particular, $\langle Qf, f\rangle$ gives the expect value of the 'frequency' or momentum of f.

More intrinsically, the two isomorphisms of H with $L^2(\mathbb{R})$, related by $\mathcal{F}$, give the spectral decomposition of P and Q respective. A state with a precise position, for example, would be a δ-function concentrated at p.

Note that for $D = df/dx$ we have $D(xf(x)) = f(x) + xD(f)$, and thus

$$[D, x] = Dx - xD = I.$$

This is important because it shows D and x cannot be simultaneously diagonalized, *but* they do commute up to lower-order terms. The same is true for polynomial operators $P_1(D)$ and $P_2(x)$.

Gaussians. The behavior of f and $\widehat{f}$ is especially easy to see when they are Gaussians. Namely if

$$g(x) = \exp(-x^2/2)$$

, then $\widehat{g} = g$; while if

$$g_a = a^{1/2} g(ax),$$

then $\|g_a\|_2 = \|g\|_2$ and

$$\hat{g_a}(t) = a^{-1/2}\widehat{g}(x/a) = g_{1/a}(t).$$

This shows concentration of position leads to uncertainty in momentum, and vice-versa.

Note. The Gaussians above do not quite give probability distributions; since $|g(x)|^2 = \exp(-x^2) = \exp(-(\sqrt{2}x)^2/2$, we have $\|g\|_2 = \|g_a\|^2 = 2^{-1/4}$.) One can put a factor for $2^{1/4}$ in front of each to get probability distributions.

Momentum. It is at first sight paradoxical that $P(f) = -idf/dx$ is a self-adjoint operator. How can $\langle P(f), f\rangle$ be real when f is a real-valued function?

The answer is: real-valued functions have expected momentum zero. That is,

$$\langle P(f), f\rangle = i\int f'(x)\overline{f}(x)\, dm(x) = i\int d(f(x)^2/2) = 0,$$

since $\overline{f} = f$. Alternatively, note that the Fourier transform of a real-valued function is always satisfies

$$\overline{\widehat{f}}t = \widehat{f}(-t),$$

and thus $|\widehat{f}(t)|^2$ is symmetric in t, so $\int t|\widehat{f}|^2 = 0$.

The Uncertainty Principle. Note that

$$[P, Q] = PQ - QP = -i\frac{d}{dx}x + ix\frac{d}{dx} = -iI.$$

Define the *variation* of P (or Q) by

$$(\Delta P)^2(f) = \langle P^2\rangle - \langle P\rangle^2 = \langle Pf, Pf\rangle - \langle Pf, f\rangle^2,$$

where $\|f\| = 1$. Then the quantities ΔP and ΔQ are the *standard deviations* (which have the same 'units' as P and Q).

Theorem 6.9 *Suppose* $[P, Q] = iI$. *Then*

$$(\Delta P)(\Delta Q) \geq 1/2.$$

Proof. If we add to P or Q multiples of I, their commutator remains the same, so we can assume $\langle Pf, f\rangle = \langle Qf, f\rangle = 0$ — i.e. the expected values of P and Q are zero.

Now we simply apply Cauchy-Schwarz:

$$\begin{aligned} 1 &= |\langle (PQ - QP)f, f\rangle| = |\langle Qf, Pf\rangle - \langle Pf, Qf\rangle| \\ &\leq 2|\langle Qf, Pf\rangle| \leq 2\|Qf\|\|Pf\| = 2(\Delta P)(\Delta Q). \end{aligned}$$

■

The uncertainty spectrum. When examining the uncertainty principle, we can focus on functions with $\|f\|_2 = 1$, and with $\langle P\rangle = \langle Q\rangle = 0$ and $\Delta P = \Delta Q$. For functions so normalized, the quadratic form

$$E(f) = \langle P^2 + Q^2\rangle$$

and the uncertainty are related by

$$(\Delta P)(\Delta Q) = E(f)/2.$$

(This is because the minimum of $|xy|$ subject to $x^2 + y^2 = E$ is $E/2$.) The eigenvalues of $P^2 + Q^2$ turn out to be $2n + 1$, $n \geq 0$, giving $1/2$ as the minimum possible uncertainty, achieved by the Gaussian distribution; but then $3/2$ among normalized functions normal to the Gaussian, and so on.

Intuitive formulation of uncertainty. Here is another way to think about the uncertainty principle.

Theorem 6.10 *If most of* $|f|^2$ *is concentrated in an interval* I, *and most of* $|\widehat{f}|^2$ *is concentrated in an interval* J, *then* $|I| \cdot |J| > 1$ *or so.*

Proof. Suppose $\|f\|_2 = 1$, most of the mass of $|f|^2$ is supported on an interval I and most of the mass of $|\widehat{f}|^2$ lives on J.

Let g be a Gaussian of height 1 and width comparable to $|I|$. Since $g \approx 1$ on most of the support of f, we have

$$\|gf\|_2 \approx 1,$$

and thus

$$\|\widehat{g} * \widehat{f}\|_2 \approx 1.$$

Now the map $\widehat{f} \mapsto \widehat{g} * \widehat{f}$ has norm 1 as an operator on $L^2(\mathbb{R})$, since it is conjugate to multiplication by g and $\|g\|_\infty \leq 1$. Thus if we replace $\widehat{f}$ by the part $\widehat{f_0}$ supported on J, we still have $\|\widehat{g} * \widehat{f_0}\|_2 \approx 1$.

On the other hand, by Cauchy-Schwarz we have

$$\|\widehat{f_0}\|_1 = \int_J |\widehat{f_0}| \cdot 1 \leq \sqrt{|J|}\|\widehat{f_0}\|_2 \leq \sqrt{|J|}.$$

We also have $\|\widehat{g}\|_2 = \|g\|_2 \approx \sqrt{|I|}$, and thus

$$1 \approx \|\widehat{g} * \widehat{f_0}\|_2 \leq \|\widehat{g}\|_2 \|\widehat{f_0}\|_1 \asymp \sqrt{|I||J|}.$$

Thus $|I||J|$ is at least about 1. ■

Tempered distributions. We now wish to extend the definition of the Fourier transform to distributions. But only certain distributions will qualify.

If $u \in C^{-\infty}(\mathbb{R}^n)$ extends from $C_c^\infty(\mathbb{R}^n)$ to a continuous linear functional on $\mathcal{S}(\mathbb{R}^n)$, then we say u is a *tempered distribution*.

Since any continuous linear functional on Schwartz functions restricts to one on the compactly supported functions, the tempered distributions are exactly the dual:

$$\mathcal{S}'(\mathbb{R}^n) = \mathcal{S}(\mathbb{R}^n)^* \subset C_c^\infty(\mathbb{R}^n)^*.$$

Growth. Every tempered distribution has a globally defined order and speed of growth. That is, there exists an N such that

$$|u(f)| \leq C \cdot \sup_x \sup_{|\alpha| \leq N} (1 + |x|^2)^N |D^\alpha f|.$$

Thus temperament is a growth condition at infinity. Note that $\mathcal{S}'$ is closed under $P(x)$ and $P(D)$ for all polynomials P, since $\mathcal{S}$ is.

One can consider ordinary functions $f(x)$ as tempered distributions via

$$\Lambda_f(g) = \int f(x)g(x)\,dm(x),$$

provided a bounded of the form above holds for all $g \in \mathcal{S}$. *Note that we use normalized measure here, not ordinary dx.*

Examples of tempered distributions.

1. Any compactly supported distribution. In particular, differential operators at a point, such that $u = D^\alpha \delta$.
2. Any finite positive measure, or more generally a measure that $\int (1 + |x|^2)^{-N}\,d\mu < \infty$ for some N.
3. Any function $g \in L^p(\mathbb{R}^n)$ for some p, $1 \le p \le \infty$.
4. More generally, any g with $(1 + |x|^2)^{-N} g$ in L^p.

Fourier transform of tempered distributions. There is no reason to expect that the Fourier transform should make sense for an arbitrary distribution. However it does for *tempered* distributions, provided we allow $\widehat{f}$ to also be a distribution.

To motivate the definition of Fourier transform, let us recall that $\mathcal{F}$ is symmetric:

$$\int \widehat{f} g = \int f \widehat{g}.$$

(Note that absence of complex conjugation.) Thus we define, for $u \in \mathcal{S}'$ and $f \in \mathcal{S}$,

$$\widehat{u}(f) = u(\widehat{f}).$$

Since $\mathcal{F}$ preserves $\mathcal{S}$, $\widehat{u}$ is *also* a tempered distribution.

Theorem 6.11 *The Fourier transform is a bijection on $L^2(\mathbb{R}^n)$, $\mathcal{S}$ and $\mathcal{S}'$.*

Examples.

1. The delta function has $\widehat{\delta} = 1$. This is because:

   $$\widehat{\delta}(f) = \delta(\widehat{f}) = \widehat{f}(0) = \int 1 \cdot f(x)\,dm(x).$$

   (It might be better to say $\widehat{\delta} = 1\,dm(x)$.)

   Similarly, $\widehat{1} = \delta_0$, and $\widehat{\delta_p} = \exp(-ipt)$.

2. Let $u = P(D)\delta$. Then $\widehat{u} = P(it)$. Note that

$$u * f = (P(D)\delta) * f = P(D)(\delta * f) = P(D)(f),$$

and thus

$$\widehat{u * f} = \widehat{P(D)f} = \widehat{u}(t)\widehat{f}(t) = P(it)\widehat{f}(t)$$

as discussed before.

3. Similarly, if $u = P(x)$, then $\widehat{u} = P(iD)\delta_0$.

4. If fact, u is a polynomial iff $\widehat{u}$ is supported at the origin, and vice-versa.

Convolution. As before,we define $(u * f)(y) = u_x(f(y - x))$. We then have:

Theorem 6.12 *If $u \in \mathcal{S}'$ and $f \in \mathcal{S}$, then $u * f \in \mathcal{S}'$.*

The usual algebraic relations extend, including $\widehat{u * f} = \widehat{u}\widehat{f}$.

As a simple example to show $u * f$ need not be better than just tempered, notice that the constant function 1 is a tempered distribution and $1 * f$ is the constant function with value $\int f$.

Note that $D(1 * f) = (D1) * f = 1 * Df = 0$ for all Schwartz functions f. The final result says $\int f'(x)\,dx = 0$.

Sobolev spaces. Note that L^2, unlike $\mathcal{S}_n$ and $\mathcal{S}'_n$, is not closed under differentiation. We will now rectify this situation by adding Sobolev spaces to the picture.

The *Sobolev space* H^s, $s \in \mathbb{R}$, is the Hilbert space consisting of tempered distributions u such that $\widehat{u}$ is a measurable function and

$$\|u\|_{H^s}^2 = \int (1 + |t|^2)^s |\widehat{u}(t)|^2 \, dm(t)$$

is finite.

Put differently, if we define a weighted L^2 norm by

$$\|f\|_{L^2_s}^2 = \int (1 + |t|^2)^s |f(t)|^2 \, dm(t),$$

then

$$\|u\|_{H^s} = \|\widehat{u}\|_{L^2_s}.$$

Examples. Thus $L^2(\mathbb{R}^n) = H^0$. The space H^n, $n \geq 0$ consists of functions such that f and all its distributional derivatives $D^\alpha f$, $|\alpha| \leq n$, are in L^2. One can think of H^s as functions with s derivatives in L^2.

Example: L^2-derivatives with f not continuous. For $0 < \alpha < 1/2$, the function $f(x) = |\log r|^\alpha$ on the unit ball in $\mathbb{R}^2$ is in H^1, but it is not continuous. (Here $r = |x|$).

Indeed, the distributional derivative $|\nabla f|$ is proportional to $|\log r|^{\alpha-1}/r$. This derivative is in $L^2(\mathbb{R}^2)$ for $0 < \alpha < 1/2$, even though $f(x)$ is not continuous (or even locally bounded) at $z = 0$.

(To see this, recall that $\int_1^\infty dx/x^\alpha$, $\int_1^\infty dx/(x\log x)^\alpha$, $\int_1^\infty dx/(x \log x(\log\log x)^\alpha)$, etc. are finite iff $\alpha > 1$.)

However if ∇f is in $L^p(\mathbb{R}^2)$, $p > 2$ then in fact f is continuous, as we will later prove.

Example: distributions. We have:

Theorem 6.13 *Every compactly supported distribution is in H^s for some s (often negative).*

Proof. Compactly supported finite measures have bounded Fourier transforms, so they are in H^{-s} for any $s > n/2$; and every compactly supported distribution is a finite sum of derivatives of measures. ■

The Sobolev theorems. Let us begin with:

Theorem 6.14 *If f is in $H^{n/2+\epsilon}$, then f is continuous.*

(In fact, $\widehat{f} \in L^1$ and hence $f \in C_0$.)

Proof. If $\widehat{f}$ is in L^1 then f is continuous. So to prove f is continuous, it suffices to verify that $\widehat{f}$ decays rapidly enough at infinity that it is in L^1. To this end we apply Cauchy-Schwarz to get

$$\int |\widehat{f}| \leq \|(1+|t|^2)^{-s/2}\|_2 \, \|\widehat{f}\|_{H^s}.$$

Now the first term on the right involves (since it is an L^2-norm) the integral of $|t|^{-2s}$ on $\mathbb{R}^n$, so it is finite once $2s > n$, i.e. for $s > n/2$. ■

By similar reasoning we find:

Theorem 6.15 *If f is in $H^{p+n/2+\epsilon}$, then f is in $C_0^p(\mathbb{R}^n)$.*

Corollary 6.16 *If $f \in H^\infty = \bigcap H^s$, then f is in $C_0^\infty(\mathbb{R}^n)$.*

Translation and norms. Unlike the usual L^2 norms, the L_s^2 are not invariant under translation. However they change by a factor with polynomial growth. That is,

$$\sup_t \frac{1+|a+t|^2}{1+|t|^2} = O(1+|a|^2),$$

so

$$\|f(t+a)\|_{L_s^2} \leq O((1+|a|^2)^{|s|/2} \cdot \|f(t)\|_{L_s^2}. \tag{6.2}$$

Orders of operators. An operator $A : \mathcal{S} \to \mathcal{S}$ is of *order* N if it maps H^s to H^{s-N} continuously, for every s.

Examples.

1. The operator $A(u) = D^\alpha(u)$ is of order $d = |\alpha|$, since $|\mathcal{F}(D^\alpha u)| \leq |t|^d \widehat{u}(t)$, and
$$|t|^{2d}(1+|t|^2)^s \leq (1+|t|^2)^{s+d}.$$

2. If $\widehat{Q}$ is in $L^\infty(\mathbb{R})$, then we have an operator order 0 defined by:
$$\widehat{Q(u)} = \widehat{Q}(t)\widehat{u}(t).$$

3. As we will see below, if f is a Schwartz function, then $A_f(u) = fu$ is an operator of order 0. Thus any linear differential operator of the form
$$A(u) = \sum_{|\alpha| \leq N} f_\alpha D^\alpha(u)$$
has order N.

4. An operator is *smoothing* (of order N) if it has order $-N$. For example, if we formally defined $(I-\Delta)^{-1}$ by
$$\widehat{(I-\Delta)^{-1}u} = \widehat{u}(t)/(1+|t|^2),$$
then this operator is smoothing of order 2, and it does in fact invert $(I-\Delta)$.

Here is the main nontrivial result we need on orders of operators:

Theorem 6.17 *For any $f \in \mathcal{S}$, the operator $A_f(u) = fu$ is of order* 0.

Proof. We have

$$\|A_f(u)\|_{H^s} = \|\widehat{f} * \widehat{u}\|_{L^2_s} \le \int_{\mathbb{R}^n} |\widehat{f}(a)| \cdot \|\widehat{u}(t-a)\|_{L^2_s}\, dm(a).$$

Now as a function of a, $\|\widehat{u}(t-a)\|_{L^2_s}$ has at worst polynomial growth, while $\widehat{f}(a)$ is rapidly decreasing. Thus the final integral converges and in fact A_f is a bounded operator on each H_s. ■

Further Sobolev Theorems; Morrey's inequality. We have seen that to get f to be continuous on $\mathbb{R}^n$, it is sufficient to get $> n$ derivatives in L^2. But supposed instead have strong information on ∇f. For example, if ∇f is in L^∞, then f is Lipschitz. More generally, we have:

Theorem 6.18 *If $\nabla f \in L^{n+\epsilon}(\mathbb{R}^n)$, then f is continuous. In fact, if $\nabla f \in L^p$, $n < p < \infty$, then f is Hölder continuous of exponent $1 - n/p$.*

Recovering f from ∇f. Let us first check this on $\mathbb{R}$. There we have, for any interval $I = [a, b]$, and $1/p + 1/q = 1$,

$$|f(a) - f(b)| \le \int_I 1 \cdot |f'(x)|\, dx \le \|\chi_I\|_q \|f'\|_p = |a-b|^{1/q} \|f'\|_p.$$

To carry out a similar calculation in $\mathbb{R}^n$, we need to relate the value of f at a point to an integral of ∇f over a set of positive measure. An elegant form of such a relationship is given by:

Proposition 6.19 *For any $f \in C_0^\infty(\mathbb{R}^n)$, we have*

$$f(y) = (\nabla f) * K, \quad \textit{where} \;\; K(x) = \frac{1}{\mathrm{area}(S^{n-1})} \frac{x}{|x|^{n-1}}.$$

Proof. This formula is obtained by integrating ∇f along radial lines. For example, at $x = 0$ we have:

$$\begin{aligned} f(0) &= c_n \int -\frac{\partial f}{\partial r}\, dr\, d\theta \\ &= c_n \int -\nabla f \cdot \frac{r\widehat{r}}{r^n}\, r^{n-1}\, dr\, d\theta \\ &= c_n \int \nabla f \cdot \frac{-x}{|x|^n}\, dx \\ &= \int \nabla f(x) \cdot K(-x)\, dx = (\nabla f * K)(0), \end{aligned}$$

with K as above. ∎

Proof of Theorem 6.18. First suppose f is smooth and supported in a ball $B(0, r)$. Since $x/|x|^n$ behaves like $1/|x|^{n-1}$, it lies in L^q locally so long as $(n-1)q < n$, i.e. $q < n/(n-1)$. While if ∇f is in L^p, with $n < p < \infty$, then p is dual to such a q (meaning $1/p + 1/q = 1$). Thus we have

$$\|f\|_\infty \le C(n, r)\|\nabla f\|_p.$$

Now suppose f is a compactly supported distribution. Then by convolution we obtain smooth functions with uniformly compact support such that $f_i \to f$ and $\nabla f_i \to \nabla f$ in L^p. These smooth functions are continuous and converge uniformly to f by the bound above, so f is also continuous.

To prove Hölder continuity, it suffices to show that for smooth f and $p > n$ we have

$$|f(a) - f(b)| \le C_n |a - b|^q \cdot \|\nabla f\|_p,$$

since then the approximating f_i will be uniformly Hölder continuous. (Throughout we use C_n to denote a constant depending on n, which can vary with context). Note that this inequality does not require compact support.

It is obtained as follows. Let $d = |a - b|$. For each $x \in \mathbb{R}^n$, $|x| = 1$, connect a to b by a piecewise linear path that runs from a to $a + rx$, then to $b + rx$, then to b. Integrating ∇f along these paths is similar to using the kernel K, but not integrating out to infinity. Using them we find

$$|f(a) - f(b)| \le \int |\nabla f| \cdot |L(x)|\, dx,$$

where L behaves like $1/|x-a|^{n-1} + 1/|x-b|^{n-1} + 1$ and is supported on $B(a, 10d)$. It remains only to estimate the L^q norm of $|L|$. This is done using the fact that, for $r = |x|$, if we integrate in polar coordinates we find:

$$\int_{B(0,d)} |r|^{-q(n-1)}\, dr\, r^{n-1}\, d\omega \le C_n d^{(n-1)(1-q)+1} = C_n d^{n-nq+q},$$

which gives

$$\|L\|_q \le C_n d^{1-n(1-1/q)} = C_n d^{1-n/p}.$$

Note that the Hölder continuity proof does not require compact support, so it also applies to general distributions f with $\nabla f \in L^p$. ■

Compactness in Sobolev spaces. Let $K \subset \mathbb{R}^n$ be a compact set. By the Arzela–Ascoli theorem, the natural map $C^{p+1}(K) \to C^p(K)$ is compact. Using Sobolev spaces we obtain a version of this compactness for fractional derivatives. (Cf. [GH, p.88] which gives a proof where K is replaced by a torus, and the Fourier transform is replaced by Fourier series.)

To state this theorem, let

$$H^s(K) = \{f \in H^s \,:\, \operatorname{supp}(f) \subset K\}.$$

We then have:

Theorem 6.20 (Rellich's Lemma) *If $r > s$, then the natural map $H^r(K) \to H^s(K)$ is compact.*

The main idea in the proof is that the Fourier transform gives a compact operator

$$\mathcal{F} : H^s(K) \to C(\mathbb{R}^n), \tag{6.3}$$

where the target is given the topology of uniform convergence on compact sets. For example, suppose $f \in L^2(\mathbb{R}^n)$, $\|f\|_2 \le 1$, and $\operatorname{supp}(f) \subset B(0,R)$. Then:

1. The Fourier transform $\widehat{f}(t)$ is uniformly bounded. Indeed, we have

$$|\widehat{f}(t)| \le \|e^{ixt}\|_2 = C_R,$$

where the norm is taken in $L^2(B(0,R))$.

2. In fact, all the derivatives of $\widehat{f}$ are uniformly bounded. For example, we have

$$|\widehat{f}'(t)| \leq \|(it)e^{ixt_1} - e^{ixt_2}\|_2 = C'_R.$$

3. Consequently, if f_n is a sequence of functions with unit L^2 norm supported in a fixed ball, then after passing to a subsequence we have a bounded, continuous function $\widehat{g}$ such that $\widehat{f_n} \to \widehat{g}$ uniformly on compact sets.

Proof of Theorem 6.20. We will give the proof when $s \geq 0$ to emphasize the main points. It suffices to treat the case where $K = B(0,R)$. Suppose f_n is a sequence of functions supported on $B(0,R)$ with $\|f_n\|_{H^r} \leq 1$. Then $\|f_n\|_2 \leq 1$ as well, since $r > s \geq 0$. Pass to a subsequence such that $\widehat{f_n}(t)$ converges uniformly on compact sets to $\widehat{g}$. Then $\widehat{g}$ is the Fourier transform of a L^2 function g, and indeed by Fatou's Lemma, we have $\|g\|_{H^r} \leq 1$.

We claim that $f_n \to g$ in H^s. To see this, consider a large ball $B(0,T)$. Then for n large, $|\widehat{f_n} - \widehat{g}|$ is as small as we like on this ball. On the other hand, the ratio $(1+|t|^2)^s/(1+|t|^2)^r$ outside this ball is very small. Since $\|f_n - g\|_{H^r} \leq 1$, this shows that $\|f_n - g\|_{H^s}$ is very small, which gives convergence in the H^s norm. ■

The case $H^s(K)$ with $s < 0$. To handle the case where $s < 0$ we need a different argument to justify compactness of the map (6.3), since we have no *a priori* control on the regularity of f. In this case we use the fact that there exists a test function ϕ such that

$$\phi f = f$$

for all $f \in H^s(K)$. (Simply arrange that $\phi = 1$ on a neighborhood of K.) Because of this, we can write

$$\widehat{f} = \widehat{\phi} * \widehat{f},$$

where $\widehat{\phi} \in \mathcal{S}$ is a fixed, smooth, rapidly decaying function. We then have

$$|\widehat{f}(0)| \leq \int |\widehat{f}(t)| \cdot |\widehat{\phi}(-t)|\, dm(t) \leq \|\widehat{f}\|_{L^2_s} \cdot \|\widehat{\phi}\|_{L^2_{-s}}.$$

By similar reason, we then have

$$\sup_{|\alpha| \leq 1} |D^\alpha \widehat{f}(t)| \leq C(|t|, K) \cdot \|f\|_{H^s},$$

where $C(r,K)$ is a continuous function of r. Thus when we apply $\mathcal{F}$ to the unit ball $H^s(K)$, we obtain a set of continuous functions that are uniformly bounded and equicontinuous on each bounded subset of $\mathbb{R}^n$. Consequently the closure of the image is compact in $C(\mathbb{R}^n)$.

Paley–Wiener theory. It is interesting to find examples where the domain and range of the Fourier transform are *different*, but we still get a bijection. (Such examples are hard to come by; for example, there is no characterization of the Fourier transform of $L^p(\mathbb{R}^n)$, $n > 1$, $p \neq 2$. In fact Fefferman's negative solution to the 'disk multiplier problem' shows there is *no* local characterization.)

Analytic extension of $\widehat{f}(t)$. The next Lemma shows we can speak unambiguously about the analytic extension of f when one exists.

Lemma 6.21 *A function $f(t)$ on $\mathbb{R}^n$ has at most one extension to a complex-analytic function on $\mathbb{C}^n$.*

Proof. Suppose $f(t)$ is analytic on $\mathbb{C}^n$ and $f = 0$ on $\mathbb{R}^n$. If we fix $t_2, \ldots, t_n \in \mathbb{R}$, then $f(t)$ is a function of $t_1 \in \mathbb{C}$, vanishing for $t_1 \in \mathbb{R}$, so f vanishes identically. Thus f does not depend on t_1. By induction, $f(t)$ is independent of t, hence constant and hence zero. ∎

Theorem 6.22 (Paley-Wiener) *The function $f(x)$ is smooth and of compact support if and only if its Fourier transform $\widehat{f}(t)$ has an analytic extension such that there exists R and C_N with*

$$|\widehat{f}(t)| \leq C_N \frac{e^{R|\operatorname{Im} t|}}{(1+|t|^2)^N} \tag{6.4}$$

for all $t \in \mathbb{C}^n$ and $N > 0$.

In fact the condition above characterizes $\widehat{f}$ with $\operatorname{supp} f \subset B(0,R)$.

Remark. The condition on $\widehat{f}$ guarantees that $\widehat{f}$ belongs to $\mathcal{S}$. (Use Cauchy's theorem to get bounds on the derivatives.)

Proof. We give the proof for $n = 1$. Suppose f is supported in $[-R,R]$. Then for $|x| \leq R$ we have

$$|e^{ixt}| = e^{\operatorname{Re} ixt} \leq e^{R|\operatorname{Im} t|}.$$

Thus for $t \in \mathbb{C}$ we have

$$|\widehat{f}(t)| = \left| \int_{-R}^{R} f(x) e^{-ixt} \, dx \right| \leq \|f\|_1 \, e^{R|\operatorname{Im} t|}.$$

Using the fact that

$$\widehat{\frac{d^n f}{dx^n}} = (it)^n \widehat{f}(t),$$

we also obtain boundedness with a polynomial denominator, depending on the L^1-norms of the derivatives of f. This completes the proof of (6.4).

Now suppose $\widehat{f}(t)$ satisfies (6.4). We first observe that for any $s \in \mathbb{R}$ we can invert the Fourier transform by the *complex* path integral

$$f(x) = \int_{\mathbb{R}+is} \widehat{f}(t) e^{ixt} \, dm(t).$$

Indeed, by Cauchy's theorem, the integral of $\widehat{f}(t)e^{ixt}$ around a rectangle is zero; and if we take a rectangle with sides $[-M, M]$ along $\mathbb{R}$ and $\mathbb{R} + is$, and with vertical sides of length $|s|$, then the integral over the vertical parts tends to zero by the rapid decay of $\widehat{f}$, giving the formula above.

Finally we fix x with $|x| > R$ and show $f(x) = 0$. (Since $\widehat{f}$ is analytic and rapidly decaying, we know already that f is in $\mathcal{S}$.) Indeed, for $x > 0$ we can take $s \gg 0$; then for $t \in \mathbb{R} + is$ we have $|e^{ixt}| \leq e^{-xs}$, while

$$|\widehat{f}(t)| \leq C_N e^{Rs} (1 + |t|^2)^{-N}.$$

Thus

$$|f(x)| \leq C_N e^{Rs} e^{-xs} \int (1 + |t|^2)^{-N} \, dt.$$

Taking N large enough, the right-hand side is integrable, and then it tends to zero as $s \to +\infty$ since $x > R$. Thus $f(x) = 0$.

Using s in the lower halfplane, we get the same conclusion for $x < 0$. ∎

Theorem 6.23 *The distribution $u(x)$ has compact support if and only if its Fourier transform $\widehat{u}(t)$ has an analytic extension satisfying, for some R and N,*

$$|\widehat{f}(t)| \leq C e^{R|\operatorname{Im} t|} (1 + |t|^2)^N$$

for all $t \in \mathbb{C}^n$.

In fact the condition above characterizes those $\widehat{u}$ such that $\operatorname{supp} u \subset B(0,R)$; however the order of u may exceed N.

The Heisenberg group $\boldsymbol{H}(\mathbb{Z})$. The group

$$H(\mathbb{Z}) = \langle a, b, c \,:\, [a,b] = c, [a,c] = [b,c] = 1 \rangle$$

is a central extension of $\mathbb{Z}^2$ to $\mathbb{Z}$. It has the remarkable property that the number of elements that can be expressed as words of length at most N in $\langle a, b, c\rangle$ grows like N^4. We have $H(\mathbb{Z}) = \pi_1(E)$ for the circle bundle $E \to S^1 \times S^1$ with first Chern class one.

The Heisenberg group $H(\mathbb{R}^n)$. This group is a central extension of the additive group $\mathbb{R}^{2n}$ by $\mathbb{R}$. The extension is defined by

$$(a,0,0)\cdot(0,b,0) = (0,b,0)\cdot(a,0,0)\cdot(0,0,a\cdot b)$$

where $(a,b) \in \mathbb{R}^{2n}$ and $a \cdot b \in \mathbb{R}$ is the central coordinate. For $n = 1$ this group can be realized as matrices with

$$(a,b,c) = \begin{pmatrix} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{pmatrix}.$$

Writing A_a, B_b and C_c for $(a,0,0)$, $(0,b,0)$ and $(0,0,c)$, we have:

$$A_a B_b = B_b A_a C_{a\cdot b}.$$

Note that $\rho_{s,t}(a,b,c) = (sa, tb, stc)$ is an automorphism of $H(\mathbb{R})$.

The Schrödinger representation. There is a beautiful connection between the Fourier transform and the Heisenberg group $H(\mathbb{R}^n)$.

There is a natural *unitary action* of $H(\mathbb{R}^n)$ on $L^2(\mathbb{R}^n)$, obtained by letting $(a,0,0)$ act by translation in position and $(0,b,0)$ act by translation in momentum. The center element $(0,c,0)$ acts by $e^{ic}I$, i.e. a multiple of the identity operator (which is central in $B(\mathcal{H})$).

In other words, we set

$$\begin{aligned} A_a f(x) &= f(x+a), \\ B_b f(x) &= e^{ibx} f(x), \quad \text{or equivalently} \\ B_b \widehat{f}(t) &= \widehat{f}(t-b), \quad \text{and} \\ C_c f(x) &= e^{ix} f(x). \end{aligned}$$

Then we have:

$$\begin{aligned} A_a B_b \cdot f(u) &= e^{iab} e^{ibx} f(x+a) \\ &= C_{ab} B_b A_a \cdot f(u). \end{aligned}$$

Theorem 6.24 (Stone-von Neumann) *Every irreducible unitary representation of the Heisenberg group is either 1-dimensional, or equivalent to the Schrödinger representation up to an automorphism $\rho_{s,t}$ of $H(\mathbb{R})$.*

The second type of representation is determined uniquely by its central character (the value h such that $\rho(C_c) = e^{ihc}$). For more details, see [Fol, 1.59].

7 Elliptic equations

In this section we discuss linear partial differential equations in general, and elliptic equations in particular, focusing on the case of the $\overline{\partial}$ equation and the Laplacian.

Fundamental solutions. First we give a general result on the *existence* of solutions. Let $P(D) = \sum_{|\alpha| \le N} a_\alpha D^\alpha$ be a linear partial different equation on $\mathbb{R}^n$ with *constant coefficients*. A distribution E on $\mathbb{R}^n$ is a *fundamental solution* for $P(D)$ if

$$P(D)E = \delta_0.$$

A basic result in the theory is:

Theorem 7.1 (Malgrange–Ehrenpreis) *Any linear differential operator $P(D) \neq 0$ has a fundamental solution.*

Corollary 7.2 *If g is a compactly support distribution, then there exists a distribution f such that $P(D)f = g$.*

Proof. Set $f = E * g$. ∎

Remarks. The solution f is only well–defined up to adding a solution to the homogeneous equation $P(D)f = 0$.

Formally, a fundamental solution should satisfy $\widehat{E}(t) = 1/P(it)$. This can in fact be used to find a fundamental solution when $1/P(it)$ is a *tempered distribution*. When E is tempered we can also solve $P(D)f = g$ for any Schwartz function g, and often for more general g.

Warning: it is important to be clear about whether one is using dx or $dm(x)$ in convolutions, to get the constants right. We general use integration against dx in this section.

Examples of fundamental solutions.

1. On $\mathbb{R}$, the fundamental solution to $DE = \delta$ is given by the Heaviside function $E = H(x) = \chi_{[0,\infty)}$. This is simply another way of saying that
$$f(y) = (g * H)(y) = \int_{-\infty}^{y} g(x)\, dx$$
solves $Df = g$.

2. More generally, $(D - \alpha)E = \delta$ has as solution $E_\alpha = e^{\alpha x}\chi_{[0,\infty)}$.
For $\operatorname{Re}\alpha < 0$, this solution is in L^1. Its Fourier transform is just what a formal calculation would suggest: $\widehat{E_\alpha} = C/(it - \alpha)$.
For $\operatorname{Re}\alpha > 0$, we get an L^1 solution by setting $E(x) = -H(-x)e^{\alpha x}$.

3. A general constant-coefficient equation $P(D)u = v$ on $\mathbb{R}$ can be solved as follows, provided $\operatorname{Re}\alpha_i \neq 0$ for all i:
Write $P(D) = \prod(D - \alpha_i)$; then
A fundamental solution is $E = E_{\alpha_1} * \cdots * E_{\alpha_n}$.
In fact:
$$P(D)E = (D - \alpha_1)E_1 * \cdots * (D - \alpha_n) * E_n = \delta * \cdots * \delta = \delta.$$

4. For the Laplacian on $\mathbb{R}^n$,
$$\Delta = \sum_1^n D_i^2,$$
a fundamental for $n > 2$ is proportional to $E = 1/|x|^{n-2}$. Note that ∇E has constant flux through each sphere $S^{n-1}(r)$ and E is harmonic outside $x = 0$.

5. For the Laplacian on $\mathbb{R}^2$, a fundamental solution is proportional to $E = \log |x|$.

6. The operator $P(D) = \overline{\partial}$ has a fundamental solution $f(z) = 1/(\pi z)$. Check the constant.

7. The wave operator
$$\Box = \frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2}$$
factors as $(D_t - D_x)(D_t + D_x)$. A typical solution to $\Box f = 0$ is $f(x,t) = g(x-t) + h(x+t)$.

   A fundamental solution for $\Box$ is proportional to the function $E(x,t) = H(t-x)H(t+x)$, the product of two Heaviside functions. This function is discontinuous along the lines $x = \pm t$, $t > 0$, and equal to one in the 'future cone' $x \in [-t, t]$, $t > 0$.

   To check that E is a fundamental solution, change coordinates so $\Box = D_x D_y$, and $E(x,y) = H(-x)H(-y)$. Then we have
$$\begin{aligned}\int f \Box E &= \int E \Box f \\ &= \int_{-\infty}^{0} \int_{-\infty}^{0} \frac{\partial f}{\partial x \partial y} \\ &= f(0).\end{aligned}$$

8. *Shocks.* Notice that the solution to $\Box E = \delta$ has singularities that are not just concentrated at the origin. This means that the solution to $\Box u = v$ may have singularities *outside* the support of the singularities of v; that is, the singularities can propagate.

9. A fundamental solution for the operator D_x on $\mathbb{R}^3$ is $H(x)\mu$ where μ is linear measure on the line $y = z = 0$. (Thus D_x also has shocks.)

10. Not all linear PDE have solutions! The famous example of Hans Lewy in $\mathbb{C} \times \mathbb{R}$, namely
$$\frac{\partial u}{\partial \overline{z}} + iz \frac{\partial u}{\partial t} = f$$
fails to have solutions for most f. (E.g. if $f = g'(t)$ then f must be real analytic.)

    See Lewy, *Annals of Math.* 66 (1957), pp. 155-158.

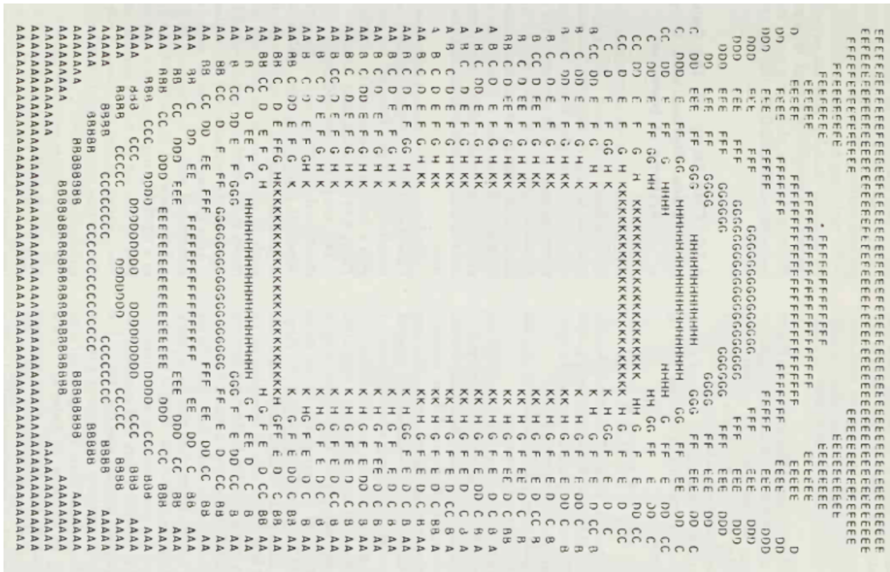

Figure 1. Numerical solution to Laplace's equation, from McCracken, *A guide to Fortran IV programming.*

Relaxation for the Laplacian. The *Dirichlet problem* is to find a solution to $\Delta u = 0$ on Ω given the values of $u|\partial\Omega$. We will not study this problem in detail, however we note that it can be solved numerically by discretization: we seek values of $u(i, j)$ on a grid, such that $u(i, j)$ is the sum of its four neighbors. Once the boundary values are fixed, one can start with any initial u and replace its value at each interior point with the required average, until the process converges.

Ellipticity: Motivation. The automatic regularity of analytic functions is familiar from complex analysis: once $f'(z)$ exists, so do all derivatives of f; in fact f is smooth. This is a consequence of the fact that the equation $df/d\overline{z} = 0$ is *elliptic.*

On domains in $\mathbb{R}^2$, the same is true for the equation

$$\Delta f = \frac{d^2 f}{dx^2} + \frac{d^2 f}{dy^2} = 0;$$

a C^2 harmonic function is automatically C^∞. On the other hand, there are other differential equation such as the wave equation

$$\frac{d^2 f}{dx^2} - \frac{d^2 f}{dy^2} = 0$$

whose solutions are not automatically smooth; for example, any function of the form $f(x-y) + g(x+y)$ satisfies the equation above.

The behavior of these equations is closely tied to the behavior of the associated polynomials – in this case, quadratic forms – x^2+y^2 versus x^2-y^2. The former is called *elliptic*, while the latter is *hyperbolic*. Note that the zeros set of an elliptic form is just the origin, while the zeros set of x^2-y^2 is two lines through $(x, y) = (0, 0)$.

We will start by discussing the regularity of Δ and $\overline{\partial}$, and then the general theory of elliptic operators. We will also touch on the general theory of solutions of linear PDEs with constant coefficients, $P(D)u = \phi$.

The Laplacian. Our main goal is to prove the following regularity theorem, which is a prototype for elliptic regularity in general.

Theorem 7.3 (Weyl's lemma) *Let f be a distribution on $\Omega \subset \mathbb{R}^n$ such that $\Delta f = 0$. Then f is a smooth function.*

For the proof we first analyze the *inhomogeneous* situation on $\mathbb{R}^n$, and then the study Ω using cutoff functions. The reason for this strategy is to analyze $\Delta f = 0$ on Ω, we will study the behavior $\Delta(\phi f)$ on $\mathbb{R}^n$, where ϕ is a test function on Ω; but the product ϕf is no longer harmonic.

Smoothing of $(I - \Delta)^{-1}$. Let $P(x) = \sum_1^n x_i^2$; then $P(D) = \Delta$ on $\mathbb{R}^n$. We can attempt to solve $\Delta f = g$ by turning it into the equation

$$\widehat{P(D)f} = P(it)\widehat{f}(t) = -\left(\sum t_i^2\right)\widehat{f}(t) = \widehat{g}(t).$$

The problem is that $1/P(it)$ is not bounded. To remedy this, we work instead with $P(D) = I - \Delta$, so that

$$P(it) = 1 + \sum t_i^2 \geq 1$$

throughout $\mathbb{R}^n$. It is then clear that we can solve the equation

$$(I - \Delta)f = g$$

whenever g belongs to the Sobolev space H^s, by setting

$$\widehat{f}(t) = \mathcal{F}((I-\Delta)^{-1}g) = \frac{\widehat{g}(t)}{1+|t|^2}.$$

This gives a continuous map

$$(I-\Delta)^{-1} : H^s \to H^{s+2}$$

for all s. In other words, $(I-\Delta)^{-1}$ is *smoothing* of order 2. From this fact we obtain:

Theorem 7.4 *Let f be a compactly supported distribution on $\mathbb{R}^n$, and suppose $g = \Delta f$ lies in H^s. Then f lies in H^{s+2}.*

Proof. Suppose f lies in $H^{s'}$. Then $(I-\Delta)f = f - g$ lies in $H^{\min(s,s')}$, so f lies in $H^{\min(s,s')+2}$. This improves the smoothness of f unless $\min(s,s') = s$, in which case it shows $f \in H^{s+2}$. ■

Fourier transform and distributions. We now turn to the proof of Weyl's theorem for distributions $f \in \mathcal{D}'(\Omega)$.

We remark that the Fourier transforms is not well adapted to the direct study of such distributions. First, the natural setting for the Fourier transform is $\mathbb{R}^n$, where translations are available. But a distribution on Ω can be wildly behaved near $\partial\Omega$ and it need not have any extension to $\mathbb{R}^n$.

And in any case, a distributions on $\mathbb{R}^n$ that is not tempered may have no reasonable Fourier transform.

Both of these difficulties are only apparent. On the positive side, any *compactly supported* distribution is tempered, so it has a Fourier transform; and any such distribution extends to all of $\mathbb{R}^n$. To study local properties of partial differential equations, such as smoothness of solutions, one can replace f by its compactly supported *localizations*, and these suffice to develop a quite general theory.

Localization. We now turn to the study of a distributional solution to $\Delta f = 0$ on a domain Ω. We exploit the fact that the issue of smoothness of f is a local one.

To begin with, let us say that f is *locally in H^s* if for any test function ϕ on Ω, we have $\phi f \in H^s$. (Note that ϕf is tempered so it definitely belongs to *some* Sobolev space.

Proposition 7.5 *If f is locally in H^s, then $D_i f$ is locally in H^{s-1}. More generally, $D^\alpha f \in H^{s-|\alpha|}$.*

Proof. For any test function ϕ, we have

$$D_i(\phi f) = \phi(D_i f) + f D_i \phi.$$

Since f is in H^s locally, the first and last terms are in H^{s-1} and H^s respectively, so the middle term is in H^{s-1}. The second statement follows by induction. ∎

Proposition 7.6 *If f is a distribution on Ω, and then $f|B$ is locally in H^s for some s.*

Proof. Let ψ be a test function with $\psi = 1$ on B. Then $\psi f \in H^s$ for some s; and $\phi f = \phi\psi f$ for any ϕ supported in B. Since multiplication by ϕ has order 0, ϕf is also in H^s. ∎

Proof of Weyl's Lemma. Suppose f is a distribution on Ω and $\Delta f = 0$.

We may assume as above that $\Omega = B$ and f is locally in H^s. It suffices to show ϕf is smooth for all test functions ϕ on B. Since $\Delta f = 0$, we have

$$g = \Delta(\phi f) = \phi \Delta f + (\nabla \phi) \cdot (\nabla f) + (\Delta \phi) f,$$

and the first term vanishes. Thus g is in H^{s-1} (applying Proposition 7.6 to ∇f). But then $\phi f \in H^{s+1}$ by our results on $\mathbb{R}^n$. Since ϕ was arbitrary, this shows that f is locally in H^{s+1}. By induction, we find that f is locally in H^∞ and hence locally smooth. But smoothness is a local property, so f is smooth. ∎

Inhomogeneous equations. By similar considerations, the discussion on $\mathbb{R}^n$ yields:

Corollary 7.7 *If g is locally in H^s on Ω, then any solution to $\Delta f = g$ is locally in H^{s+2}.*

Interior bounds for harmonic functions. A basic property of harmonic functions f is that a uniform bound on $|f|$ on Ω gives a uniform bound on all its derivatives, provided we restrict to a compact set $K \subset \Omega$. More precisely we have:

Theorem 7.8 *Let $K \subset \Omega$ be compact set. Then there exists a constant $C(K, N)$ such that for any bounded harmonic function $f \in C^\infty(\Omega)$, we have*

$$\sup_K |D^\alpha f| \le C(K, |\alpha|) \sup_\Omega |f|.$$

Proof. Let $S \subset C^0(\Omega)$ be the set of bounded, continuous solutions to $\Delta f = 0$. Then S is closed (any C^0 limit is a distributional solution), and all elements of f are smooth. Thus we have a well-defined operator $D_N : S \to C(K)^M$ that records all the derivatives $D^\alpha f$ with $|\alpha| \le N$. Since all elements of S are smooth, D_N has a closed graph, and hence D_N is bounded. ∎

Bounding f by Δf. It is also true that, when f is compactly supported, the size of Δf controls the size of f. This is an effective version of the following simple fact:

Proposition 7.9 *A compact supported harmonic function (or distribution) f is identically zero.*

Proof. Compact support implies that $\widehat{f}(t)$ is smooth, and harmonic implies that $|t|^2 \widehat{f}(t) = 0$; hence $\widehat{f} = f = 0$. ∎

Theorem 7.10 *Let f be a distribution on $\mathbb{R}^n$ with* $\operatorname{supp}(f) \subset B(0, R)$. *We then have, for each s,*

$$\|f\|_{H^{s+2}} \le C(s, R) \cdot \|\Delta f\|_{H^s}.$$

Proof. Since $(I - \Delta)^{-1}$ is smoothing of order 2, we have

$$\|f\|_{H^{s+2}} \le C(\|\Delta f\|_{H^s} + \|f\|_{H^s}).$$

Suppose the Theorem is false. There we can find distributions with $\|f_i\|_{H^{s+2}} = 1$ and $\|\Delta f_i\|_{H^s} \to 0$. By compactness of the inclusion $H^{s+2}(B(0, R))$ into $H^s(B(0, R))$ (see Theorem 6.20), we can assume that $f_i \to g$ in H^s. Then, by the equation above, we have $\|g\|_{H^s} \ge 1/C > 0$. But we also have $\Delta g = 0$, contradicting the fact that a compactly supported harmonic function is zero. ∎

Corollary 7.11 *Let f_n be a sequence of distributions supported in $B(0,R)$. If $\Delta f_n \to 0$, then $f_n \to 0$ in the space of distributions.*

The maximum principle. For the Laplacian, we have the stronger *maximum principle*, which says that $\sup_K |f| = \sup_{\partial K} |f|$ if $\Delta f = 0$; as well as the mean value property, which say that $f(x)$ is the average of f over the sphere $|x-y| = r$.

Because of the latter, if convolve a harmonic function f with a spherically symmetric bump function of mass one, we find $f * \psi_r = f$ (!) This offers an alternate explanation of the smoothness of harmonic functions. However these principles are hard to use for general elliptic operators.

Elliptic operators in general; the symbol. Here is a general definition. Given a domain Ω, and $f_\alpha \in C^\infty(\Omega)$, consider the differential operator of order N defined by

$$P(x,D) = \sum_{|\alpha| \le N} f_\alpha(x) D^\alpha.$$

The *symbol* of $P(D)$ is the homogeneous polynomial in t, defined for each $x \in \Omega$ by

$$P_N(x,t) = \sum_{|\alpha| = N} f_\alpha(x)(it)^\alpha.$$

We say $P(x,D)$ is an *elliptic operator* if for every x, the only real zero of $P_N(x,t)$ is at $t = 0$.

All of the theorems proved above for Δ generalize, in a suitable sense, to elliptic differential operators.

Non-example: the wave operator. The operator $P(D) = \Box = d^2/dx_2^2 - d^2/dx_1^2$ has principal symbol $P_2(it) = t_2^2 - t_1^2$, and this polynomial vanishes on the lines $t_1 = \pm t_2$, so $P(D)$ is not elliptic. The failure of ellipticity is consistent with the irregularity of solutions.

It is worthwhile to compare the factorizations, in $\mathbb{R}^2$:

$$\Box = (D_x - D_y)(D_x + D_y) \quad \text{and} \quad \Delta = (D_x - iD_y)(D_x + iD_y).$$

The fact that $D_x - D_y$ has a large, unruly kernel is responsible for the wildness of solutions to the wave equation $\Box f = 0$. The fact that the Laplacian is tame implies that each of its two factors must be well–behaved as well. In fact these are the operators ∂ and $\overline{\partial}$.

The $\overline{\partial}$ operator. To give the idea of the proof of elliptic regularity the general case, consider next the $\overline{\partial}$ operator. This is given by:

$$P(D)f = \frac{\partial f}{\partial \overline{z}} = \frac{1}{2}\left(\frac{\partial f}{\partial x_1} + i\frac{\partial f}{\partial x_2}\right),$$

where $z = x_1 + ix_2$. The smooth solutions to $P(D)u = 0$ are holomorphic functions. We will see that the same is true for distributional solutions. (One way to see this would be to show that $\partial\overline{\partial}u = C\Delta u = 0$; but we will give a more general argument.)

The symbols of $\boldsymbol{\Delta}$ and $\boldsymbol{\overline{\partial}}$. Note that $\overline{\partial}$ has a *complex* symbol: $P(D) = D_1 - iD_2$, so $\widehat{P}(t) = P(it) = it_1 + t_2$. This operator is indeed elliptic, since $P(it)$ vanishes only when $t_1 = t_2 = 0$.

The case of the Laplacian $P(D) = \sum D_i^2$ is particularly simple to analyze, because its symbol $\widehat{P}(t) = -|t|^2$ is *real and negative*, so we could remove its zeros by forming $1 - \widehat{P}(t)$. But for the $\overline{\partial}$ operator, the function $\widehat{P}(t) + Q = it_1 + t_2 + Q$ has a zero no matter what value we choose for Q. To invert it, we instead choose $Q = P(it)/|P(it)|$. Then $|P + Q| = 1 + |P| \geq 1$ everywhere, and we can complete the argument as before.

Homogeneous elliptic operators. Here is a general statement. To treat the $\overline{\partial}$ operator, and similar *homogeneous* elliptic operators, we prove the following.

Theorem 7.12 *Let $P_N(D)$ be homogeneous elliptic operator of degree N on $\mathbb{R}^n$, with constant coefficients, and suppose f and g are compactly supported distributions satisfying*

$$P_N(D)f = g.$$

Then if $g \in H^s$, we have $f \in H^{s+N}$.

Proof. On the level of the Fourier transform, the operator $Pf = P_N(D)f$ is given by

$$\widehat{f}(t) \mapsto \widehat{P}(t)\widehat{f}(t),$$

where $\widehat{P}(t) = P_N(it)$. Its inverse is formally given by multiplication by $1/\widehat{P}(t)$, but this quantity always blows up at $t = 0$. To assist in inverting P, we define a new operator

$$Q : \mathcal{S}' \to \mathcal{S}'$$

of degree zero which is given at the level of the Fourier transform by multiplication by

$$\widehat{Q}(t) = \frac{\widehat{P}(t)}{|\widehat{P}(t)|}.$$

Then $(P+Q)$ is invertible; in fact, we have

$$|\widehat{P} + \widehat{Q}| = 1 + |\widehat{P}| \geq 1 + C|t|^N$$

and hence the operator $(P+Q)^{-1}$, defined by

$$\widehat{(P+Q)^{-1}}f = (\widehat{P} + \widehat{Q})^{-1}\widehat{f},$$

is smoothing of order N.

Now suppose we have compactly supported distributions f and g in $H^{s'}$ and H^s such that $Pf = g$. Then we have $(P+Q)f = g + Qf$, and thus

$$f = (P+Q)^{-1}(g + Qf).$$

Since Q has order 0, while $(P+Q)^{-1}$ is smoothing of order N, we find $u \in H^{\min(s,s')+N}$. This improves the smoothness of f unless $\min(s,s') = s$, in which case it shows that $f \in H^{s+N}$. ∎

Question. How is ellipticity used in the argument above? By construction, $|\widehat{P} + \widehat{Q}| \geq 1$ almost everywhere (since the zero set of $\widehat{P}$ has measure zero), so its reciprocal is certainly a bounded measurable function, giving a well–defined operator. This would work for any polynomial $P_N(D)$.

Ellipticity is used to insure that $|\widehat{P} + \widehat{Q}| \geq C|t|^N$ for large t, so that its reciprocal will be *smoothing* of order N.

Bounds on compact sets. Generalizing Theorem 7.10, we also have:

Theorem 7.13 *Let $P_N(D)$ be a homogeneous elliptic differential operator of order N. Then for any distribution f supported on $B(0,R)$, and any s, we have*

$$\|f\|_{H^{s+N}} \leq C(s,R) \cdot \|P_N(D)f\|_{H^s}.$$

The proof follows the same lines, and uses the fact that $P_N(D)f = 0$ implies $f = 0$ when f has compact support.

Elliptic regularity in general. Here is the general statement of elliptic regularity.

Theorem 7.14 *Let $P = P(D, x)$ be an elliptic operator of order N on Ω. Then if a distribution f on Ω satisfies*

$$Pf = 0,$$

then in fact f is smooth ($f \in C^\infty(\Omega)$).

More generally, if g is locally in H^s and

$$Pf = g,$$

then f is locally in H^{s+N}.

Commutators. We will prove the elliptic regularity theorem under the simplifying assumption that the *principal symbol* $P_N(D, x)$ has constant coefficients.

Let us begin by formalizing a theorem on interchanging differentiation and multiplication.

Proposition 7.15 *If $P(D, x)$ is a linear differential operator of order N on Ω, and ϕ is a test function, then the operator*

$$[P(D, x), \phi]f = P(D, x)(\phi f) - \phi P(D, x)f$$

has order $N - 1$.

Proof. By Leibniz's rule, the terms in $P(D, x)(\phi f)$ other than $\phi P(D, x)f$ all involve lower order derivatives of f, multiplied by smooth functions. ∎

Proof of Theorem 7.14. We assume the principal symbol $P_N(D)$ is constant. Let $f, g \in C^{-\infty}(\Omega)$ satisfy

$$P(D, x)f = g,$$

where g is locally in H^s. Our aim is to show f is locally in H^{s+N}.

Shrinking Ω slightly, we can assume that f is locally in $H^{s'}$ for some s'. Given a test function ϕ, let us now estimate the smoothness of ϕf. Writing $P(D, x) = P_N(D) + R(D, x)$, where $R(D, x)$ has order $N - 1$, we have

$$P_N(\phi f) = (P - R)(\phi f) = \phi(Pf) + [P, \phi]f - R(\phi f).$$

Now we have an equation with a constant coefficient operator $P_N(D)$, and where both sides are compactly supported. Thus we can apply Theorem 7.12. The operator $[P, \phi] - R$ has order $N-1$, so the last two terms lie in $H^{s'-(N-1)}$, while $Pf = g$ is locally in H^s; so we conclude that ϕf itself lies in $H^{\min(s+N, s'+1)}$. Thus we can either improve our assumed regularity of f, or we find ϕf has order $s+N$. This means f is locally in H^{s+N}. ■

Corollary 7.16 *Let P be an elliptic operator and suppose $K \subset \Omega$ is a compact set. Then there exist constants such that for any solution to $Pf = 0$ on Ω, we have*

$$\sup_K |D^\alpha f| \leq C(K, |\alpha|) \sup_\Omega |f|.$$

Cauchy's integral formula. A familiar example of the Corollary above, for the $\overline{\partial}$ operator, can be proved directly using Cauchy's integral formula:

$$\frac{f^{(n)}(z)}{n!} = \frac{1}{2\pi} \int_\gamma \frac{f(\zeta)\, d\zeta}{(\zeta - z)^{n+1}}.$$

Namely this shows that if $f(z)$ is analytic on a domain Ω, then

$$|f^{(n)}(z)| \leq \frac{n! \sup |f|}{d(z, \partial\Omega)^n},$$

as can be seen by setting $\gamma = \partial B(z, r)$ with $r = d(z, \partial\Omega) - \epsilon$. In particular, if $K \subset \Omega$ is compact then $\inf_K d(z, \partial\Omega) > 0$ and hence we obtain uniform bounds on the derivatives of f on K.

Finiteness. As an application of the regularity corollary, we find:

> *For a vector bundle $E \to M$ on a* compact *manifold, the sections of E satisfying an elliptic differential equation span a finite-dimensional space.*

As a concrete example, we have:

Theorem 7.17 *For any holomorphic vector bundle $E \to M$ over a compact complex manifold, the space of global holomorphic sections $V = \mathcal{O}_E(M)$ is finite-dimensional.*

Proof. Putting any metric on E, we obtain a norm on V by $\|\sigma\| = \sup|\sigma(x)|$. By elliptic regularity, a bound on the sup-norm of σ gives a bound on its gradient, and thus the unit ball in V is compact. Therefore V is finite-dimensional. ■

The Atiyah–Singer *index theorem* aims to compute the dimension of this space. It is a generalization of the Riemann-Roch theorem.

Serre duality and Riemann-Roch. The only step in the proof of Riemann-Roch for Riemann surfaces that is *not* just formal manipulation of sheaves is the Serre duality statement:

$$H^{0,1}(X) \cong \Omega(X)^*$$

i.e. the $\overline{\partial}$-cohomology is dual to the space of holomorphic 1-forms (and its generalization to line bundles).

Proof. The space of smooth $(0,1)$-forms, $C^{0,1}$, is dual to the space of $(1,0)$-distributions, $D^{1,0}$, by

$$\langle \alpha, \beta \rangle \mapsto \int_X \alpha \wedge \beta.$$

Using the estimate we used in the construction of fundamental solutions, one can see that for f with compact support, f is controlled by $\overline{\partial} f$, and thus $\overline{\partial} C^{0,0}$ is a *closed* subspace of $C^{0,1}$.

The quotient space $H^{0,1} = C^{0,1}/\overline{\partial} C^{0,0}$ is therefore dual to the subspace of distributional $(1,0)$ forms such that $\langle \alpha, \overline{\partial}\beta \rangle = 0$ for all smooth functions β. But this is the same as saying that $\langle \overline{\partial}\alpha, \beta \rangle = 0$ for all smooth functions β, which means exactly that the $(1,1)$-form $\overline{\partial}\alpha$ vanishes as a distribution.

By regularity of the $\overline{\partial}$ equation, α is thus a holomorphic 1-form, i.e. $(H^{0,1}(X))^* = \Omega(X)$. ■

Existence of fundamental solutions. We now return to the proof of Theorem 7.1. Let $P(D)$ be a linear differential operator on $\mathbb{R}^n$ with constant coefficients. We wish to show there is a distribution E such that

$$P(D)E = \delta_0.$$

To show E exists, we just need to show:

Theorem 7.18 *The map $P(-D)f \mapsto f(0)$ is continuous on the image of $P(-D)$ in $C_0^\infty(\mathbb{R}^n)$.*

Once we know this, by the Hahn-Banach theorem, this map will extend to a continuous linear functional on the whole space and hence to the required distribution E.

Rapid convergence near $\mathbb{R}^n$. Recall that if f is smooth and compactly supported, then $\widehat{f}(t)$ extends to an entire function of $t \in \mathbb{C}^n$. To apply the Fourier transform, we complement the Paley-Wiener theorem as follows.

Theorem 7.19 *If $f_i \to 0$ in $C_c^\infty(\mathbb{R}^n)$, then for $r > 0$ and $N > 0$, we have*

$$\sup_{t \in \mathbb{R}^n + iB(0,r)} (1 + |t|^2)^N |\widehat{f_i}(t)| \to 0$$

as $i \to \infty$.

Proof. For $N = 0$ use the fact that $\operatorname{supp} f_i \subset B(0,R)$ for some R, and that $|e^{-ixt}| \le M_R$ for $x \in B(0,R)$ and $t \in \mathbb{R}^n + iK$, to conclude that

$$\widehat{f_i}(t) = \int f_i(x) \exp(-ixt)\, dm(t) \le M_R \|f_i\|_1 \to 0.$$

To obtain rapid convergence, differentiate $f_i(t)$ with respect to t and apply the same argument. ■

When this condition is satisfied, we say $\widehat{f_i} \to 0$ *rapidly near* $\mathbb{R}^n$.

A problem in entire functions. We now return to the continuity of $P(-D)f \mapsto f(0)$. Recall that

$$f(0) = \int_{\mathbb{R}^n} \widehat{f}(t)\, dm(t).$$

Passing to Fourier transforms as above, it suffices to prove the following:

Theorem 7.20 *Fix a polynomial $P(t) \neq 0$, and let $g_i(t)$ be a sequence of entire functions such that $P(t)g_i(t) \to 0$ rapidly near $\mathbb{R}^n$. Then $\int_{\mathbb{R}^n} |g_i(t)|\, dt \to 0$.*

We are thus reduced to a problem in complex variables.

Proof: The one-dimensional case. Here is a proof of this theorem in the one-dimensional case (on $\mathbb{R}$) that appeals to the maximum principle.

Choose a compact ball $B \subset \mathbb{C}$ containing all the zeros of P. Then for any entire function $g(t)$, we have

$$\sup_B |g| \leq (\inf_{\partial B} |P|)^{-1} \sup_{\partial B} |Pg|.$$

In particular, $\widehat{f_i} \to 0$ uniformly on B. On the other hand, $|P|$ is bounded below outside B, and thus $\widehat{f_i}(t) \to 0$ rapidly near $\mathbb{R}$. In particular $\int_{\mathbb{R}} |\widehat{f_i}| \to 0$ (since $\int (1+|t|^2)^{-1} < \infty$). ■

The n-dimensional case. We will reduce the case of $\mathbb{R}^n$ to the 1–dimensional case, using the fibration

$$S^1 \to S^{2n+1} \to \mathbb{P}^n(\mathbb{C}).$$

We will show:

Theorem 7.21 *Let $P(t)$ be a nonzero polynomial on $\mathbb{C}^n$. Then for any entire function $g(t)$, we have*

$$|g(t)| \leq C_P \int_{|z|=1} |(Pg)(t+z)| \, |dz|.$$

(Here the integral is over a $2n-1$-dimensional sphere.)

From this theorem it is evident that rapid convergence of $P\widehat{f_i} \to 0$ near $\mathbb{R}^n$ implies the same for $\widehat{f_i}$, and thus implies the existence of a fundamental solution.

Proof: $n = 1$. First suppose $n = 1$, $P(t) = c\prod_1^N (t - a_i)$. We make use of a clever trick: the polynomial $Q(t) = c\prod_1^N (1 - \overline{a}_i t)$ has $Q(0) = c$ and $|Q(t)| = |P(t)|$ when $|t| = 1$. Thus we have

$$\begin{aligned} |g(0)| &= \frac{|Q(0)g(0)|}{|c|} \leq \frac{1}{2\pi|c|} \int_{|z|=1} |Q(z)g(z)| \, |dz| \\ &= C_P \int_{|z|=1} |P(z)g(z)| \, |dz| \end{aligned}$$

where C_P depends only on the leading coefficient of P (not the location of its zeros). Since the leading coefficient is translation invariant, the same result holds for $g(t)$, establishing the bound for $n = 1$.

Proof: general n. For the general case, suppose $P(t)$ has degree N, and let $P_N(t)$ be the part that is homogeneous of degree N. Then for almost every complex line L through z, the restriction $P|L$ is a polynomial of degree N with leading coefficient $c(L)$. As L varies in the projective space $\mathbb{P}^{n-1}$, the coefficient $c(L)$ varies continuously, so it is bounded.

Now for each L, we have

$$2\pi|c(L)g(t)| \leq \int_{S^1(L)} |(Pg)(z+t)|\,|dz|;$$

integrating over projective space, we get

$$|g(t)| \int_{\mathbb{P}^{n-1}} |c(L)| \leq C_n \int_{|z|=1} |(Pg)(z+t)|\,|dz|,$$

giving the required bound. (Here we have used the fact that volume measure on the sphere decomposes as arclength on circles times volume measure on projective space.) ■

Remark. We could also have used just one specific L, depending on P, with $c(L) \neq 0$, to conclude that $Pg_i \to 0$ rapidly near $\mathbb{R}^n$ implies $g_i \to 0$ in the same way.

Introduction to singular integral operators.

1. *Translation and dilation.* An operator that commutes with translations is given by convolution, $Tf = f * K$. When does T commute with dilations? That is, when does $T(f(ax)) = (Tf)(ax)$? We need to have:

$$\begin{aligned}
(Tf)(ax) &= \int f(ax-y)K(y)\,dy \\
&= \int f(ax-ay)K(ay)d(ay) \\
&= \int f(ax-ay)a^nK(ay)dy = \\
T(f(ax)) &= \int f(ax-ay)K(y)\,dy.
\end{aligned}$$

For invariance to hold, we need (at least formally) to have $K(ay) = a^{-n}K(y)$, i.e. $K(y)$ should be *homogeneous of degree* $-n$. Better put, the *measure* $K(y)\,dy$ should be dilatation invariant.

2. *Calderón-Zygmund operators.* References: [St1], [St2].

   These operators, also called singular integral operators, are defined by
$$(Tf)(x) = f * K$$
   where $K(x)$ is a homogeneous kernel of *degree* $-n$, smooth outside $x = 0$, *and* (this is crucial)
$$\int_{S^{n-1}} K(x)\, dx = 0.$$
   These operators commute with both dilatation and translation, as mentioned above.

   Since $|K(x)|$ is integrable at neither zero nor infinity, even when f is a Schwartz function the operator needs to be defined as a principal value, i.e.
$$(Tf)(y) = \lim_{r \to 0} \int |x| > r f(y - x) K(x)\, dx.$$
   Notice that the average of K over a sphere *must* be zero for this integral to even have a chance of converging. In fact the cancellation leads to convergence, and we get
$$T : \mathcal{S}(\mathbb{R}^n) \to \mathcal{S}(\mathbb{R}^n).$$

3. *L^2/L^p-theory.* By the general properties of homogeneous functions, the Fourier transform $\widehat{K}$ of K is of *degree zero*; and by smoothness of K, it is bounded on the sphere, so $\widehat{K} \in L^\infty$. This shows $Tf = f * K$ extends to a bounded operator on $L^2(\mathbb{R}^n)$.

   The main result in the theory is:

   Theorem. Any Calderón-Zygmund operator T extends to a bounded operator
$$T : L^p(\mathbb{R}^n) \to L^p(\mathbb{R}^n)$$
   for $1 < p < \infty$. Moreover T preserves all smoothness classes $C^{k+\alpha}$ with $0 < \alpha < 1$.

4. *The Hilbert transform.* The only example of a Calderón-Zygmund operator on $\mathbb{R}$, up to scale, is the Hilbert transform
$$Hf(x) = f * K(x) = \sqrt{\frac{2}{\pi}} f(x) * \frac{1}{x}.$$

This important operator is related to harmonic conjugation.

Note: the constant factor depends on our definition of convolution; following Royden we convolve using the normalized measure $dm(x) = dx/\sqrt{2\pi}$.

Theorem. The Fourier transform of $K(x) = (2/\pi)^{1/2}x^{-1}$ is $\widehat{K}(t) = -i\,\mathrm{sign}(t)$.

The computation of the Fourier transform rests on the following important integral:

Lemma. For $t > 0$ we have:

$$\int \frac{e^{itx}}{x}\,dx = \pi i\,\mathrm{Res}_{x=0}\,\frac{e^{itx}}{x} = \pi i.$$

Here the integral is taken in terms of the principal value. For $t < 0$ we get $-\pi i$.

Proof. Approximate $\mathbb{R}$ with by segments $[-R, -r] \cup [r, R]$. Adding a pair of half-circles c and C of radii r and R, we obtain a closed contour in $\mathbb{H}$. By Cauchy's integral theorem, the integral around the contour is zero. Since e^{itz} tends to zero rapidly as $\mathrm{Im}\,z \to \infty$, the integral along C is negligible, which around c we pick up $(-1/2)$ of the residue of the integrand at $z = 0$. ∎

Remark: equivalently we have shown:

$$\int_{\mathbb{R}} \frac{\sin(xt)}{x}\,dx = \pi$$

for all $t > 0$.

Returning to the Fourier transform, we find for $t > 0$,

$$\widehat{K}(t) = \frac{1}{\sqrt{2\pi}}\sqrt{\frac{2}{\pi}}\int \frac{e^{-ixt}}{x}\,dx = \frac{-\pi i}{\pi} = -i.$$

For $t < 0$ we get $\widehat{K}(t) = i$.

5. *The Hilbert transform and holomorphic functions.* Now suppose $\widehat{f} \in L^1$ is supported on $[0, \infty)$; then

$$f(t) = \int e^{ixt}\widehat{f}(t)\,dm(t)$$

extends to a *holomorphic* function on the upper halfplane $\mathbb{H} = \{t : \mathrm{Im}(t) > 0\}$. At the same time we have $Tf = -if$.

On the other hand, if $\widehat{f}$ is supported on $(-\infty, 0]$, then f extends to be holomorphic on $-\mathbb{H}$ and $Tf = if$.

Now suppose $f(x)$ is a real-valued function, extending to a *harmonic* function on $\mathbb{H}$ (as it will, e.g. if $f \in C_0^\infty(\mathbb{R})$). Then there is a *harmonic conjugate* $g(x)$, also extending to a harmonic function on $\mathbb{H}$, such that $f(x) + ig(x)$ is *holomorphic* on $\mathbb{H}$. Similarly, $f - ig$ extends to a holomorphic function on $-\mathbb{H}$. Thus we have

$$2H(f) = H(f + ig + f - ig) = -i(f + ig) + i(f - ig) = 2g.$$

In other words, $H(f) = g$ is (the boundary values of) the harmonic conjugate of f.

6. *Residues and the Hilbert transform.* For a more direct analysis of H, note that for suitable functions $f(z)$ analytic in $\mathbb{H}$ we have, at least formally,

$$0 = \int_{\mathbb{R}} \frac{f(z)}{z}\,dz + \int_C \frac{f(z)}{z}\,dz = (-\pi Hf)(0) - \pi i f(0),$$

where C is an infinitesimal half-circle in $\mathbb{H}$ oriented clockwise, and we have implicitly closed the loop $\mathbb{R} \cup C$ with a large circle near infinity. Thus $H(f) = -if$.

7. *L^∞ and BMO.* Using the Hilbert transform it is easy to see that Calderón-Zygmund operators do not have to preserve L^∞. In fact, just consider an analytic $F(z) = f(z) + ig(z)$ on $\mathbb{H}$ such that g is bounded but f is not.

The most basic such example is

$$F(z) = \log(z) : \mathbb{H} \to \{z \,:\, 0 < \mathrm{Im}\, z < \pi\}.$$

Then we have

$$g(x) = \begin{cases} 0 & \text{if } x > 0, \\ \pi & \text{if } x < 0, \end{cases}$$

while $f(x) = \log|x|$. The function $f(x)$, while not in L^∞, is the basic example of a function of *bounded mean oscillation*, meaning

$$\|f\|_{BMO} = \sup_B \frac{1}{|B|}\int_B |f - f_B| < \infty,$$

where the sup is over all balls B, $|B|$ is the measure of B, and f_B is the average of f over B.

It turns out all Calderón-Zygmund operators *do* preserve the space BMO, so BMO is the correct replacement for L^∞ in the theory.

8. *Conformal invariance.* Another characteristic feature of Calderón-Zygmund operators is that they *commute with dilations*. That is, $(Tf)(ax) = T(f(ax))$. This is because, to make convolution with a function f natural, K should transform like a measure. Since

$$K = K(x)|x|^n \frac{dx}{|x|^n},$$

and the first term is a scale-invariant function while the last is a scale-invariant measure, K and hence T commute with dilations.

Conversely, any operator that commutes with both translation and dilatation and is sufficiently smooth, must be a Calderón-Zygmund operator.

This scale-invariance explains why these operators arise in Yang-Mills theory (self-dual 2-forms are conformally invariant) and in complex analysis (holomorphic functions are conformally invariant).

It also explains the importance of the borderline norm L^∞, which is a conformally invariant norm on functions. The BMO-norm is also conformally invariant, and the BMO-functions are preserved by Calderón-Zygmund operators.

One can think of BMO as a 'quantum' replacement of L^∞. It is an appropriate space for the study of 'critical phenomena', where the same pattern appears at all scales; such scale-invariance is characteristic of phase transitions and quantum field theory.)

9. *Bounds for $\overline{\partial}$.* One of the main uses of Calderón-Zygmund operators is to obtain bounds on the solutions to differential equations.

For example, define $Tf = \partial\overline{\partial}^{-1} f$ for functions on $\mathbb{C}$. Since $\overline{\partial}$ goes over to $it/2$ and $\overline{\partial}$ to $i\bar{t}/2$ upon Fourier transform, we have

$$\widehat{Tf} = \frac{\bar{t}}{t}\widehat{f}.$$

Since the factor on the right is homogeneous of degree zero, T is a Calderón-Zygmund operator. *In fact T is an isometry.*

By the general theory we may conclude, for example, that if $\overline{\partial}u = v \in L^p$ then all the derivatives of u are in L^p, for $1 < p < \infty$. Summing up we have:

Theorem. For any compactly supported smooth function $f(z)$ on $\mathbb{C}$, we have:

$$\|\partial f\|_2 = \|\overline{\partial} f\|_2,$$

and for $1 < p < \infty$ we have:

$$\|\partial f\|_p \le C_p \|\overline{\partial} f\|_p.$$

10. *Failure at $p = 1$.* These bounds almost always fail at $p = 1$. For example, the function $f(z) = C/z$ (for suitable C) has $\overline{\partial} f = \delta$, so it is a limit of functions with $\|\overline{\partial} f\|_1 = 1$. On the other hand, $\partial f = -C/z^2$ is not integrable.

11. *Bounds for Δ.* As another example, from the Laplacian we obtain a Calderón-Zygmund operator by:

$$Tf = D^2\Delta^{-1} f = \frac{\partial^2 \Delta^{-1} f}{\partial x_i \partial x_j}.$$

This matrix-valued operator is given at the level of Fourier transforms by

$$\widehat{Tf} = \frac{t_i t_j}{|t|^2}\widehat{f},$$

so it is also Calderón-Zygmund .

Theorem. For any compactly supported smooth function f on $\mathbb{R}^n$, and $1 < p < \infty$, we have:

$$\|D^2 f\|_p \le C_p \|\Delta f\|_p.$$

That is, the Laplacian controls the full matrix of second derivatives.

12. *Quasiconformal maps.* An orientation-preserving homeomorphism $f : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ is *quasiconformal* if its derivatives are in L^2 and there exists a $k < 1$ such that
$$\left|\frac{\partial f}{\partial \overline{z}}\right| \leq k \left|\frac{\partial f}{\partial z}\right|.$$
Noting that
$$Df(v) = f_z v + f_{\overline{z}} \overline{v},$$
we see that Df sends circles to ellipses of bounded eccentricity, preserving orientation. Also we see that Jacobian Jf derivative is comparable to either partial derivative, so for a bound region U we have, at least formally,
$$\text{area}(f(U)) = \int_U |Jf| \geq C \int_U |Df|^2.$$
This is why it is naturally to require derivatives in L^2.

13. *Solution to the Beltrami equation.* In 2 dimensions, quasiconformal maps are very flexible.

    Theorem. Let μ be a measurable function on $\mathbb{C}$ with $\|\mu\|_\infty < k < 1$ and with $\operatorname{supp} \mu \subset B(0, R)$. Then there is a unique homeomorphism $f : \mathbb{C} \to \mathbb{C}$, with $f(z) = z + O(1/z)$ for $|z| \gg 0$, such that
$$\frac{\partial f}{\partial \overline{z}} = \mu \frac{\partial f}{\partial z}$$
as distributions.

    Remark: Covering the sphere with two balls, we also obtain, for any $\mu \in L^\infty(\widehat{\mathbb{C}}, d\overline{z}/dz)$ with $\|\mu\|_\infty < k < 1$, the existence of a quasiconformal homeomorphism with dilatation μ.

    Sketch of the proof. Let us associate to $g \in C_c^\infty(\mathbb{C})$ the unique solution to the equation $\overline{\partial} f = g$ with $f(z) \sim z + O(1/z)$ as $z \to \infty$.

    Then the Calderón-Zygmund operator $T = \partial \overline{\partial}^{-1}$ satisfies
$$T(f_{\overline{z}}) = f_z - 1.$$

    The Beltrami equation is $f_{\overline{z}} = \mu f_z$, where $\|\mu\|_\infty < 1$. As noted above, T is an isometry on $L^2(\mathbb{C})$. By the general theory of Calderón-Zygmund operators, for $p > 2$ we have $\|T\|_{L^p} = C_p \to 1$ as $p \to 2$.

Thus given μ, we can choose p with $\|\mu\|_\infty C_p < 1$. Then it is straightforward to solve the equation

$$v = f_{\overline{z}} = \mu f_z = \mu T(v) + \mu.$$

Namely

$$v = \mu + \mu T(\mu) + \mu T(\mu T(\mu)) + \cdots ,$$

which converges in L^p by contraction of μT (assuming μ is compactly supported).

Then we can integrate v to get f. Since v is in L^p, $p > 2$, the integrated function f is in L^p; in fact it is Hölder-continuous.

To prove f is a homeomorphism, we approximate μ by smooth functions and observe that smooth solutions are *diffeomorphisms* and f^{-1} is equicontinuous. Thus both f and f^{-1} have convergent subsequences, so in the limit of measurable μ we still have a continuous inverse for f.

14. *The Uniformization Theorem.* Cor. For any smooth metric g on S^2, or even measurable metric with bounded eccentricity, there exists a *conformal* homeomorphism $f : (S^2, g) \to \widehat{\mathbb{C}}$; in the sense that the derivatives of f are in L^2, and Df sends g-circles to standard tangent circles for $\widehat{\mathbb{C}}$.

8 The prime number theorem

> *The merits of Wiener's method lie in its great power and generality, and the light which it throws on the whole subject; not in simplicity.*
>
> —Hardy, 1949, p.302.

In this section we connect the study of $L^1(\mathbb{R})$ under convolution, the Fourier transform, and the prime number theorem.

The prime number theorem, proved by Hadamard and de la Vallé Poussin, is the following statement.

Theorem 8.1 *The number of primes $p \leq n$ satisfies*

$$\pi(n) \sim \frac{n}{\log n}.$$

Roughly speaking, the probability that n is prime is asymptotic to the reciprocal of the number of digits of n in base e.

Remark. The *Riemann hypothesis* states that the error term in the prime number theorem is $O(n^{1/2+\epsilon})$. To state this properly, however, one must replace the main term with the (offset) *logarithmic integral*,

$$\mathrm{Li}(x) = \int_2^x \frac{dt}{\log t} \sim \frac{x}{\log x} \sum_0^\infty \frac{k!}{(\log x)^k}.$$

For $n = 10,000$, the nth prime is $p = 104,729$. We find $p/\log p$ and $\mathrm{Li}(p)$ differ from n by 9% and 0.3% respectively.

Heuristic arguments. Let us give a plausibility argument for the prime number theorem. By Euclid's argument (consider $\prod_P p + 1$), they are infinitely many primes. One can go further and show, as Euler did, that $\sum 1/p$ diverges. This fact also shows, however, that the primes have density zero among the integers. So $\sum 1/p$ is like $\sum 1/n$, except the nth prime must be $\gg n$. The simplest series that still diverges is $\sum 1/(n \log n)$, suggesting that the nth prime is about $n \log n$.

Here is another argument, showing the prime number theorem is at least self–consistent. Suppose we take x 'at random'. If it is composite, then it is divisible by a prime $p < x$. If we regard divisibility by different p as independent events, each with probability $1/p$, then the probability that x is prime is $1/\delta$, where

$$\delta(x) = \prod_{p<x} \left(1 - \frac{1}{p}\right)^{-1}$$

Since $\log(1 - 1/p)$ is approximately $-1/p$ when p is large, we can estimate

$$\log \delta(x) \approx \int_2^x \frac{dt}{t\,\delta(t)}.$$

Since the integral on the right is $\log\log x + C$, the guess $\delta(x) = \log x$ is at least self–consistent. (However this reasoning is too rough to show $\delta(x) = A \log x$ is not equally plausible.)

Approach via harmonic analysis. Our goal is to situate the prime number theorem in a more general Tauberian setting, and give Wiener's proof of the theorem. Useful references include: Rudin, Chapter 9; Hardy [Har, Ch. 12]; Wiener [Wie].

Warmup on the circle. Let I be a closed subspace of $C(S^1)$, invariant under multiplication by z and z^{-1}. Then $I \subset M_t$ for some $t \in S^1$, where $M_t = \{f(z) \,:\, f(t) = 0\}$.

Proof. The subspace I is actually an *ideal* in $C(S^1)$, so it is contained in a maximal ideal, and these are all of the form A_t. ■

Convolution and ideals. Recall that $A = L^1(\mathbb{R})$ forms a commutative Banach algebra with respect to convolution and the usual L^1 norm. A subspace $I \subset A$ is an (algebraic) *ideal* if $A * I \subset I$. In this setting it is useful to also require that I is *closed*; then A/I has the structure of a Banach algebra as well.

It is easy to see that:

I is a closed ideal $\iff$ I is translation–invariant.

As an example,

$$M_0 = \left\{ f \in A \,:\, \int f = 0 \right\}$$

is a (proper, closed) ideal, since $\int f * g = (\int f)(\int g)$. By the same taken,

$$M_t = \left\{ f \,:\, \int f(x) \exp(-ixt)\, dx = 0 \right\}$$

is an ideal for each $t \in \mathbb{R}$. In fact these are all *maximal ideals*; we have $A/M_t \cong \mathbb{C}$, and the quotient map is given by $f \mapsto \widehat{f}(t)$.

Using the theory of Banach algebras, we will later show:

Theorem 8.2 (Wiener) *Every proper ideal $I \subset A$ is contained in M_t for some t.*

The idea of this theorem is that the Fourier transform actual gives the canonical representation of A as continuous functions on its space of maximal ideals.

Corollary 8.3 *Given $f \in A = L^1(\mathbb{R})$, we have $\overline{f * A} = A$ if and only if $\widehat{f}(t)$ has no zeros for $t \in \mathbb{R}$.*

The same result holds on $\mathbb{R}^n$; we will only need the case above.

Tauberian theorems. Suppose that, given $g \in L^\infty(\mathbb{R})$, we wish to show

$$\lim_{x \to +\infty} g(x) = L$$

exists. It may be easier to prove a smoothed version of this result: $(f * g)(x) \to L$ for some $f \in L^1(\mathbb{R})$.

A *Tauberian theorem* allows one to remove the smoothing, and conclude that $g(x)$ itself converges, under some hypothesis on g. For our purposes a useful condition is that $g(x)$ is *slowly oscillating*. This means that for all $\epsilon > 0$, there exists an $M, \delta > 0$ such

$$|x - y| < \delta \quad \text{and} \quad x, y > M \implies |g(x) - g(y)| < \epsilon.$$

For example, if $|g'(x)|$ is bounded, or $g(x)$ is uniformly continuous, then g is slowly oscillating. However slow oscillation also allows discontinuities or jumps in the value of $g(x)$, provided they are small when x is large.

Theorem 8.4 (Wiener, Pitt) *Suppose $g(x)$ is slowly oscillating, $\int f = 1$,*

$$(f * g)(x) \to L,$$

and $\widehat{f}(t) \neq 0$ for all t. Then $g(x) \to L$.

Proof. It is convenient to normalize so $L = 0$. Then the set of $\phi \in L^1(\mathbb{R})$ such that $(\phi * g)(x) \to 0$ forms a closed ideal containing f. By the nonvanishing of $\widehat{f}(t)$, it coincides with the whole space.

Now we use the (ϵ, δ) pairs from slow oscillation. Let ϕ_n be a bump function supported on an interval of length $< \delta$, with $\int \phi_n = 1$. Then we have:

$$0 = \lim(\phi_n * g)(x) \leq \liminf g(x) - \epsilon.$$

A similar, reverse inequality holds for $\limsup g(x)$. It follows that $g(x) \to 0$. ∎

Non-example. Let $g(x) = \sin(2\pi x)$, $f(x) = \chi_{[0,1]}(x)$. Then g is slowly varying and $f * g = 0$. But $g(x)$ does not tend to zero. The source of the problem is that $\widehat{f}(t)$ is equal to $C(e^{-it} - 1)/t$, which has periodic zeros.

Wiener's Theorem is a converse: if f 'detects all periodic oscillations', then convergence of g follows form that of $f * g$.

The Mellin transform: Harmonic analysis on $\mathbb{R}_+$. For arithmetic purposes and other applications, it is useful to replace the group $(\mathbb{R}, +)$, with Haar measure dx, with the group $(\mathbb{R}_+, \cdot)$, with Haar measure dt/t.

These two groups are related by $t = \exp(x)$. Thus for $f(t)$ a function on $(0, \infty)$, we define

$$\|f\|_p = \left(\int_0^\infty |f(t)|^p \, dt/t \right)^{1/p}.$$

Convolution is given by

$$(f * g)(u) = \int_0^\infty f(t) g\left(\frac{u}{t}\right) \frac{dt}{t},$$

and the *Mellin transform* of $f(t)$ is given by

$$Mf(s) = \int f(t) t^{-is} \frac{dt}{t}\cdot$$

Note that the Mellin transform and the Fourier transform are related by

$$Mf(t) = \sqrt{2\pi}\mathcal{F}(f(e^x)).$$

Thus the Melllin transform gives an isomorphism

$$M : L^2(\mathbb{R}_+) \to L^2(\mathbb{R}),$$

up to a constant factor that can be absorbed into the definition of the norm.

Properties of the Mellin transform. As usual we have

$$M(f * g) = (Mf)(Mg).$$

We also have, by change of variables for $a > 0$,

$$f(at) * g(t) = f(t) * g(at) = (f * g)(at). \tag{8.1}$$

A key transformation property for us is the following: if $f_a(t) = f(at)$, then

$$Mf_a(s) = a^{is} Mf(s), \tag{8.2}$$

as can be seen by changing variables.

In this setting we will *often* consider $Mf(s)$ as a function of a complex variable, and we will often compute $Mf(s)$ when f is not in L^1. In this case the values of $Mf(s)$ on $\mathbb{R}$ may be obtained by analytic continuation.

Examples of Mellin transforms.

1. Let us being with an improper example. Let $f(t) = \chi_{[1,\infty)}(t)$. Even though f is not in L^1, we can formally compute:

$$Mf(s) = \int_1^\infty t^{-1-is}\,dt = \frac{1}{is}.$$

This formula is correct provided $\mathrm{Re}(1+is) > 1$, which is enough to make the integral converge. We have then formally continued it to the whole plane.

Now let $g(t) = f(t) - f(at)$. We then have

$$Mg(s) = (1-a^{is})Mf(s) = \frac{1-a^{is}}{is}.$$

This function g is now in L^1, its Mellin transform is an entire function, and we have

$$\int_0^\infty g(t)\,dt/t = Mg(0) = \log(1/a)$$

by L'Hôpital's rule. This can be verified directly, e.g. when $a > 1$ we have $g(t) = \chi_{[1,1/a]}$.

2. *The gamma function.* The kernel $f(t) = e^{-t}$ comes up frequently. Note that this is an *additive* homomorphism on the *multiplicative* group $\mathbb{R}_+$. Thus its Mellin transform is like a Gauss sum; it is formally given by:

$$Mf(s) = \int_0^\infty e^{-t}t^{-is-1}\,dt = \Gamma(-is).$$

In this case the integrand blows up at $t = 0$, so we only get convergence when $\mathrm{Re}(is+1) < 1$.

Now recall that the classical Γ function can be defined by

$$\Gamma(s) = \int_0^\infty e^{-t}t^{s-1}\,dt;$$

it satisfies $\Gamma(1) = 1$, and integrating by parts one sees that

$$\Gamma(s+1) = s\Gamma(s), \quad \text{which implies} \quad \Gamma(n+1) = n!.$$

In fact $\Gamma(s)$ analytically continues from $\mathrm{Re}\, s > 0$ to the entire complex plane, where it has no zeros but simple poles at $s = 0, -1, -2, \ldots$, with residue 1 at $s = 0$.

3. The kernel $f(t) = te^{-t}$ is closely related to the gamma function but it has the advantage that it is in L^1. In general, multiplication by t shifts the Mellin in a simple way; in the case at hand, we have

$$Mf(s) = \int e^{-t} t^{-is}\, dt = \Gamma(1 - is).$$

Now $Mf(0) = 1$ (we have integrability), and we continue to have $Mf(s) \neq 0$.

4. Similarly for $f(t) = t\chi_{[0,1]}(t)$ we find

$$Mf(s) = \int_0^1 t^{-is}\, dt = \frac{1}{1 - is} \neq 0$$

for all $s \in \mathbb{R}$.

5. *Step-like functions.* To see how the Riemann zeta–function can enter the discussion, consider

$$f(t) = [t]t^{-1},$$

where $[t]$ is the largest integer not exceeding t. In general, we have

$$\int_0^\infty [t] g(t)\, dt = \sum_{n=1}^\infty \int_n^\infty g(t)\, dt;$$

applying this identity, we find

$$Mf(s) = \sum_n \int_n^\infty t^{-2-is}\, dt = \sum \frac{n^{-1-is}}{1 + is} = \frac{\zeta(1 + is)}{1 + is}\cdot$$

Here we recall that

$$\zeta(s) = \sum_1^\infty \frac{1}{n^s} = \prod_p \frac{1}{1 - p^{-s}}\cdot$$

The ζ function is initially defined by $\operatorname{Re} s > 1$, but it admits an analytic continuation to the whole complex plane with a simple pole, of residue 1, at $s = 1$. An analytic continuation to $\operatorname{Re} s > 0$ can be seen by writing

$$\zeta(s) = \int_1^\infty \frac{dt}{t^s} + \sum_n \left(\frac{1}{n^s} - \int_n^{n+1} \frac{dt}{t^s} \right) = \frac{1}{s - 1} + h(s).$$

Here each term in the sum is $O(1/n^{s+1})$, so $h(s)$ is holomorphic for $\operatorname{Re} s > 0$.

6. *Corrected step functions.* Let $f_a(t) = [at](at)^{-1} = f_1(at)$ for $a > 0$. These functions all satisfy Wiener's condition that $Mf_a(s) \neq 0$ for all s. Unfortunately, none of them are in L^1. To correct this, consider, for $a, b > 0$, the function

$$f(t) = 2f_1(t) - f_a(t) - f_b(t). \tag{8.3}$$

This does lie in L^1, and its Mellin transform is given by

$$Mf(s) = (2 - a^{is} - b^{is})\frac{\zeta(1+is)}{1+is}$$

by equation (8.2) . In fact, using the fact that the residue of $\zeta(s)$ at $s = 1$ is one, we find

$$Mf(0) = \log 1/(ab).$$

Also, provided $(\log a)/(\log b) \notin \mathbb{Q}$, there is no $s \in \mathbb{R}$ such that

$$a^{is} + b^{is} = 2$$

(note that both terms on the left have absolute value 1.) This shows:

> *For typical a and b, the function $f(t)$ satisfies the hypothesis of Wiener's theorem.*

Arithmetic functions. We now move towards the proof of the prime number theorem. We will find that Wiener's theorem can be applied, after we replace $\mathbb{R}$ with $\exp(\mathbb{R}) = \mathbb{R}_+$. The reason for this change of coordinates is that the theory of primes is bound up with the *multiplicative* theorem of $\mathbb{Z}$.

To illustrate the ideas, we will consider two useful functions on $\mathbb{Z}_+$. The *Möbius function* is defined by

$$\mu(p_1 \dots p_e) = (-1)^e,$$

and $\mu(n) = 0$ otherwise. In other words, $\mu(n) \neq 0$ if and only if n is square–free, in which case it measures the parity of the number of prime divisors of

n. Note that the probability that a random number n is square–free is given by
$$\prod_p \left(1 - \frac{1}{p^2}\right) = \frac{1}{\zeta(2)} = \frac{6}{\pi^2} > 0.$$

The second function is the *von Mangoldt* function, defined by
$$\Lambda(p^k) = \log p$$
and $\Lambda(n) = 0$ otherwise.

Sums over divisors. These two functions have beautiful behavior when summed over divisors. Namely, for $n \geq 1$ we have
$$\sum_{d|n} \mu(d) = \begin{cases} 1 & \text{if } n = 1, \text{ and} \\ 0 & \text{otherwise;} \end{cases}$$
and
$$\sum_{d|n} \Lambda(d) = \log n.$$
Both properties follow easily from unique factorization. The first one is based on the inclusion–exclusion identity
$$(1-1)^n = \sum_0^n (-1)^k \binom{n}{k} (1)^k = 0,$$
which holds when $n > 0$.

Remark: Dirichlet convolution. The sums above are special cases of *Dirichlet convolution* of arithmetic functions $f : \mathbb{Z}_+ \to \mathbb{Z}$. This convolution is defined by
$$(f * g)(n) = \sum_{d|n} f(d)g(n/d).$$
One generally restricts attention to *multiplicative functions*, which satisfy $f(ab) = f(a)f(b)$ provided $\gcd(a, b) = 1$.

Such functions form a ring under convolution, with $\mu * 1 = e$. Here e is the identity element and $1(n) = 1$. The Möbius inversion formula says that if
$$f(n) = \sum_{d|n} g(d),$$

then

$$g(n) = \sum_{d|n} \mu(n/d) f(d).$$

It is a consequence of the fact that $f = g * 1$, so $f * \mu = g * 1 * \mu = g * e = g$.

Dirichlet series. We can also associate a Dirichlet series to an arithmetic function. For example, we have

$$\sum \frac{\mu(n)}{n^s} = \frac{1}{\zeta(s)} = \frac{1}{\sum n^{-s}}.$$

The von Mangoldt function is directly related to $\zeta(s)$ as well; we have:

$$\sum \Lambda(n) n^{-s} = -\frac{\zeta'(s)}{\zeta(s)}.$$

Both properties follow from the Euler product formula

$$\zeta(s) = \prod_p \frac{1}{1 - p^{-s}}.$$

Formulations of the prime number theorem. We now let

$$\psi(x) = \sum_{n \le x} \Lambda(n).$$

(It is understood that indices like d, n, m etc. are always ≥ 1.) We are expecting the primes p to occur with density $1/\log(x)$ near a given value of x, so if we weight them by $\log p$, the result should be a measure that is approximately uniform. In fact this statement is equivalent to the prime number theorem.

Theorem 8.5 *If $\psi(x) \sim x$, then $\pi(x) \sim x/\log x$.*

Proof. Suppose $\psi(x) \sim x$. Pick a small $\epsilon > 0$. Then we also have

$$\psi_\epsilon(x) = \sum_{x^{1-\epsilon} < p < x} \Lambda(p) \sim x.$$

Indeed the sum of the terms smaller than $x^{1-\epsilon}$ is bounded by $x^{1-\epsilon} \log x \ll x$, while the excluded terms coming from proper powers of primes, $\Lambda(p^k)$ with $k > 1$, contribute at most $O(x^{1/2} \log x)$.

On the other hand, for p in the range above, $\Lambda(p)$ is within a factor of $(1-\epsilon)$ of $\log x$. Thus

$$\pi(x) - \pi(x^{1-\epsilon}) = \sum_{x^{1-\epsilon}<p<x} 1$$

is very well approximated by

$$\frac{\psi_\epsilon(x)}{\log x} \sim \frac{x}{\log x}.$$

Since $\pi(x^{1-\epsilon}) < x^{1-\epsilon} \ll x/\log x$, we can also drop this quantity, and the prime number theorem follows. ∎

By more subtle reasoning, one can also show (see Hardy) that if we let

$$M(x) = \sum_{n\le x} \mu(n),$$

then to prove the prime number theorem it suffices to show that

$$M(x) = o(x),$$

i.e. $M(x)/x \to 0$.

Smoothed arithmetic function. It is hard to see how to bring a sum over divisors into play with the function $\psi(x) = \sum_{n\le x} \Lambda(x)$, which only uses additive inequality. To do this, we make the following important observation.

Let $f(n)$ be a function on integers $n \ge 1$, and let

$$F(x) = \sum_{n\le x} f(n).$$

We then have:

$$\sum_{m=1}^{\infty} F\left(\frac{x}{m}\right) = \sum_{dm\le x} f(d) = \sum_{n\le x}\sum_{d|n} f(d). \tag{8.4}$$

In other words, we can organize a sum over the pairs (d,m) with $md \le x$ into a sum along the hyperbolas defined by $dm = n$ for each $n \le x$. On each hyperbola we have a sum over the divisors of n.

Here is a typical application which will be the starting point for our proof of the prime number theorem: we have

$$\sum_m \psi\left(\frac{x}{m}\right) = \log[x]!. \tag{8.5}$$

Indeed, we can rewrite this sum as

$$\sum_{n\le x}\sum_{d|n}\Lambda(n) = \sum_{n\le x}\log n.$$

The analogous result for the Möbius function is:

$$\sum_m M(x/m) = 1, \tag{8.6}$$

since the sum can be rewritten as

$$\sum_{n\le x}\sum_{d|n}\mu(n) = \sum_{n\le x}\delta_{n1}.$$

Stirling's formula. As is well–known, we have

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n.$$

We need less than this for our applications, namely

$$\log n! = n\log n - n + o(n).$$

This is easily proved by integrating $\log x$ from 1 to n and estimating the error.

From Stirling to the PNT. The main point in the study of prime numbers is that there products must neatly and uniquely express each positive integer. This point is captured by Stirling's formula. With it, as we will see below, one can show that $\sum_{n\le x}\Lambda(n) = O(x)$. This bound is not obvious because the individual terms have size $\log n$, not 1. It already shows the primes avoid a certain kind of clumping. The idea of

Wiener's approach is to show that a careful consideration of exactly what clumping is ruled out by Stirling's formula leads directly to the prime number theorem, *via* the fact that a certain kernel has nonvanishing Mellin transform. This last point relies on properties of $\zeta(1+is)$.

Ingham's Tauberian theorem. We now turn Wiener's argument into a general statement, due to Ingham.

Theorem 8.6 *Suppose $g(x) \geq 0$ is an increasing function supported on $[1,\infty)$, and*

$$G(x) = \sum_m g\left(\frac{x}{m}\right)$$

satisfies

$$G(x) = Ax \log x + Bx + o(x). \tag{8.7}$$

Then $g(x)/x \to A$ as $x \to \infty$.

Corollary 8.7 *The prime number theorem holds.*

Proof. Take $g(x) = \psi(x) = \sum_{n \leq x} \Lambda(n)$, and conclude from (8.5) that $\psi(x) \sim x$. ∎

Alternatively, one can base the proof of the prime number theorem on:

Corollary 8.8 *We have $M(x) = \sum_{n \leq x} \mu(n) = o(x)$.*

Proof. Take $g(x) = M(x) + x$, use the fact that $\sum M(x/m) = 1$ to deduce that $G(x) = x \log x + o(x)$, and apply Ingham's theorem. Here we have added x to $M(x)$ to insure that $g(x)$ is *increasing*. ∎

Note that we expect $x \log x$ growth in $G(x)$. For example, if $g(x) = x$ then $G(x) = x \sum_{m \leq x} 1/m \sim x \log x$.

Boundedness. We first observe that (8.7) implies $g(x) = O(x)$, i.e. $g(x)/x$ is bounded. This argument will show that $\psi(x) = O(x)$, which is enough to prove Chebyshev's theorem $\pi(x) = O(x/\log x)$.

To see this, note that

$$G(x) - 2G(x/2) = g(x) - g(x/2) + g(x/3) - \cdots \geq g(x) - g(x/2),$$

which together with (8.7) implies

$$g(x) = O(x + g(x/2));$$

iterating gives $g(x) = O(x)$.

Convolution on $\mathbb{R}_+$. Next we would like to bring convolution and the Mellin transform into play. Thus we switch variables from x to t. Recall that $g(t)$, and hence $G(t)$, both vanish on $[0, 1]$.

To see that convolution is involved, note that for any $g(t)$ supported on $[1, \infty)$, we have for $u > 0$:

$$\int_0^\infty [t] g\left(\frac{u}{t}\right) \frac{dt}{t} = \sum_n \int_n^\infty g(u/t)\, dt/t.$$

Changing variables, we find

$$\int_n^\infty g(u/t)\, dt/t = \int_1^\infty g(u/nt)\, dt/t,$$

and hence:

$$\int [t] g(u/t)\, dt/t = \int_1^u G(u/t)\, dt/t = \int_1^u G(t)\, dt/t.$$

This last quantity we understand well via (8.7), while the first quantity can be regarded as a convolution: if we set $h(t) = g(t)/t$ (which is bounded), and $f_1(t) = [t]t^{-1}$, then

$$(f_1 * h)(u) = \int [t] t^{-1} (t/u) g(u/t)\, dt/t = \frac{1}{u} \int_1^u G(t)\, dt/t = H(u). \qquad (8.8)$$

We are thus in a position to apply Wiener's theorem — except that $f_1(t)$ is not in $L^1(\mathbb{R}_+)$.

Use of the behavior of $G(t)$. Using our assumption that

$$G(t) = At \log t + Bt + o(t),$$

we can easily see from (8.8) that

$$(f_1 * h)(u) = H(u) = u^{-1}(Au \log u - Au + Bu + o(u)) = A \log u - A + B + o(1).$$

Now choose $a, b > 0$ and set

$$f(t) = 2f_1(t) - f_a(t) - f_b(t)$$

as in equation (8.3), where $f_a(t) = f_1(at)$. Then $f \in L^1(\mathbb{R}_+)$, and we have

$$Mf(0) = \log(1/(ab)).$$

We also have, using equation (8.1),

$$(f_a * h)(u) = (f_1 * h)(au) = H(au).$$

Now

$$H(au) = H(u) + A \log a + o(1),$$

and therefore

$$(f * h)(u) = 2H(u) - H(au) - H(bu) = A \log(1/ab) + o(1).$$

This shows:

$$(f * h)(u) \to A \cdot \int f(t)\, dt/t.$$

i.e. $h(u) = g(u)/u \to A$, after averaging by f.

Completion of the proof. We can now complete the proof of Ingham's theorem, and hence of the prime number theorem. Since $Mf(s)$ is never zero, by Wiener's theorem we have $(f * h) \to L$ for any f with $\int f(t)\, dt/t = 1$. Since $g(t)$ is increasing,

$$rh(rt) = \frac{g(rt)}{t}$$

is an increasing function of r, with the value $h(t)$ when $r = 1$. Thus if f_1 is supported on $[r, 1]$ and f_2 is supported on $[1, 1/r]$, we have

$$r(f_1 * h)(u) \le h(u) \le (1/r)(f_2 * h)(u)$$

for all u. This shows that any limit of $h(u_n)$ with $u_n \to \infty$ lies in $[rA, A/r]$. Letting $r \to 0$ we find $h(u) \to A$. ■

Zeros of $\boldsymbol{\zeta}$. To keep our proof self-contained, we sketch a proof that $\zeta(s)$ has no zeros on the line $\sigma = 1$, where $s = \sigma + it$.

First we note that $\zeta(s)$ has continues to a meromorphic function on the region $\mathrm{Re}(s) > 0$, with a simple pole at $s = 0$ and elsewhere holomorphic. This continuation can be obtained by noting that

$$\begin{aligned}
\zeta(s) &= \sum \frac{1}{n^s} \\
&= \int_1^\infty \frac{dx}{x^s} + \sum_1^\infty \frac{1}{n^s} - \int_n^{n+1} \frac{dx}{x^s} \\
&= \frac{1}{s-1} + O\left(\sum \frac{1}{|n^{1+s}|}\right)
\end{aligned}$$

gives a convergent expression for $\zeta(s)$ in the plane $\mathrm{Re}(s) > 0$.

The analysis of zeros is easier by considering

$$-\frac{\zeta'(s)}{\zeta(s)} = \sum \Lambda(n) n^{-s},$$

where $\Lambda(n)$ is the von Mangoldt function. This formula is proved by taking the logarithmic derivative of the Euler product formula $\zeta(s) = \prod(1-p^{-s})^{-1}$.

Now let $\sigma = 1+\epsilon$. Since ζ has a pole at $s=1$, we have

$$\sum \Lambda(n) n^{-\sigma} = \frac{1}{\epsilon} + O(1). \tag{8.9}$$

On the other hand, if ζ has a simple zero at $1+it$, then

$$\sum \Lambda(n) n^{-\sigma} n^{-it} = \frac{-1}{\epsilon} + O(1). \tag{8.10}$$

But this will imply (see below) that $\zeta(s)$ has a pole at $1+2it$; indeed, we will see that

$$\sum \Lambda(n) n^{-\sigma} n^{-2it} = \frac{1}{\epsilon} + O(1),$$

and that is a contradiction. The case of a multiple zero is similar.

Here are the details of the final step. The idea is to think of (8.10) dividing by the normalizing factor (8.9) as the integral of $f(n) = n^{-it}$ against a probability measure dm on $\mathbb{N}$. Then

$$|f| = 1 \quad \text{and} \quad \int f\, dm = -1 + O(\epsilon), \tag{8.11}$$

so f is close to -1 most of the time. This should imply that f^2 is close to 1 and hence $\int f^2$ is close to 1, as desired. To justify this, observe that equation (8.11) implies that

$$\int (f+1)(\overline{f}+1)\, dm = \int |f|^2 + \int f + \int \overline{f} + \int 1 = O(\epsilon),$$

and hence $\|f+1\|_2 = O(\sqrt{\epsilon})$. Then by Cauchy–Schwarz we have

$$\left| \int (f^2 - 1) \right| dm \le \|f-1\|_2 \cdot \|f+1\|_2 = O(\sqrt{\epsilon}),$$

so $\int f^2\, dm$ is close to 1 as claimed.

Next steps: primes in arithmetic progressions. A natural sequel to the theory of the primes as a whole, which we will not pursue in detail here, is the study of primes in arithmetic progressions. This is carried out with the aid of *Dirichlet L–functions* of the form

$$L(s) = \sum_n \frac{\chi(n)}{n^s} = \prod (1 - \chi(p)p^{-s})^{-1},$$

where $\chi : (\mathbb{Z}/q)^* \to S^1$ is a *character* of the multiplicative group mod q. The end result is that, provided $\gcd(a, q) = 1$, there are infinitely many primes with $p = a \bmod q$, and they account for a fraction $1/\phi(q)$ of all primes.

The simplest example arises when $q = 4$ and $\chi(1) = 1$, $\chi(-1) = 3$. We then have

$$L(s) = \frac{1}{1^s} - \frac{1}{3^s} + \frac{1}{5^s} - \frac{1}{7^s} + \cdots$$

Note that $L(1) \neq 0$; it is a convergent alternating series. Taking logarithms, we find that for $s > 1$ we have

$$\log L(s) = \sum_p \chi(p)p^s + O(1).$$

Here we have used the fact that

$$-\log(1 - x) = x + x^2/2 + x^3/3 + \cdots = x + O(x^2),$$

and for $s > 1$ we have

$$\sum (1/p)^{2s} < \sum 1/n^2 = O(1).$$

This shows that

$$\sum_{p=1 \bmod 4} p^{-s} - \sum_{p=3 \bmod 4} p^{-s}$$

remains bounded as $s \to 1$. Since $\sum 1/p^s \to \infty$, we conclude that there are infinitely primes congruent to $1 \bmod 4$ and infinitely many congruent to $3 \bmod 4$.

Classical Tauberian Theorems. The problem of assigning a plausible value to divergent series has many interesting solutions. For example, by suitable arguments one can show that:

$$\begin{aligned} 1 - 1 + 1 - 1 + \cdots &= 1/2, \\ 1 - 2 + 3 - 4 + \cdots &= 1/4, \\ 1 + 2 + 3 + 4 + \cdots &= -1/12, \quad \text{and} \\ 1 - 1! + 2! - 3! + \cdots &= 0.59634\ldots \end{aligned}$$

The third formula was obtained by Euler by analytically continuing $\zeta(s)$ to $s = -1$. Similar (but easier) methods allow us to evaluate $\lim s_n$, $s_n = \sum_0^n a_i$, in the first two cases as well.

Say $s_n \to S$ (A) (for Abel), or $\sum a_i = S\,(A)$, if

$$\sum a_n r^n = (1-r)\sum s_n r^n \to S$$

as $r \to 1$. For example, from

$$\sum_0^\infty (n+1)r^n = \frac{d}{dr}\frac{1}{1-r} = \frac{1}{(1-r)^2}$$

we find that $1 - 2 + 3 - \cdots = 1/4\,(A)$.

Say $s_n \to S$ (C) (for Césaro) if

$$\frac{1}{N+1}\sum_0^N s_n \to S$$

as $N \to \infty$. For example, we have $1 - 1 + 1 - 1 + \ldots = 1/2\,(C)$.

Here are two results which hold for any sequence s_n.

1. If $s_n \to S$ in the ordinary sense, then we also have $s_n \to S\,(A)$ and $s_n \to S\,(C)$.

2. If $s_n \to S$ (C) then $s_n \to S\,(A)$.

A converse to such a result (under some additional condition) is called a *Tauberian theorem.*

For example, here are two classical results.

Theorem 8.9 *If $s_n \to S$ (A) and $a_n = O(1/n)$, then $s_n \to S$.*

Theorem 8.10 *If $s_n \to S$ (A) and $s_n = O(1)$, then $s_n \to S$ (C).*

Proof via Wiener's method. As an illustration of the connection with Wiener's method, we will prove Theorem 8.10. In fact for bounded s_n, convergence (A) and (C) are equivalent and both are 'universal' in the sense that they involve kernels with nowhere zero Mellin transforms.

To see this, let us shift perspective to the study, for a given kernel $k(t)$ in $L^1(\mathbb{R}_+, dt)$ and $g(t)$ bounded, the averaging process

$$A(L) = \frac{1}{L}\int_0^\infty g(t)k(t/L)\,dt$$

with $L \to \infty$. For example, if $k(t) = \chi_{[0,1]}(t)$, then

$$A(L) = \frac{1}{L}\int_0^L g(t)\,dt,$$

which is the integral analogue of Césaro summation. Note that $(1/L)\int k(t/L)\,dt$ is independent of L. We can rewrite this integral as

$$A(L) = \int g(t)k(t/L)(t/L)\,dt/t = (f * g)(L),$$

where

$$f(1/t) = tk(t).$$

Now the Mellin transform of $f(1/t)$ is just $Mf(-s)$, so Wiener's condition on f is just:

$$Mf(-s) = \int_0^\infty k(t)t^{is}\,dt \neq 0$$

for all $s \in \mathbb{R}$. We have already seen that this condition holds for $k(t) = \chi_{[0,1]}(t)$ and $k(t) = e^{-t}$, where we get $Mf(-s) = 1/(1+is)$ and $\Gamma(1+is)$ respectively.

Next, note that for $k(t) = e^{-t}$, if we choose L so that $e^{-1/L} = r < 1$, and set $g(t) = s_{[t]}$, then we have $1/L \sim 1 - r$ as $r \to 1$, and hence

$$A(L) \sim (1-r)\sum s_n r^n$$

which is the integral form of Abel's summation method. Since the corresponding kernel satisfies Wiener's condition, we find that Abel convergence and Césaro convergence are equivalent for bounded s_n. (It is routine to check that once the theorem is proved for integrals, it also holds for sums.) ■

Asymptotic series. As just a hint of the broader theory of divergent series, going beyond those that are Abel or Césaro summable, we note, (following Hardy [Har, §2.4]) that Euler showed

$$1 - 1!x + 2!x^2 - 3!x^3 + \cdots = \int_0^\infty \frac{e^{-t}\,dt}{1+xt},$$

as can be seen formally by expanding the integrand into a power series in x and integrating term by term, recalling that $\int t^n e^{-t}\,dt = n!$. In particular it is reasonable to write

$$\sum_0^\infty (-1)^n n! = 0.596347\ldots$$

Appendix: Weyl's law, the heat kernel and Tauberian theorems.

Theorem 8.11 *Let M^n be a closed Riemannian manifold, and let $0 = \lambda_0 < \lambda_1 \le \lambda_2 \le \cdots$ be the eigenvalues of $-\Delta$ on M. Then*

$$N(\Lambda) = |\{k : \lambda_k \le \Lambda\}| \sim C_n \operatorname{vol}(M^n)\Lambda^{n/2}.$$

Example. On the unit circle S^1, the eigenfunctions $\phi_k = \exp(ikx)$ have eigenvalues $\lambda_k = k^2$ under $-\Delta = -d^2/dx^2$. Thus the number of eigenvalues up to Λ is asymptotic to a constant time $\Lambda^{1/2}$.

Discussion: energy. Let us define the energy of a function $f \in C^\infty(M) \subset L^2(M)$ by

$$\begin{aligned} E(f) &= \int_M |\nabla f|^2 = -\int_M f\Delta f \\ &= \langle f, -\Delta f\rangle. \end{aligned}$$

Then $N(\Lambda)$ is the dimension of the largest subspace of f for which $E(f) \le \Lambda\|f\|^2$.

One can make an elementary lower bound for $N(\Lambda)$ by packing M with $\asymp \operatorname{vol}(M)/r^n$ disjoint r-balls, and putting on each one a function f with $\|f\|_2 = 1$, with $\|f\|_\infty \asymp r^{-n/2}$, and with $\|\nabla f\|_\infty \asymp r^{-n/2-1}$. Then $E(f) \asymp r^{-2}$, which shows $N(r^{-2})$ is at least on the order of r^{-n}, and thus

$$N(\Lambda) \ge c_n \operatorname{vol}(M)\Lambda^{n/2}.$$

As in the case of the prime number theorem, the hard part is to show $N(\Lambda)/\Lambda^{n/2}$ actually tends to a limit.

Proof. Consider the heat equation

$$\frac{df}{dt} = \Delta f.$$

On $\mathbb{R}^n$ the fundamental solution is given by $f_t = K_t * f_0$ where the *heat kernel* is given by

$$K_t(x) = c_n t^{-n/2} \exp(-x^2/(4t)).$$

To remember this formula: the standard deviation should grow like $\sqrt{t}$, as in the drunkard's walk; at the factor of $t^{-n/2}$ keeps the total mass constant.

On the manifold M^n the heat kernel can be expressed in terms of the orthonormal eigenfunctions $\Delta\phi_k = -\lambda_k\phi_k$ by

$$K_t(x,y) = \sum_k e^{-\lambda_k t}\phi_k(x)\phi_k(y).$$

For small time, $K_t(x,x)$ behaves as on $\mathbb{R}^n$, and thus the trace of the heat kernel satisfies

$$\mathrm{Tr}(K_t) = \int_M K_t(x,x)\,dx = \sum e^{-\lambda_k t} \sim c_n t^{-n/2}\,\mathrm{vol}(M).$$

Applying a Tauberian theorem as $t \to 0$ lets us recover the distribution of the eigenvalues λ_k. Namely we use the kernel $K(u) = \exp(-u^{2/n})$, $u = t^{n/2}$ and set $\alpha_k = \lambda_k^{n/2}$. Let $f(u)\,du$ denote the measure with a delta mass at each α_k. Then we have:

$$\int_0^\infty f(t)uK(tu)\,dt = \sum uK(\alpha_k u) = t^{n/2}\sum e^{-t\lambda_k} \to c_n\,\mathrm{vol}(M).$$

One can show $MK(s) \neq 0$ for all s, by explicit computation (the Mellin transform comes from the Γ-function, which has no zeros). Indeed, for $K(t) = \exp(-t^\alpha)$ with $\alpha = 1/\beta > 0$, we can set $u = t^\alpha$, $dt = \beta u^{\beta-1}du$, to obtain

$$MK(s) = \int \exp(-t^\alpha)t^{is}\,dt = \int \exp(-u)u^{i\beta s}\beta u^\beta\,(du/u) = \beta\Gamma(\beta(1+is)) \neq 0$$

for $s \in \mathbb{R}$. Indeed, $t \mapsto t^\alpha$ corresponds to the dilation $\log t \mapsto \alpha\log t$ on $(\mathbb{R}, +)$, which changes the Fourier transform without introducing or removing zeros.

By Wiener's Tauberian theorem (with smoothing), we conclude that

$$\int_0^N f(t)\,dt = |\{k : \alpha_k \le N\}| \sim c_n' \operatorname{vol}(M) N,$$

which implies Weyl's law by taking $N = \Lambda^{n/2}$. ∎

9 Banach algebras

In this section we discuss the abstract theory of Banach algebras, and especially the Gelfand representation of commutative Banach algebras as continuous functions on a compact Hausdorff space. We will see that the Fourier transform is a particular instance of this representation, and use the general theory to prove Wiener's Tauberian theorem (used previously to study primes).

Definition. A *Banach algebra A* is a *complex* Banach space equipped with the structure of an algebra (usually with identity but often non-commutative) such that

$$\|ab\| \le \|a\| \cdot \|b\|.$$

Examples.

1. $C(X)$, where X is a compact Hausdorff space. For X discrete we obtain $\mathbb{C}^n$.

2. $\ell^p(\mathbb{Z})$, $1 \le p \le \infty$, with pointwise multiplication.

3. $P(K)$, where $K \subset \mathbb{C}^n$ is compact and $P(K)$ is the uniform closure of the algebra of polynomials. All elements of $P(K)$ are holomorphic on $\operatorname{int}(K)$.

4. The group ring $\mathbb{C}[G]$ of a finite group G, with $\|\sum a_g \cdot g\| = \sum |a_g|$. Note that the multiplication law is a discrete version of convolution.

5. $L^1(\mathbb{R}^n)$ and $\ell^1(\mathbb{Z})$ with convolution.

6. $M(\mathbb{R}^n)$, the algebra of signed measures (with bounded total variation), under convolution.

7. The bounded operators on a Hilbert space, $\mathcal{B}(\mathcal{H})$. For a finite-dimensional Hilbert space we obtain the matrices $M_n(\mathbb{C})$.

8. $L^\infty(E)$, where $E \subset \mathbb{R}^n$ is measurable.

9. $H^\infty(\Delta)$: the bounded analytic functions on the unit disk.

10. $A(\overline{\Delta}) \subset C(\overline{\Delta})$: the analytic functions which extend continuously to the closed disk.

11. $P(K) \subset C(K)$, where $K \subset \mathbb{C}^n$ is compact: the closure of the polynomials in the uniform topology. (Note that $A(\overline{\Delta}) = P(\overline{\Delta})$.)

12. Non-examples: $L^p[0,1]$ with pointwise multiplication, $1 \leq p < \infty$.

Goals. We will eventually see that most *commutative* Banach algebras look like $C(X)$; that is, there is a canonical compact Hausdorff space X associated to A, and a natural map $A \to C(X)$.

This result (the Gelfand-Naimark representation theorem) has several applications:

1. In the case $A = L^1(\mathbb{R}^n)$ we will have $X = \mathbb{R}^n \cup \{\infty\}$ and the map $A \to C(X)$ will be the *Fourier transform*.

2. For a compact set $K \subset \mathbb{C}^n$, the space X for $P(K)$ will be the *polynomial convex hull* of K.

3. For $A = \ell^1(\mathbb{Z})$ we will find $A \subset C(S^1)$ by $a_n \mapsto \sum a_n z^n$.

4. Finally, for a single operator $T \in \mathcal{B}(\mathcal{H})$, the closed subalgebra A of $\mathcal{B}(\mathcal{H})$ generated by T and T^* will play a crucial role in the spectral analysis of T, and we will have $X = \sigma(T)$.

Fixing the norm. If A has a continuous multiplication, then we can map each $a \in A$ into $T_a \in \mathcal{B}(A)$, the algebra of bounded operators on A, with $T_a(b) = ab$. Since A has identity element 1, this map is an embedding, and indeed $\|T_a\| \geq \|a\|/\|1\|$, so the image is closed. Thus $\|T_a\| \asymp \|a\|$ and now the norm is sub-multiplicative. This shows:

A Banach space with a continuous algebra structure has an equivalent submultiplicative norm; and

Any Banach algebra can be realized as a closed subalgebra of $\mathcal{B}(A)$ for some Banach space A.

For example, it is clear that multiplication is continuous in $C^k(\Omega)$ and $C^\alpha(\Omega)$, where Ω is a domain in $\mathbb{R}^n$. By the above result, the norm can be chosen to be submultiplicative.

Adjunction of the identity. Given a Banach algebra A_0 without an identity element (such as $L^1(\mathbb{R}^n)$ under convolution), we can add in the identity to obtain an algebra $A = A_0 \oplus \mathbb{C} \cdot 1$, which becomes a Banach algebra with the new norm $\|(a, \lambda)\| = \|a\| + |\lambda|$.

For example, the uniform algebra $A_0 = C_0(\mathbb{R}^n)$ of functions tending to zero on $\mathbb{R}^n$ is thereby extended to the algebra of asymptotically constant functions. The algebra $L^1(\mathbb{R}^n)$ is extended by adding in the δ-function at zero.

Units. If $x \in A$ has a right *and* left inverse in A, then they are equal, x^{-1} is unique and we say x is a *unit*. The set of all units is denoted $A^\times$.

Example: the shift operator $S(a_i) = (a_{i+1})$ on $\ell^2(\mathbb{N})$ has a one-sided inverse $T(a_i) = (0, a_0, a_1, \ldots)$ satisfying $ST(a) = a$; but $TS \neq \mathrm{id}$. If, however, ST and TS are both invertible, then so are S and T. Indeed, $((TS)^{-1}T)S = I$ and $S(T(ST)^{-1}) = I$.

Topology of units. Let $A^\times$ denote the set of invertible elements in A.

Theorem 9.1 *The open ball of radius* 1 *about the identity consists of units; that is,* $\{1 + x \,:\, \|x\| < 1\} \subset A^\times$.

Proof. We have $(1 - x)^{-1} = 1 + x + x^2 + \cdots$ and the series converges since A is a Banach algebra. ∎

Theorem 9.2 *The set $A^\times$ is open.*

Proof. If y is near $x \in A^\times$ then y/x is near 1 and thus invertible; therefore y is invertible. ∎

Theorem 9.3 *The inverse map $x \mapsto x^{-1}$ is continuous on $A^\times$.*

Proof. The power series shows continuity near 1, which implies continuity near every point. ■

The spectrum. For any element x in an algebra $A/\mathbb{C}$, we define the *spectrum* of x (relative to A) as:

$$\sigma(x) = \{\lambda \in \mathbb{C} \,:\, (\lambda - x) \not\in A^\times\}.$$

The *spectral radius* is given by

$$\rho(x) = \sup\{|\lambda| \,:\, \lambda \in \sigma(x)\}.$$

Examples.

1. For $f \in A = C(X)$, the spectrum $\sigma(f) = f(X)$ detects the *range* of f.

2. For $T \in M_n(\mathbb{C})$, the spectrum $\sigma(T)$ is the finite set of *eigenvalues* λ_i of T, and hence
$$\rho(T) = \sup |\lambda_i|.$$
One can show quite generally, e.g. using Jordan normal form, that
$$\rho(T) = \sup \|T^n\|^{1/n}.$$
Note that for $T = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ we have $\rho(T) = 1$ even though $\|T^n\| \asymp n$.

3. The spectrum of an operator on Hilbert space is like a continuous version of the eigenvalues.

   What is the spectrum of the operator $S(a_0, a_1, \ldots) = (0, a_0, a_1, \ldots)$ on $\ell^2(\mathbb{N})$?

   Note that $\|S^n\| = 1$ for all n, so the spectral radius is 1. On the other hand, note that S has *no eigenvectors.*

   To compute $\sigma(S)$, note that $\ell^2(\mathbb{N})$ can be identified with $H^2(\Delta)$ by sending (a_n) to $f(z) = \sum a_n z^n$. Then $S(f(z)) = zf(z)$. It is then clear that $S - \lambda$ fails to be surjective for all $\lambda \in \Delta$, and hence $\sigma(S) = \overline{\Delta}$.

The spectrum and complex analysis. One of the most important and characteristic results in the theorem of Banach algebras is:

Theorem 9.4 *The spectrum $\sigma(x)$ is always nonempty.*

Proof. If the spectrum is empty, then for any $\phi \in A^*$, $f(\lambda) = \phi(1/(\lambda - x))$ is an analytic function of $\lambda \in \mathbb{C}$ that tends to zero at infinity, and hence is identically zero. Setting $\lambda = 0$ we find that $\phi(1/x) = 0$ for all elements of A^*. By the Hahn–Banach theorem, this implies that $1/x = 0$, a contradiction. ■

For a more detailed study of the spectrum, we will use the power series:

$$\frac{1}{\lambda - x} = \frac{1}{\lambda}\frac{1}{1 - x/\lambda} = \frac{1}{\lambda}\sum_0^\infty \frac{x^n}{\lambda^n}. \tag{9.1}$$

Note that if $|\lambda| > \|x\|$, this power series clearly converges, and so $\lambda \notin \sigma(x)$. This shows:

Proposition 9.5 *We have $\rho(x) \le \|x\|$. In particular the spectrum of x is a nonempty, compact subset of $\mathbb{C}$.*

Example. For $f \in A = C^k[0,1]$, $k \ge 1$, we have $\rho(f) = \sup |f|$ but $\|f\|$ involves $\sup |D^\alpha f|$. Thus we often have $\rho(f) \ll \|f\|$.

The spectral radius formula. Leveraging the argument further we will obtain an exact formula for the spectral radius. We begin by noting:

Proposition 9.6 *For any $x \in A$, the sequence $\|x^n\|^{1/n}$ converges.*

Proof. The proof is based on the simple fact that the norms form a sub-multiplicative sequence, i.e.

$$\|x^{a+b}\| \le \|x^a\| \cdot \|x^b\|.$$

Supposer $\|x^a\| = R^a$. Writing $n = am + b$ with $0 \le b < a$, we then have

$$\|x^{am+b}\| \le R^{am}\|x\|^b;$$

taking the nth root and passing to the limit then easily gives

$$\limsup \|x^n\|^{1/n} \le R = \|x^a\|^{1/a}.$$

Now take the lim inf on the right to conclude that the lim sup and lim inf agree. ■

Theorem 9.7 *For any $x \in A$, we have $\rho(x) = \lim \|x^n\|^{1/n}$.*

Proof. The proof will be based on two basic facts from complex analysis about the series $f(z) = \sum a_n/z^n$ and the quantity $R = \limsup |a_n|^{1/n}$:

(i) If $|z| > R$ then the series converges; and

(ii) If $f(z)$ extends to an analytic function on the region $|z| > S$, then $S \geq R$.

Applying (i) we conclude that the power series (9.1) converges for $|\lambda| > \limsup \|x^n\|^{1/n}$ and hence:

$$\rho(x) \leq \limsup \|x^n\|^{1/n}.$$

To use principle (ii), choose as before $\phi \in A^*$ and consider the analytic function $f(\lambda) = \phi(1/(\lambda - x))$, which we can expand as a power series:

$$f(\lambda) = \frac{1}{\lambda} \sum_0^\infty \frac{\phi(x^n)}{\lambda^n}.$$

Since this is analytic for $|\lambda| > \rho(x)$, by (ii) we have:

$$\limsup |\phi(x^n)|^{1/n} \leq \rho(x).$$

Now we use the uniform boundedness principle to replace $|\phi(x^n)|$ with $\|x^n\|$. To do this, we just observe that the above inequality implies that for each $\delta > \rho(x)$ we can find an M_ϕ such that

$$\sup_n |\phi(x^n/\delta^n)| \leq M_\phi.$$

By the uniform boundedness principle, this implies that $\|x^n/\delta^n\| \leq M$, and hence

$$\limsup \|x^n\|^{1/n} \leq \delta.$$

Since $\delta > \rho(x)$ was arbitrary, the proof is complete. ■

Quotient fields. Already the fact that the spectrum is nonempty has profound consequences. Here are the main ones.

Theorem 9.8 *If A is a division algebra (every nonzero element has an inverse), then $A = \mathbb{C}$.*

Proof. For every $x \in A$, there is a $\lambda \in \mathbb{C}$ such that $\lambda - x$ is non-invertible, and hence $x = \lambda$. ■

Note that the quaternions are *not* a Banach algebra (over $\mathbb{C}$): this is because its center is $\mathbb{R}$, not $\mathbb{C}$.

Multiplicative linear functionals. A complex linear *algebra* homomorphism $\phi : A \to \mathbb{C}$ is called a multiplicative linear functional. (By definition, $\phi(1) = 1$.)

Such a linear map ϕ is *automatically* continuous on A, since it is nonzero on the open set $A^\times$. In fact we have

$$|\phi(x)| \le \rho(x) \le \|x\|$$

and hence $\|\phi\| \le 1$. (To see this, write $|\phi(x)| = \lim |\phi(x^n)|^{1/n}$.)

The spectrum of an algebra. Next, we let

$$X = \operatorname{Spec} A = \{\phi \in A^* \,:\, \phi \text{ is multiplicative}\},$$

endowed with the weak* topology. It is easy to see that X is closed and hence by Alaoglu's theorem, X is a compact Hausdorff space canonically associated to A.

Prime example. An important example arises when $A = C(Y)$ is already the Banach algebra of continuous functions on a compact Hausdorff space Y. We then have:

Theorem 9.9 *The spectrum of $A = C(Y)$ is naturally isomorphic to Y.*

Proof. The point evaluations give a natural map $Y \to \operatorname{Spec} A$, which is injective by Urysohn's lemma and easily seen to be continuous.

For the converse, given $\phi \in \operatorname{Spec} A$ let $Z \subset Y$ be the largest set where all $f \in M = \operatorname{Ker}(\phi)$ vanish. (It is the intersection of the zero sets of each f in the kernel). If Z is empty then using compactness we can find $f_1, \ldots, f_n \in M$ such that $g = \sum |f_i|^2 > 0$ at all points; then $g \in M$ as well, and $(1/g) \in A$, so $1 \in M$, which contradicts the assumption that M is a proper ideal.

Thus we can find at least one point with $p \in Z$. Then we have $\operatorname{Ker}(\phi) \subset \operatorname{Ker}(\phi_p)$ which implies $\phi = \phi_p$. ∎

This example show that the algebra A knows about the space Y. In fact, since the topology on A is not used in the definition of $\operatorname{Spec} A$, we can recover the set $X = \operatorname{Spec} A$ from A itself. Then the topology on is also determined by A, simply by taking the weakest topology such that each $a \in A$ defines a continuous function on $\operatorname{Spec} A$.

The idea of moving the algebra A to the fore, and regarding the topological space $\operatorname{Spec} A$ as a 'feature' of the algebra, is the basis for Grothendieck's theory of schemes.

Commutative Banach algebras. For the remainder of this section we assume that A is a *commutative* Banach algebra. Commutativity has several useful consequences. First, it brings A closer to algebras like $C(X)$. Second, it means every $a \in A$ that is not invertible generates an ideal aA, which is contained in a maximal ideal M. We have seen that $A/M \cong \mathbb{C}$. So we have:

Theorem 9.10 *Every maximal ideal M in a commutative Banach algebra has the form $M = \operatorname{Ker}(\phi)$ for some $\phi \in \operatorname{Spec}(A)$.*

The Gelfand transform. We now have a natural surjective map

$$A \to \widehat{A} \subset C(X) = C(\operatorname{Spec} A),$$

sending each $a \in A$ to the function

$$\widehat{a}(\phi) = \phi(a)$$

on X. This map is the *Gelfand transform* on A.

The Gelfand transform satisfies $\widehat{ab} = \widehat{a} \cdot \widehat{b}$ because each $\phi \in X$ is multiplicative. The function $\widehat{a}(x)$ is continuous on all of A^* in the weak* topology, so it is certainly continuous on $X = \operatorname{Spec} A$. It is immediate from the definitions that:

Theorem 9.11 *The algebra $\widehat{A}$ separates points and contains the identity. Thus if $\widehat{A}$ is closed under complex conjugation, it is dense in $C(X)$.*

Perhaps the most important feature of the Gelfand transform $\widehat{a}$ is that it detects invertibility of a, and hence it also detects the spectrum of a.

Theorem 9.12 *The element $a \in A$ is invertible if and only if $\widehat{a}(x) \neq 0$ for all x.*

Proof. Clearly if a is invertible then $\phi(a)\phi(a^{-1}) = 1$ and hence $\phi(a) \neq 0$.

Conversely, suppose a is not invertible. Then the proper ideal aA is contained in a maximal ideal $M = \operatorname{Ker} \phi$ for some $\phi \in X$, and hence $\widehat{a}(\phi) = 0$. ∎

Corollary 9.13 *If $\widehat{a} \in A$ has no zeros, then $1/\widehat{a}(x) \in \widehat{A}$.*

Corollary 9.14 *We have $\rho(a) = \sup |\widehat{a}(x)|$, and $\sigma(a) = \widehat{a}(X)$.*

Algebras and general topology. Let $c_0 \subset \ell^\infty(\mathbb{N})$ be the Banach algebra of sequences with $a_n \to 0$. Then adjunction of the identity yields the algebra $A = c$ of convergent sequences (with the sup norm). It is then easy to see that $\operatorname{Spec} A = \widehat{N} = \mathbb{N} \cup \{\infty\}$ is the 1 point compactification of the natural numbers, and indeed $A = C(\widehat{N})$.

More generally, suppose X is a locally compact Hausdorff space, and $A_0 = C_0(X)$ with the sup–norm. Then A_0 is a commutative Banach algebra *without* an identity element. If we adjoin the identity, it is readily verified that $A = C(\widehat{X})$, where $\widehat{X} = X \cup \{\infty\}$ is the one–point compactification of X.

For a more extreme example, let $A = \ell^\infty(\mathbb{Z})$. Then $X = \operatorname{Spec} A$ contains $\mathbb{Z}$ itself as a dense subspace. The space X is the *Stone-Čech compactification* of $\mathbb{Z}$ — a very large space. For example, the sequence $x_n = n \in \mathbb{Z}$ has no convergent subsequence, even thought X is compact. Every bounded function on $\mathbb{Z}$ has a *continuous* extension to X. The space X is the largest compactification of $\mathbb{Z}$, in the sense that every other compactification is a quotient of it.

In general, if X is a normal topological space, its Stone-Čech compactification can be defined as $\beta X = \operatorname{Spec} A$, where A is the Banach algebra of all bounded, continuous functions on X.

A more standard definition is $\beta X = \overline{\iota(X)}$, where $\iota : X \to [0,1]^C$ and C is the set of continuous maps $f : X \to [0,1]$. The map ι sends x to the sequence $(x_f) = (f(x))$. In both cases, compactness of βX ultimately rests on Tychonoff's theorem, that $[0,1]^S$ is compact for any set S.

For an even more exotic space, try to visualize $\operatorname{Spec} L^\infty[0,1]$. (There is no constructive way to define even one point in this space!)

The radical. The (Jacobsen) *radical* of a commutative Banach algebra A can be described equivalently as follows:

1. $\operatorname{rad}(A) = \bigcap M$ over all maximal ideals $M \subset A$ (a purely algebraic definition); or

2. $x \in \operatorname{rad}(A)$ iff $\rho(x) = 0$ (i.e. the spectrum of x is trivial); or

3. $x \in \operatorname{rad}(A)$ iff $\|x^n\|^{1/n} \to 0$.

Since $\rho(f) = \sup |\widehat{f}|$, we have:

Theorem 9.15 *The kernel of the Gelfand representation $A \mapsto \widehat{A}$ is the radical of A.*

Examples: All nilpotent elements are in the radical. The operator $I \in A = \mathcal{B}(C[0,1])$ given by

$$(If)(x) = \int_0^x f(t)\,dt$$

is in the radical of A. Indeed,

$$|I^n(f)(x)| \leq \|f\| x^n/n! \leq \|f\|/n!.$$

Note that the matrix for I on the polynomials $\mathbb{C}[x^n]$ with basis x^n has entries $I_{ij} = 1/j$ if $j = i+1$, and 0 otherwise. This matrix is 'ω-nilpotent'.

The algebra determines the topology. A Banach algebra is *semisimple* if its radical is trivial.

Theorem 9.16 *Let A and B be semisimple commutative Banach algebras. Then any algebraic homomorphism $f : A \to B$ is continuous.*

In particular, if A and B are isomorphic as $\mathbb{C}$-algebras, then they are isomorphic as Banach algebras.

Proof. First note that, from the simple fact $\rho(x) \leq \|x\|$, we have $|\phi(x)| \leq \|x\|$ for any $\phi \in \operatorname{Spec} A$. This shows any multiplicative linear functional is automatically continuous.

Now we apply the closed graph theorem. Suppose $a_n \to a$ and $f(a_n) \to b$. Then for any $\phi \in \operatorname{Spec} B$ we have $\phi \circ f \in \operatorname{Spec} A$ and so both are continuous. Thus $\phi(f(a)) = \phi(b)$. Since B is semisimple and equality holds for all ϕ, we conclude that $f(a) = b$, and by the closed graph theorem f is continuous. ■

Absolutely convergent Fourier series. Here is a famous result that is difficult to prove directly, but which flows easily from the general theory of Banach algebras.

Let $A = \ell^1(\mathbb{Z})$ under convolution. Note that this algebra has an identity element, the δ function at $n = 0$; and that $\widehat{\mathbb{Z}} = \operatorname{Hom}(\mathbb{Z}, S^1) = S^1$. We then have a Fourier transform

$$\mathcal{F} : \ell^1(\mathbb{Z}) \to C(S^1)$$

given by $\mathcal{F}(a_n) = \widehat{a}(\lambda) = \sum a_n \lambda^n$. This map is injective and its image, $\widehat{A}$, consists exactly of those continuous functions whose Fourier series is *absolutely convergent.* (Note: C^1 functions have this property, but most continuous functions do not.)

Theorem 9.17 *The Fourier transform on $A = (\ell^1(\mathbb{Z}), *)$ agrees with the Gelfand transform; in particular,* $\operatorname{Spec} A = S^1$.

Proof. We have a natural map $e : \mathbb{Z} \to \ell^1(\mathbb{Z})$ sending $i \in \mathbb{Z}$ to the sequence $a_n = \delta_{in}$. This map satisfies $e(i+j) = e(i) * e(j)$, and $a = \sum a_n e(n)$ for any $a \in \ell^1(\mathbb{Z})$. Thus any $\phi \in \operatorname{Spec}(A)$ is uniquely determined by the value $\lambda = \phi(e(1))$, and satisfies $\phi(e(n)) = \lambda^n$. Since $\phi(e(n))$ is bounded, we have $|\lambda| = 1$, and therefore $\phi(a) = \sum a_n \lambda^n$. Conversely, all linear functionals ϕ of this form are multiplicative. ∎

Note: as in the case of the Fourier transform, while $\mathcal{F}$ sends $\ell^2(\mathbb{Z})$ to $L^2(S^1)$, it is very hard to characterize the functions in the image of $\ell^1(\mathbb{Z})$. Nevertheless we have the following striking result, which defeated direct attempts at proof:

Corollary 9.18 (Wiener) *If $f(\lambda) = \sum a_n \lambda^n$ has an absolutely convergent Fourier series (so in particular f is continuous), and f vanishes nowhere on the circle, then $1/f$ also has an absolutely convergent Fourier series.*

Proof. Since $f = \widehat{a}$ is nowhere vanishing, its inverse $1/f$ also belongs to $\widehat{A}$. ∎

It is interesting to reflect on the fact that proof of this rather concrete result is a rather elaborate argument by contradiction, in which the axiom choice is used to construct a maximal ideal containing f!

Divergent Fourier series. It is easy to see from the preceding result that a generic function in $C(S^1)$ does not have Fourier series in $\ell^1(\mathbb{Z})$. Otherwise we would have $\ell^1(\mathbb{Z}) \cong C(S^1)$ and hence $\sum |a_n| \le C\|f\|_\infty$. This would imply the same for L^∞; but the Fourier series of a step function behaves like $\sum a_n z^n / n$, so this is false.

The Fourier transform and the Gelfand transform. We now turn from Fourier series to the Fourier transform, and show it too can be identified with the Gelfand transform.

Lemma 9.19 *Every continuous homomorphism $\phi : \mathbb{R}^n \to S^1$ is given by $\phi(x) = \exp(ix \cdot t)$ for some $t \in \mathbb{R}^n$.*

Theorem 9.20 *Every nonzero multiplicative linear functional $\phi : (L^1(\mathbb{R}^n), *) \to \mathbb{C}$ is given by $\phi(f) = \widehat{f}(t)$ for some $t \in \mathbb{R}^n$.*

Proof. There is a $g \in L^\infty(\mathbb{R}^n)$ such that $\phi(f) = \int fg$ for all $f \in L^1(\mathbb{R}^n)$. Choose $F \in L^1(\mathbb{R}^n)$ such that $\phi(F) = \int fg = 1$. Let $F_a(x) = F(x+a)$ and define a continuous function on $\mathbb{R}^n$ by

$$G(a) = \phi(F_a) = \phi(F(x+a)).$$

We claim that $G(a) = g(a)$ a.e., and that $G(a+b) = G(a)G(b)$. These two properties mean that $g : \mathbb{R}^n \to S^1$ is a continuous homomorphism, and thus $g(x) = \exp(-itx)$ for some t, and hence $\phi(f) = \widehat{f}(t)$.

To see $G = g$ a.e., construct a sequence of functions $k_n \geq 0$ in L^1 such that $\int k_n = 1$ and $k_n^0 \to \delta_0$ as a distribution. Let $k_n^a(x) = k_n(x+a)$. Then on the one hand we have

$$\phi(k_n^a) = \int k_n^a g \to g(a)$$

for almost every a, by properties of measurable functions; and on the other hand, we have

$$\phi(k_n^a) = \phi(k_n^a)\phi(F) = \phi(k_n^a * F) \to \phi(F_a) = G(a)$$

for all $a \in \mathbb{R}^n$. This shows that $g(a)$ is continuous, and that $\phi(k_n^a) \to g(a)$ for all a.

To prove that $g(a)$ is a homomorphism, just observe that $k_n^a * k_n^b \to \delta_{a+b}$, and thus

$$g(a+b) = \lim \phi(k_n^a * k_n^b) = \lim \phi(k_n^a)\phi(k_n^b) = g(a)g(b).$$

∎

Letting $\widehat{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$ denote the one–point compactification of $\mathbb{R}$, we have:

Corollary 9.21 *The Gelfand transform on $A = (L^1(\mathbb{R}^n), *) \oplus \mathbb{C}\delta_0$ agrees with the Fourier transform, and sends A isomorphically to a subalgebra $\widehat{A}$ of $C(\widehat{\mathbb{R}})$.*

The fact that the radical of A is trivial can be proved using the inverse Fourier transform on $\mathcal{S}'$, which contains $L^1(\mathbb{R}^n)$. Elementary proofs that $\widehat{f} = 0$ implies $f = 0$ can also be given.

Proof of Wiener's Tauberian theorem. One of the motivations for developing the theory of Banach algebras was to give a conceptual proof of Wiener's theorem. We now develop this proof. (See [Nai], §7.4, §15.1–5, §13.1–8, and Theorem 8 of §13.8.) The proof depends on some special properties of the algebra $A_0 = (L^1(\mathbb{R}^n), *)$ which follow from the propositions below.

Proposition 9.22 *If $\widehat{f}$ is a smooth, compactly supported function on $\mathbb{R}^n$, then $f \in L^1(\mathbb{R}^n)$.*

Proof. We have $\widehat{f} \in \mathcal{S} \subset L^1(\mathbb{R}^n)$, and the space of Schwartz functions is invariant under the Fourier transform (and its inverse). ■

Proposition 9.23 *Every $f \in L^1(\mathbb{R}^n)$ is a limit of functions whose Fourier transforms have compact support.*

Proof. Let $\widehat{g}$ be a smooth, compactly supported function such that $\widehat{g}(0) = 1$. Then $g \in L^1(\mathbb{R}^n)$, and we have $\int g = \widehat{g}(0) = 1$. Let $g_r = r^n g(x/r)$. Then $f_r = g_r * f \to f$ as $r \to \infty$, and $\widehat{f_r} = \widehat{f}\widehat{g}_r$ has compact support since $\widehat{g}_r$ does. ■

Ideals in $L^1(\mathbb{R}^n)$. Here are some algebraic consequences of these propositions. Recall that $\operatorname{Spec} A_0 = \mathbb{R}^n$. Let

$$I_0 = \{a \in A_0 \,:\, \widehat{a} \text{ has compact support}\}.$$

Clearly I_0 is an ideal in A_0. The preceding proposition shows it is dense: we have

$$\overline{I_0} = A_0.$$

Next, for any set $E \subset \operatorname{Spec} A_0$ let

$$I(E) = \{a \in A_0 \,:\, \widehat{a}|E = 0\}.$$

This is clearly a closed ideal in A_0.

Proposition 9.24 *If K is compact, then* $\operatorname{Spec} A_0/I(K) = K$.

Proof. The spectrum of $A_0/I(K)$ consists of those t such that $\operatorname{Ker}(\phi_t) \supset I(K)$. Clearly every $t \in K$ has this property. On the other hand, if $t \notin K$, then there exists a compactly supported smooth function with $\widehat{a}(t) = 0$ but $\widehat{a}|K = 1$, so $a \in \operatorname{Ker}(\phi_t)$ but $a \notin I(K)$. ■

Local invertibility. The next step is key. It shows that the image of L^1 under the Fourier transform is 'locally' closed under $\widehat{f} \mapsto 1/\widehat{f}$. More precisely, if $\widehat{f}|K$ is nowhere vanishing, then there is a $\widehat{g}$ such that $\widehat{g}\widehat{f} = 1$ on K.

The proof is similar in spirit to Wiener's result on absolutely convergent Fourier series.

Proposition 9.25 *If K is compact and nonempty, then $A_0/I(K)$ is a Banach algebra with identity.*

Proof. There exists a compactly supported smooth function with $\widehat{a}|K = 1$, and hence $a \in A_0$ projects to the identity in $A_0/I(K)$. ■

Theorem 9.26 *Let $J \subset A_0$ be an ideal that is not contained in* $\operatorname{Ker}(\phi_t)$ *for any $t \in \mathbb{R}^n$. Then J contains I_0, and hence $\overline{J} = A_0$.*

Proof. Let $K \subset \mathbb{R}^n$ be a nonempty compact set. Then J must map surjectively to $A_0/I(K)$, otherwise it would be contained in a maximal ideal of the form $\operatorname{Ker} \phi_t$ with $t \in K$. Thus there exists an $f \in J$ with $\widehat{f}|K = 1$. Consequently, whenever $g \in L^1(\mathbb{R}^n)$ and $\widehat{g}$ is supported in K, we have $\widehat{g} = \widehat{f}\widehat{g}$ and thus $g = f * g \in J$. This shows $I_0 \subset J$. ■

Corollary 9.27 (Wiener's Tauberian theorem) *If $\widehat{f}(t) \neq 0$ for all t, then $L^1 * f$ is dense in $L^1(\mathbb{R}^n)$.*

Regular algebras. We remark that in the general theory of Banach algebras with identity, one can define a new closure operation on $\operatorname{Spec} A$ by $\overline{E} = \operatorname{Spec} A/I(E)$. We say A is *regular* if this new topology agrees with the weak* topology on $\operatorname{Spec} A$.

Regularity is equivalent to the condition that for any closed set $F \subset \operatorname{Spec} A$ and $t \notin F$, there exists an $a \in A$ with $\widehat{a}|F = 1$ and $\widehat{a}(t) = 0$. One can compare this condition to the notion of a T_3 or regular topological space.

The arguments above show that $A = (L^1(\mathbb{R}^n), *) \oplus \mathbb{C}\delta$ is a regular Banach algebra. Note that many Banach algebras are not regular; for example, in the disk algebra $A(\Delta)$, once a function is constant on an open set it must be constant throughout the disk.

Comparison of Wiener's theorems. There is a closed connection between the two results of Wiener above. Both rest on the fact that if $\widehat{f} \in \widehat{A}$ is nowhere zero, then $1/\widehat{f} \in \widehat{A}$.

In the case of Fourier series, this principle is exploited to show that if f has an absolutely convergent Fourier series, then so does $1/f$.

In the case of the Fourier transform, this principle is used to show that if K is compact and $\widehat{f}|K$ has no zeros, then $1/\widehat{f} \in \widehat{A}|K$. This in turn implies that the ideal generated by f contains all g with $\widehat{g}$ supported in K. The proof is completed by showing such g are dense in $L^1(\mathbb{R}^n)$.

Functions in one and several complex variables. We now turn briefly to some interactions with analytic functions.

Convex hulls. We begin with the notion of the *polynomial* convex hull $\widehat{K}$ of a compact set $K \subset \mathbb{C}^n$. This is defined by $z \in \widehat{K}$ if and only if

$$|p(z)| \leq \sup_{w \in K} |p(w)|$$

for all polynomials $p \in \mathbb{C}[z_1, \ldots, z_n]$. For example, if $K = S^1 \subset \mathbb{C}$ then $\widehat{K} = \overline{\Delta}$. More generally, if $K = (S^1)^n \subset \mathbb{C}^n$, then $\widehat{K} = \overline{\Delta}^n$.

Now let $P(K)$ be the closure, in $C(K)$, of the polynomial functions. This is a Banach subalgebra for the sup-norm. In general, the spectrum of a subalgebra can be bigger or smaller than the spectrum of the original algebra. In this case we have:

Theorem 9.28 *The spectrum of $P(K)$ is naturally identified with $\widehat{K}$.*

Proof. First, from the definition of the convex hull, we find there is a natural identification between $P(K)$ and $P(\widehat{K})$. Thus point evaluations on $\widehat{K}$ give multiplicative linear functionals on $P(K)$. Since the coordinate functions separate points, different points give different functionals.

For the converse, given $\phi \in \operatorname{Spec}(P(K))$, let $w \in \mathbb{C}^n$ be the point with coordinates $\phi(z_i)$. Since ϕ is a homomorphisms of algebras, $\phi(p) = p(w)$ for all polynomials $p \in P(K)$. Thus

$$|p(w)| = |\phi(p)| \leq \rho(p) \leq \|p\| = \sup_K |p(z)|$$

for all polynomials, and thus w is in the convex hull $\widehat{K}$. ∎

Invertibility. Here is a somewhat different take on the proof above. If $w \notin \widehat{K}$ then there is a polynomial such that $|p(z)| \leq 1$ on K but $p(w) = \lambda$, $|\lambda| > 1$. But then the usual power series

$$(\lambda - p)^{-1} = \frac{1}{\lambda} \sum_0^\infty \lambda^{-n} p^n$$

exhibits $q = \lambda - p$ as a uniform of polynomials on K, and hence an element of $P(K)$. Thus we have a polynomial q with $q(w) = 0$ and $1/q \in P(K)$, so $w \notin \operatorname{Spec} P(K)$.

The notion of convexity. Note the comparison between the polynomial hull and the convex hull. Indeed the convex hull is what we obtain if we restrict to *linear polynomials* in the definition of $\widehat{K}$. This shows that $\widehat{K}$ lies *inside* the usual convex hull of K.

Open Problem. If K is a union of disjoint balls in $\mathbb{C}^n$, does $\widehat{K} = K$?

A nullstellensatz for the disk algebra. Let $A = P(\overline{\Delta}^n)$; then A is the uniform algebra of all functions continuous on the closed polydisk and holomorphic on its interior.

Theorem 9.29 *If $f_1, \ldots, f_n \in A$ have no common zero, then there exist $g_1, \ldots, g_n \in A$ such that $\sum f_i g_i = 1$.*

Proof. We have seen that the maximal ideals of A coincide with the points of the closed polydisk. By assumption the ideal $I = (f_1, \ldots, f_n)$ is contained in no maximal ideal (note: we are *not* taking the closure of I!) Thus $I = A$ and therefore $1 \in I$. ∎

Spectral radius and subalgebras. Let A be a commutative Banach algebra. We begin with:

Lemma 9.30 *If $x_n \in A^\times$ converges to $x \in \partial A^\times$, then $\|x_n^{-1}\| \to \infty$.*

Proof. Otherwise, we can pass to a subsequence such that $\|x_n^{-1}\|$ is bounded. Then
$$x_n^{-1}(x_n - x) = 1 - x_n^{-1}x \to 0.$$
Since $A^\times$ contains the unit ball about the identity, we would then find that $x_n^{-1}x \in A^\times$ and hence $x \in A^\times$, a contradiction. ■

Now let $B \subset A$ be a closed subalgebra. Given $x \in B$, let $\sigma_A(x)$ and $\sigma_B(x)$ denote its spectrum as an element of A and B respectively, and similarly for $\rho_A(x)$ and $\rho_B(x)$.

In general, the spectrum can change. For example, $B = A(\overline{\Delta})$ is a closed subalgebra of $A = C(S^1)$, and the function $f(z) = z$ satisfies
$$\sigma_A(f) = S^1 \subset \sigma_B(f) = \overline{\Delta}.$$
We will show:

Theorem 9.31 *For $B \subset A$, $\sigma_B(x)$ is obtained from $\sigma_A(x)$ by filling in some bounded holes. More precisely, we have*
$$\sigma_B(x) \supset \sigma_A(x) \quad \textit{and} \quad \partial\sigma_B(x) \subset \partial\sigma_A(x).$$

Corollary 9.32 *The spectral radius of x is independent of the choice of the algebra containing x.*

(Of course we also know this by the spectral radius formula.)

The result above is immediate from the corresponding result on the level of algebras:

Theorem 9.33 *For $B \subset A$, we have*
$$B^\times \subset A^\times \quad \textit{and} \quad \partial B^\times \subset \partial A^\times.$$

Proof. The first inclusion is immediate: if $x \in B$ has an inverse in B, it remains invertible in A. For the second assertion, if $x \in \partial B^\times$ then there exist invertible $x_n \to x$ in B with $\|x_n^{-1}\| \to \infty$; thus the same is true in A, so x cannot be invertible (since inversion is continuous). ■

Holomorphic functional calculus. The algebra $A = C(X)$ is closed under post-composition with arbitrary continuous maps $f : \mathbb{C} \to \mathbb{C}$; i.e. if $a \in A$ then $f(a) \in A$. In fact it is only necessary that f be defined near the spectrum $\sigma(a)$.

For a general Banach algebra, we can clearly form $f(a)$ for any *polynomial* function $f(z)$. We can also make sense of $f(a)$ if $f(z)$ is a *rational function* with poles outside $\sigma(a)$.

Using Cauchy's integral formula and completeness of A, we can then extend this 'functional calculus' to define $f(a)$ when f is holomorphic near $\sigma(a)$. Given a bounded region $\Omega \subset \mathbb{C}$, let $A(\overline{\Omega}) \subset C(\overline{\Omega})$ denote the uniform algebra of functions analytic in Ω and continuous on $\overline{\Omega}$.

Theorem 9.34 *Let a be an element of a Banach algebra a, let $\Omega \supset \sigma(a)$ be a bounded domain in $\mathbb{C}$ containing the spectrum of a. Then there is a unique continuous algebra homomorphism*

$$(A(\overline{\Omega}), z) \to (A, a).$$

Because of this theorem, we can make sense of $\sqrt{a}$, $\log a$, etc. whenever $\sigma(a) \subset (0, \infty)$, or more generally when the spectrum is contained in the right halfplane.

Warning on square-roots. It is not true in general that $f(a)$ makes sense for all $f \in C(\sigma(a))$. For example, if $a(x) = x$ in $A = C^1[0, 1]$, then $\sqrt{a(x)}$ is not in A. However, $\sqrt{r + a} \in A$ for all $r > 0$, by the holomorphic functional calculus.

Integration in a Banach algebra. To define $f(a)$, choose any contour $\gamma \subset \Omega$ surrounding $\sigma(a)$ (this contour may have many components). Then the expression:

$$f(a) = \frac{1}{2\pi i} \int_\gamma \frac{f(\zeta)\, d\zeta}{\zeta - a} \in A$$

is well-defined and independent of the choice of contour.

To see this, first note that

$$(\zeta - a)^{-1} : \Omega - \sigma(a) \to A$$

is an *analytic function*, in the sense that it admits a local power series expansion, or in the sense that $\phi((\zeta - a)^{-1})$ is analytic in ζ for any $\phi \in A^*$. In particular, $(\zeta - a)^{-1}$ is smooth enough that the integral is convergent.

Similarly, if we apply $\phi \in A^*$ to both sides of the equation, then the integral becomes that of an analytic function, and is thus independent of the choice of γ. By the Hahn-Banach theorem, $f(a)$ itself is independent of the choice of γ.

Evaluation of integrals. We now check that the definition of $f(a)$ by algebra agrees with the definition via the Cauchy integral formula.

Theorem 9.35 *If $f(z) = P(z)/Q(z)$ is a rational function with poles outside of $\sigma(a)$, then $f(a) = P(a)/Q(a)$.*

Proof. By partial fractions, it suffices to check the theorem for $f(z) = (\alpha - z)^n$, where $n \in \mathbb{Z}$ and $\alpha \notin \sigma(a)$. Now using the fact that

$$\frac{1}{\zeta - a} = \frac{1}{\alpha - a} + \frac{\alpha - \zeta}{(\zeta - a)(\alpha - a)},$$

we find for any $f \in A(\overline{\Omega})$ we have

$$f(a) = \frac{1}{2\pi i(\alpha - a)} \left(\int_\gamma f(\zeta)\, d\zeta + \int_\gamma \frac{(\alpha - \zeta) f(\zeta)\, d\zeta}{\zeta - a} \right).$$

Since $f(\zeta)$ is analytic on Ω, the first term above is zero. We recognize the second term as $(\alpha - a)^{-1} g(a)$, where $g(z) = (\alpha - z) f(z)$. Thus we have shown:

$$g(a) = (\alpha - a) f(a).$$

Thus it suffices to check that $f(a) = 1$ when $f(z) = 1$.

But this is easy by power series: we can push the contour off to a neighborhood of infinity and conclude that:

$$f(a) = \frac{1}{2\pi i} \int_\gamma \frac{d\zeta}{\zeta - a} = \frac{1}{2\pi i} \int_{|\zeta| = R} \left(1 + \frac{a}{\zeta} + \frac{a^2}{\zeta^2} + \cdots \right) \frac{d\zeta}{\zeta} = 1 + O(1/R).$$

Letting $R \to \infty$ we obtain $f(a) = 1$. ∎

Density of rational functions. Finally we extend the calculus by continuity to holomorphic f. To this end we note:

Theorem 9.36 *Rational functions are dense in $A(\overline{\Omega})$.*

Proof. A dense set of $f \in A(\overline{\Omega})$ are analytic on a neighborhood of $\overline{\Omega}$. Let $\gamma \subset \mathbb{C} - \overline{\Omega}$ be a contour surrounding Ω and contained in the region where f is analytic. Then Cauchy's integral formula

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

expresses f as a uniform limit of finite sums of rational functions with poles on γ. ∎

Holomorphic functional calculus: conclusion. Let

$$M = \sup_\gamma \|(\zeta - a)^{-1}\|.$$

Then we have

$$\|f(a)\| \le (2\pi)^{-1} M \operatorname{length}(\gamma) \sup_\gamma |f(z)|.$$

Since rational functions are dense, the algebra map $f \mapsto f(a)$ extends by continuity to all of $A(\overline{\Omega})$. The requirement that $f(z) = z$ maps to a uniquely determines the values of this homomorphism on the rational functions, and thus it is unique. ∎

***C^** algebras.** We now turn to the study of Banach algebras with the additional structure of an *involution* $a \mapsto a^*$. The axiom for these C^* algebras are meant to abstract the main properties of the adjoint operation on $\mathcal{B}(\mathcal{H})$.

For a general operator $T \in \mathcal{B}(\mathcal{H})$ we have

$$\|T\|^2 = \sup_{|x|=|y|=1} |\langle Tx, Ty\rangle| = \sup |\langle T^*Tx, y\rangle| = \|T^*T\|. \tag{9.2}$$

This equation works in any Hilbert space, and it shows the problem of computing the norm reduces to the case of self-adjoint operators (such as TT^*).

A *C^*-algebra* is a Banach algebra equipped with a map $a \mapsto a^*$ satisfying:

1. $a^{**} = a$;
2. $(ab)^* = b^* a^*$;
3. $(a+b)^* = a^* + b^*$;
4. $(\lambda a)^* = \overline{\lambda} a^*$; and

5. $\|aa^*\| = \|a\|^2$.

Prime examples. The algebra $A = C(X)$ with complex conjugation as $*$ is a C^* algebra. We will shortly see that all commutative C^* algebras have this form. It is immediate that $\|f^2\| = \sup|f^2| = \sup|f|^2 = \|f^*f\|$.

The central non-commutative example is $A = \mathcal{B}(\mathbb{H})$, with adjoint as $*$; see equation (9.2).

Note that if a compact Hausdorff space X is endowed with a measure μ, then we have a natural inclusions:

$$C(X) \subset L^\infty(X,\mu) \subset \mathcal{B}(L^2(X,\mu)).$$

Thus $C(X)$ can often itself be considered as an operator algebra.

Warning. Suppose $h : X \to X$ is a homeomorphism of order 2 ($h(h(x)) = x$). Then if we define $a^*(x) = \overline{a(h(x))}$, we get an involution of $C(X)$ that satisfies all the hypotheses *except the last.*

The last property is thus crucial to the theory of C^* algebras.

We also note that if $e \in A$ is the identity, then $a^* = (ea)^* = a^*e^*$ for all $a \in A$, so $e^* = e$.

Reality of the spectrum. It is a familiar fact that the spectrum of a self–adjoint operator on a finite–dimensional Hilbert space is *real.* This is immediate: if $Tv = \lambda v$, and $\langle v, v\rangle = 1$, then

$$\lambda = \langle Tv, v\rangle = \langle v, T^*v\rangle = \langle v, Tv\rangle = \overline{\lambda}.$$

It is more subtle to prove this result for operators on a general Hilbert space. In fact it holds within any C^* algebra.

Theorem 9.37 *Let A by a C^* algebra, and suppose $a \in A$ satisfies $a^* = a$. Then we have $\sigma(a) \subset \mathbb{R}$.*

Proof. The proof is tricky. Suppose $x+iy \in \sigma(a)$. Let $e \in A$ be the identity element. We then have, for any $Y \in \mathbb{R}$, $x + iy + iY \in \sigma(a + iYe)$. Since $(a + iYe)^* = a - iYe$, we find:

$$\begin{aligned} |x + iy + iY|^2 &= x^2 + (y + Y)^2 \le \rho(a + iYe)^2 \le \|a + iYe\|^2 \\ &= \|(a + iYe)(a - iYe)\| = \|a^2 + Y^2 e\| \le \|a\|^2 + Y^2. \end{aligned}$$

The last part of the computation ratifies our intuition that a and iYe should be ‘orthogonal’. Letting $Y \to \infty$ (or $-\infty$) we find that $y = 0$. ∎

Commutative C^* algebras. Using reality of the spectrum, we can now establish a close connection between C^* algebras and topological spaces in the case where A is commutative.

Theorem 9.38 *Let A be a commutative C^* algebra. Then Gelfand transform gives an isomorphism*

$$A \cong \widehat{A} = C(\operatorname{Spec} A).$$

The involution $*$ is sent to complex conjugation, and

$$\|a\| = \sup |\widehat{a}| = \rho(a)$$

for all $a \in A$.

Proof. Let us first show that $*$ is sent to conjugation. Since any $a \in A$ has the form $a = x + iy$ where x and y are self–adjoint, it suffices to check this assertion for x and y. But for self–adjoint elements, the assertion is immediate because their spectrum is real.

Next, we show that $\rho(a) = \|a\|$ for all $a \in A$. Since self–adjoint elements satisfy $\|a^2\| = \|a\|^2$, this equality holds for them by the spectral radius formula. To handle an arbitrary element of A, just observe that

$$\rho(a)^2 = \rho(aa^*) = \|aa^*\| = \|a\|^2.$$

Here the first equality uses the fact that $\rho(a) = \sup |\widehat{a}|$ and the fact that $*$ goes over to complex conjugation; the second uses the fact that aa^* is self-adjoint; and the third is the basic identity for a C^*-algebra.

In summary, the map $A \to \widehat{A} \subset C(X)$ is norm-preserving, and hence injective. The image is a closed subalgebra of $C(X)$, invariant under conjugation and separating points; thus $\widehat{A} = C(X)$. ∎

Corollary 9.39 *Let A be a C^* algebra generated by a single, self–adjoint element a, in the sense that $\mathbb{C}[a]$ is dense in A. Then* $\operatorname{Spec} A = \sigma(a)$*, and the Gelfand transform*

$$A \cong C(\sigma(a)),$$

sends a to the function $f(\lambda) = \lambda$.

Proof. Consider the continuous map $\pi : \mathrm{Spec}(A) \to \sigma(a)$ given by $\phi \mapsto \phi(a)$. Since $\mathbb{C}[a]$ is dense in A, this map is injective. It is also surjective, since each $\lambda \in \sigma(a)$ gives rise to an ideal $A \cdot (\lambda - a)$ which is contained in some maximal ideal of A. Finally the function $f \in C(\sigma(T))$ determined by a is given by $f(\lambda) = \phi_\lambda(a) = \lambda$. ∎

The category of C^* algebras. The preceding discussion can be formalized to show that the study of *commutative* C^* algebras is *the same* as the study of compact Hausdorff spaces.

Theorem 9.40 *The functor $A \mapsto \mathrm{Spec}\, A$ gives an isomorphism between the category of commutative C^*-algebras and the category of compact Hausdorff spaces, with the arrows reversed.*

The morphisms are algebraic $*$-homomorphisms in the first case (they are automatically continuous because $\rho(a) = \|a\|$), and continuous maps in the second case.

When applied to a map $f : A \to B$, the functor gives the continuous map $\widehat{f} : \mathrm{Spec}\, B \to \mathrm{Spec}\, A$ defined by

$$\widehat{f}(\phi : B \to \mathbb{C}) \;=\; (\phi \circ f : A \to \mathbb{C}).$$

Robustness of the C^*-spectrum. The follow result on spectra holds in both the commutative and non-commutative settings.

Theorem 9.41 *If $B \subset A$ is an inclusion of C^*-algebras, then $\sigma_A(x) = \sigma_B(x)$. In fact $B^\times = A^\times \cap B$.*

Proof. If x is self-adjoint, then $\sigma_A(x) \subset \mathbb{R}$ has no holes to fill, and thus $\sigma_B(x) = \sigma_A(x)$.

Now consider $x \in B$, invertible in A. Then x^*x in also in B and invertible in A; but since it is self-adjoint, it is invertible in B. That is, $yx^*x = 1$ for some $y \in B$; but then $yx^* = x^{-1}$. ∎

Non-commutative topology. Any non-commutative C^* algebra is isomorphic to a closed subalgebra of $\mathcal{B}(\mathcal{H})$. Such algebras are described by Connes as 'non-commutative topological spaces'.

The idea is that when a (possibly strange) space X is defined from a well-behaved space Y by taking the quotient by an equivalence relation, one can build a C^* algebra A using both Y and the equivalence relation. The continuous functions on X will correspond to the center of A.

The simplest case arise when Y consists of n points, identified to a single point comprising X. Then $A = M_n(\mathbb{C})$. Note that as an algebra, A is generated by the diagonal matrices (that is, $C(Y)$) and the matrix of a cyclic permutation (that is, the operator correspond to a dynamical system with orbit space X.)

As a second example, let $R : S^1 \to S^1$ be an irrational rotation of the circle $S^1 = \mathbb{R}/\mathbb{Z}$, where $R(x) = x+\theta$. Then the topological space $X = S^1/\langle R\rangle$ carries no real-valued functions except for the constants, and thus $C(X)$ does not reveal anything about X.

Suppose, however, we consider over X the bundle of Hilbert spaces $H \to X$ with fiber $H_x \ell^2(x+\mathbb{Z}\theta)$ over $[x] \in X$. That is, H_x is the square-summable functions on the orbit of X. Then we can consider the C^*-algebra A of continuous sections of the associated bundle of operator algebras $\mathcal{B}(\mathcal{H})$.

Each continuous function $f : S^1 \to \mathbb{C}$ provides an element of A that acts on H_x by multiplication. That is, T_f sends the basis element $e_n \in H_x$ concentrated at the point $x + n\theta$ to $f(x + n\theta)e_n$.

The irrational rotation R also acts on H_x, as the shift. Thus A contains the non-commutative semidirect product $C(S^1) \ltimes \mathbb{Z}$.

10 Operator algebras and the spectral theorem

In this section we will use the theory of Banach algebras to study operators on Hilbert space, and attempt to classify them up to conjugacy. We will concentrate on self–adjoint operators, and then extend the discussion to unitary and normal operators.

Operators on Hilbert space. Let $\mathcal{H}$ be a complex Hilbert space, and let $\mathcal{B}(\mathcal{H})$ be the algebra of bounded operators $T : \mathcal{H} \to \mathcal{H}$, with the adjoint defined by

$$\langle Tx, y\rangle = \langle x, T^*y\rangle$$

and with the norm defined by

$$\|T\| = \sup_{|x|=1} \|Tx\|.$$

Then $\mathcal{B}(\mathcal{H})$ is a C^*-algebra, as we saw above (equation (9.2); that is,

$$\|T\|^2 = \sup_{|x|=|y|=1} |\langle Tx, Ty\rangle|. = \|T^*T\|.$$

Recall that an operator T is:

- *self–adjoint* if $T^* = T$;
- *unitary* if $T^{-1} = T^*$ (so $\langle Tx, Ty\rangle = \langle x, y\rangle$);
- *normal* if T and T^* commute; and
- a *projection* if $T^2 = T = T^*$.

As an example, consider $A = L^\infty[0,1]$ as an algebra of operators acting on $\mathcal{H} = L^2[0,1]$ by $T_f(g) = fg$. Then T_f is self–adjoint if and only if f is real, and T_f is unitary if and only if f takes values in S^1. Indeed, we have $T_f^* = T_{\overline{f}}$, and hence T_f is normal for all f. Finally T_f is a projection (to $L^2(E)$) if and only if $f = \chi_E$ for some measurable set E.

Self–adjoint case. Now let us focus on the case where T is self–adjoint. The most important feature of such an operator is that if S is a subspace of $\mathcal{H}$ and $T(S) \subset S$, then $T(S^\perp) \subset S^\perp$. When $\mathcal{H}$ is finite–dimensional, one can use this fact to easily prove:

Theorem 10.1 *If $T \in \mathcal{B}(\mathbb{C}^n)$ is self–adjoint, then $\mathbb{C}^n$ has an orthonormal basis of eigenvectors satisfying $Te_i = \lambda_i e_i$, with $\lambda_i \in \mathbb{R}$.*

Corollary 10.2 *A self–adjoint operator on $\mathbb{C}^n$ is uniquely determined, up to isometry, by its spectrum $\sigma(T) = \{\lambda_1, \ldots, \lambda_n\}$ and by the multiplicities of its eigenvalues.*

We aim to formulate and prove a suitable generalization of this result to the infinite–dimensional case.

Positivity. Let T be a self–adjoint operator. It then follows easily from the C^* property that

$$\sigma(T) \subset \mathbb{R} \quad \text{and} \quad \|T\| = \rho(T)$$

Let $[\rho_-(T), \rho_+(T)]$ denote the smallest interval containing the spectrum of T. Then $\rho(T) = \max(|\rho_-(T)|, |\rho_+(T)|)$. We wish to relate these values more directly to the operator T.

The argument will rely implicitly on the idea of a *positive* self–adjoint operator; this is one such that $\langle Tx, x\rangle \geq 0$ for all x. When T is positive, it defines a new inner product by $\langle Tx, y\rangle$ on $\mathcal{H}/\operatorname{Ker}(T)$ and we will use Cauchy–Schwarz on this inner product.

The following result is readily verified in the finite–dimensional case and in the case of $T_f \in L^\infty[0,1]$.

Theorem 10.3 *If T is self–adjoint, then $\rho_+(T) = \sup_{|x|=1}\langle Tx, x\rangle$.*

The corresponding formula for $\rho_-(T)$ follows, and hence:

Corollary 10.4 *We have $\|T\| = \rho(T) = \sup_{|x|=1} |\langle Tx, x\rangle|$.*

Corollary 10.5 *A self–adjoint operator T is positive if and only if its spectrum lies in $[0,\infty)$.*

Proof of the Theorem. Let $M(T) = \sup_{|x|=1}\langle Tx, x\rangle$. We wish to show $M(T) = \rho_+(T)$. Clearly for λ real, both sides translate by λ if we replace T by $T + \lambda$. Thus it suffices to treat the case where $\rho_+(T) = \|T\|$ and where $\langle Tx, x\rangle \geq \langle x, x\rangle$ for all x, since both properties can be achieved by adding a large positive constant to T.

Under these assumptions it is clear that $M(T) \leq \|T\| = \rho_+(T)$. Also, $\langle Tx, y\rangle$ defines a Hermitian inner product on $\mathcal{H}$. By the Cauchy–Schwarz inequality, we then have for all x, y,

$$|\langle Tx, y\rangle|^2 \leq \langle Tx, x\rangle\langle Ty, y\rangle.$$

Taking the sup over all x, y of norm 1, we find $\|T\|^2 \leq M(T)^2$, and hence $M(T) = \|T\| = \rho_+(T)$. ∎

Remark. Another approach to Corollary 10.4 involves writing

$$\operatorname{Re}\langle Tx, y\rangle = \frac{1}{4}\left(\langle T(x+y), x+y\rangle - \langle T(x-y), x-y\rangle\right).$$

Compact operators. Recall that T is *compact* if $\overline{T(B)}$ is compact, where $B \subset \mathcal{H}$ is the unit ball. Here is the spectral theorem for such operators.

Theorem 10.6 *Let T be a compact self-adjoint operator. Then T has a finite or countable set of orthonormal eigenvectors*

$$Te_i = \lambda_i e_i$$

with nonzero eigenvalues $\lambda_i \to 0$, and $\mathrm{Ker}(T) = \{x : \langle x, e_i \rangle = 0 \ \forall i\}$.

The list of eigenvalues, with multiplicities, is thus again a complete invariant of T. The proof we will rely on the following general facts.

Lemma 10.7 *If $T \in \mathcal{B}(\mathcal{H})$ is a compact operator and $x_n \to y$ in the weak topology, then $T(x_n) \to T(y)$ in the norm topology.*

Lemma 10.8 *A compact self–adjoint operator has $\rho_+(T)$ as an eigenvalue.*

Proof. Adding a constant to T if necessary, we may assume that $\|T\| = \rho_+(T)$. By Theorem 10.3, we have a sequence of unit vectors $x_n \in \mathcal{H}$ with $\langle Tx_n, x_n \rangle \to \|T\|$. Pass to a subsequence such that $x_n \to y$ weakly in L^2; then $T(x_n) \to T(y)$ strongly, by the Lemma above. We then have:

$$\langle Ty, y \rangle = \langle Ty - Tx_n, y \rangle + \langle Tx_n, y - x_n \rangle + \langle Tx_n, x_n \rangle.$$

On the right, the first term tends to zero since $T(x_n) \to T(y)$ in norm. The second term is the same as $\langle x_n, T(y - x_n) \rangle$, so it tends to zero for the same reason. The last term converges to $\|T\|$, so we have found y with $|y| \le 1$ and

$$\langle Ty, y \rangle = \|T\| \le \|y\| \cdot \|Ty\|.$$

Since this gives follows the case of equality in Cauchy–Schwarz, we have $Ty = \lambda y$ for some λ. Then clearly $\lambda = \|T\|$. ∎

Proof of Theorem 10.6. Using the preceding Lemma we can construct by induction a sequence of eigenvectors as required, with $|\lambda_1| = \|T\|$ and $|\lambda_i|$ equal to the norm of T restricted to the perp of $(e_1, \ldots, e_{i-1})$. Since T is compact, the sequence of vectors $T(e_i) = \lambda_i e_i$ must have a convergent subsequence; since the e_i are orthogonal, this implies that $\lambda_i \to 0$. ∎

Continuous functional calculus. For a more sophisticated viewpoint on the spectral theorem, one can use the continuous functional calculus. Since $\mathcal{B}(\mathcal{H})$ is a C^* algebra, so is any closed sub $*$-algebra, and hence Corollary 9.39 immediately yields:

Theorem 10.9 *Let T be self-adjoint. Then there is a canonical isomorphism*

$$C(\sigma(T)) \to A \subset \mathcal{B}(\mathcal{H}),$$

where A is the smallest closed algebra containing T. This isomorphism sends $P(\lambda)$ to $P(T)$ for any polynomial P.

Projections. Now let T be a self–adjoint operator such that $\sigma(T)$ consists of the origin together with a sequence of points $\lambda_i \to 0$. Let $f_i \in C(\sigma(T))$ be the continuous function equal to 1 at λ_i and 0 elsewhere. Then clearly $z = \sum \lambda_i f_i(z)$ in $C(\sigma(T))$. On the other hand, $P_i = f_i(T)$ is clearly a projection, and by the functional calculus,

$$T = \sum \lambda_i P_i.$$

In other words, the spectral theorem expresses T as a weighted sum of commuting projections. Since $z f_i(z) = \lambda_i f_i(z)$, we have $T(P_i(x)) = \lambda_i P_i(x)$ for all x. Thus P_i is nothing but projection onto the λ_i eigenspace of T.

Example of a compact operator. Let I be the operator on $\mathcal{H} = \mathcal{B}(L^2[0,1])$ defined by

$$(If)(x) = \int_0^x f(t)dt.$$

Then If is Hölder continuous of exponent $1/2$; indeed, by Cauchy-Schwarz we have

$$|(If)(x) - (If)(y)|^2 \le \|f\|_2^2 |x-y|.$$

Thus I is a compact operator.

Let π be projection onto the functions of mean zero. Then $T = i\pi I\pi$ is a compact self-adjoint operator, and the spectral decomposition of T is given by $e_n(x) = \exp(2\pi i n x)$, $n \in \mathbb{Z} - 0$, and $\lambda_n = 1/(2\pi n)$.

Topologies on $\boldsymbol{\mathcal{B}(\mathcal{H})}$. To approach general self–adjoint operators, it will be useful to distinguish 3 different notions of convergence in $\mathcal{B}(\mathcal{H})$.

- We say $T_n \to T$ *in norm* if the operator norms $\|T_n - T\| \to 0$.

- We say $T_n \to T$ *strongly* if $T_n(x) \to T(x)$ for all $x \in \mathcal{H}$.
- We say $T_n \to T$ *weakly* if for all $x, y \in \mathcal{H}$, we have

$$\langle T_n x, y\rangle \to \langle Tx, y\rangle.$$

Note that the strong topology is the same as pointwise convergence of the map T_n, as functions on $\mathcal{H}$, while the weak topology comes from convergence of the 'matrix coefficients' of T.

By suitable applications of the uniform boundedness principle, one can show that if $T_n \to T$ strongly or weakly, then $\|T_n\|$ is bounded.

The strong and weak topologies on $\mathcal{B}(\mathcal{H})$ are generally not metrizable; however, if $\mathcal{H}$ is separable, they are metrizable when restricted to *bounded sets*.

Example. Let us regard $A = L^\infty[0,1]$ as an algebra of operators on $\mathcal{H} = L^2[0,1]$. Then the operator norm and the L^∞ norm agree on A. The three topologies behave as follows:

1. $f_n \to f$ in norm, as operators, if and only if $\|f_n - f\|_\infty \to 0$;
2. $f_n \to f$ strongly, as operators, if and only if $\|f_n\|_\infty$ is bounded and $f_n \to f$ in measure; and
3. $f_n \to f$ weakly, as operators, if and only if $f_n \to f$ in the weak* topology.

For the last assertion we regard $L^\infty[0,1]$ as $L^1[0,1]^*$.

Strong convergence means for all $\phi \in L^2[0,1]$ we have

$$\|(f_n - f)\phi\|_2^2 = \int |f_n - f|^2 \cdot |\phi|^2 \to 0.$$

The function $f_n(x) = \exp(inx)$ do not converge in either the norm or strong topologies, but it converges to 0 in the weak topology.

Note that $L^\infty[0,1]$ itself is weakly closed in $\mathcal{B}(\mathcal{H})$. Moreover, we have:

Proposition 10.10 *The closure of $C[0,1]$ in the strong or weak operator topologies is all of $L^\infty[0,1]$.*

In fact, the weak closure of any *convex subset* of $\mathcal{B}(\mathcal{H})$ agrees with its strong closure. This is because a linear function on $\mathcal{B}(\mathcal{H})$ is strongly continuous if and only if it is weakly continuous.

von Neumann algebras. For the continuous function calculus, and the preceding special cases of the spectral theorem, it was enough to work in the C^* algebra $A \cong C(\sigma(T))$ generated by T. Here A is the *norm closure* of the algebra generated by T and T^*.

For the general spectral theorem, we will need the *Borel functional calculus.* The new operators $f(T)$, where f is a *Borel* function on $\sigma(T)$, will reside in the *strong closure* M of A.

This strong closure is an example of a *von Neumann algebra.* Just as commutative C^* algebras encode topology, commutative von Neumann algebras encode measure theory.

One can define von Neumann algebras in three different ways, which agree by the following result.

Theorem 10.11 *Let $M \subset \mathcal{B}(\mathcal{H})$ be an algebra such that $M = M^*$. Then the following conditions are equivalent.*

- *M is weakly closed.*
- *M is strongly closed.*
- *M agrees with its double commutant, M''.*

Here M' is the set of operators in $\mathcal{B}(\mathcal{H})$ that commute with every element of M.

We will not need this result, rather we wish to draw attention to the commutative von Neumann algebra M generated by T as the natural setting for the discussion of the spectral theorem for T.

The spectral theorem. We now turn to the spectral theorem for a general self–adjoint operator.

In the infinite-dimensional setting, T may have *no eigenvectors.* (Consider, for example, $T(\phi) = t\phi(t)$ acting on $L^2[0,1]$.)

Instead, the decomposition into projections $T = \sum \lambda_i \pi_i$ available in the case of compact operators is replaced by an integral, and the eigenspaces of T for discrete eigenvalues are replaced approximate eigenspaces $\mathcal{H}_{[a,b]}$, where $\|T - \lambda\| \le |a - b|$ for all $\lambda \in [a,b]$. (In the example above, $H_{[a,b]} = L^2[a,b] \subset L^2[0,1]$.)

Here is the general statement.

Theorem 10.12 (Spectral Theorem) *Let $T \in \mathcal{B}(\mathcal{H})$ be a self-adjoint operator on a separable Hilbert space $\mathcal{H}$. Then:*

(A) The operator T can be expressed as a direct integral

$$T = \int_{\mathbb{R}} \lambda \, d\pi_\lambda,$$

where π_λ is an increasing family of projections.

(B) The continuous functional calculus $C(\sigma(T)) \to \mathcal{B}(\mathcal{H})$ extends to the bounded Borel functions $L_b^\infty(\sigma(T))$ in a unique way such that if $f_n \to f$ pointwise, and $\sup \|f_n\|_\infty < \infty$, then the corresponding operators satisfy $f_n(T) \to f(T)$ strongly.

(C) There is a finite or countable collection of probability measures μ_i on $\sigma(T)$ and an isomorphism $\mathcal{H} \to \oplus L^2(\sigma(T), \mu_i)$ such that the action of $C(\sigma(T))$ is realized by multiplication. In particular, $T(f(\lambda)) = \lambda f(\lambda)$.

Parts (A) and (C) provide a 'continuous diagonalization' of T. In part (A), we have $\pi_\lambda = 0$ for $\lambda < \rho_-(T)$ and $\pi_\lambda = I$ for $\lambda > \rho_+(T)$.

Spectral measures. The crucial and beautiful idea behind the proof of the spectral theorem is that each unit vector $x \in \mathcal{H}$ defines a probability measure on $\sigma(T)$. Considered as a linear functional on $C(\sigma(T))$, this probability measure is given by

$$\mu_x(f) = \int_{\sigma(T)} f(\lambda) \, d\mu_x(\lambda) = \langle f(T)x, x \rangle.$$

Here $f(T)$ is defined using the continuous functional calculus. Note that if $f(\lambda) \geq 0$ then $f = g^2$ and hence $\langle f(T)x, x\rangle = \langle g(T)x, g(T)x\rangle \geq 0$, so μ_x is indeed a *positive* linearly functional (and hence a measure).

Since $\int \mu_x = \langle x, x \rangle = 1$, we have a probability measure.

Quantum interpretation! The measure μ_x is exactly the distribution of the observable T in the state x.

Example: if $\mathcal{H} = L^[0, 1]$ (particles in a box) and $T(\phi) = x\phi$ is the position operator, then $\sigma(T) = [0, 1]$ and the measure μ_ϕ is given by $|\phi(x)|^2 \, dx$.

Proof of (C). Let A be the smallest closed algebra in $\mathcal{H}(\mathcal{H})$ containing T and the identity. Then A is a C^* algebra; in fact we have $A \cong C(\sigma(T))$.

Let $x \in \mathcal{H}$ be a unit vector. We can then consider the subspace $A \cdot x \subset \mathcal{H}$ and define a map

$$\pi : A \cdot x \to L^2(\sigma(T), \mu_x)$$

by $\pi(f(T)x) = f(\lambda)$. This map is well-defined and *preserves norms*: in fact,

$$\|f(T)x\|^2 = \langle f(T)x, f(T)x \rangle = \langle f(T)^* f(T)x, x \rangle = \int |f(\lambda)|^2 \, d\mu_x.$$

Thus we obtain an *isometry* between the closure $\mathcal{H}_x = \overline{A \cdot x}$ and $L^2(\sigma(T), \mu_x)$ (using the fact that continuous functions are dense in L^2).

Clearly the isometry π sends the action of T on $\mathcal{H}_x$ to the action of multiplication by λ on $L^2(\sigma(T), \mu_x)$. Thus if $\mathcal{H}_x = \mathcal{H}$, the proof of (C) is complete.

Otherwise, we pass to the complement $\mathcal{H}^\perp$, which is also preserved by T, and continue. In this way we obtain a sequence of orthogonal subspaces $\mathcal{H}_i \subset \mathcal{H}$, and measure μ_i, such that the action of T goes over to multiplication by λ on $\mathcal{H}_i \cong L^2(\sigma(T), \mu_i)$.

Since $\mathcal{H}$ is separable, there is an orthonormal basis $\langle x_1, x_2, \ldots \rangle$ spanning $\mathcal{H}$. We start with $x = x_1$ and then in the construction of $\mathcal{H}_{i+1}$ choose $x = x_n$, for the first n not already in the span of $\mathcal{H}_1, \ldots, \mathcal{H}_i$. Then in the end we obtain a finite or countable direction sum, $\mathcal{H} = \oplus \mathcal{H}_i$, completing the proof. ∎

Measure classes and multiplicities. A pair of finite, nonzero measures μ_1, μ_2 on $\mathbb{R}$ are in the same *measure class* if they have the same sets of measure zero. In this case the Hilbert spaces $\mathcal{H}_i = L^2(\mathbb{R}, \mu_i)$, $i = 1, 2$ are naturally isomorphic, via the map $S : \mathcal{H}_1 \to \mathcal{H}_2$ defined by

$$S(\phi) = \left(\frac{d\mu_1}{d\mu_2} \right)^{1/2} \phi.$$

This norm–preserving isomorphism respects the action of multiplication by bounded functions. Thus the model for $T \in \mathcal{B}(\mathcal{H})$ provided by (C) above is not unique.

Instead, we can attach to T three invariants:

- Its spectrum $\sigma(T) \subset \mathbb{R}$;
- The *spectral measure* (class) $[\mu] = [\sum 2^{-i} \mu_i]$ on $\sigma(T)$; and

- The *multiplicity function* $m(x) = \sum \chi_i \in L^\infty(\sigma(T), \mu)$, where $\chi_i(x) = 0$ or 1 and μ_i is in the same measure class as $\chi_i \mu$.

Provided we regard $[\mu]$ as a *measure class*, the data $(\sigma(T), [\mu], m)$ is *canonically* determined by T, and it suffices to reconstruct T completely, up to conjugacy in $\mathcal{B}(\mathcal{H})$.

Hilbert space bundles. Here is explicit reconstruction of T acting on $\mathcal{H}$ from $(\sigma(T), [\mu], m)$.

First, note that one can naturally attach a Hilbert to a *measure class* $[\mu]$. This can done by considering *half-densities* instead of functions; that is, objects whose square is a measure. Then we let $L^2([\mu])$ denote the space of all half-densities ϕ such that $[|\phi|^2] \ll [\mu]$ (that is, $|\phi|^2$ is absolutely continuous with respect to μ), with $\langle \phi, \psi \rangle = \int \phi \overline{\psi}$.

Now let $\mathcal{H}_m \to \sigma(T)$ be the bundle of Hilbert spaces with dimension $m(\lambda)$ over λ. (We can write $\mathcal{H}_m$ as a union of $E_n \times \mathbb{C}^n$ over the sets $E_n = m^{-1}(n)$, with the obvious convention for $n = \infty$.) Let $L^2([\mu], m)$ be the space of half-densities $\phi(\lambda)$ with values in $H(\lambda)$, and inner product

$$\langle \phi, \psi \rangle = \int_{\sigma(T)} \langle \phi(\lambda), \psi(\lambda) \rangle_{H(\lambda)}.$$

(Here the inner product of two sections ϕ and ψ gives a measure on $\sigma(T)$.) Then we have a natural action of $L^\infty(\sigma(T), [\mu])$ on $L^2([\mu], m)$ by $T_f(\phi) = f\phi$. The operator $T_\lambda(\phi) = \lambda\phi(\lambda)$ then provides a reconstruction of the operator T. Indeed, part (C) of the spectral theorem easily implies:

Theorem 10.13 *The data $(\sigma(T), \mu, m)$ determines T up to an isometry of $\mathcal{H}$; in fact, there is an isometry of $\mathcal{H}$ to $L^2([\mu], m)$ sending T to multiplication by λ.*

Examples.

1. Let $T \in \mathcal{B}(\mathbb{C}^n)$ be self-adjoint, with $\sigma(T) = \{\lambda_1, \ldots, \lambda_n\}$. Then $[\mu]$ is represented by a sum of δ masses on the eigenvalues of T, and $m(\lambda_i) = \dim \operatorname{Ker}(T - \lambda_i)$.

   Note that in general T requires several μ_i in part (C), even when $\mathcal{H}$ is finite dimensional. In the extreme case where $T = \lambda_1 I$, what is required is $\mu_1, \ldots, \mu_n$ all concentrated at λ_1. The choice of measures is the same as the choice of an orthonormal basis for $\mathcal{H} = \mathbb{C}^n$. Thus the map from $\mathcal{H}$ to $\oplus L^2(\sigma(T), \mu_i)$ is not unique.

2. Suppose $f : [0,1] \to [a,b]$ is a homeomorphism. Then $f \in L^\infty[0,1]$ determines a self–adjoint multiplication T_f operator on $\mathcal{H} = L^2[0,1]$. The spectrum $\sigma(T) = [a,b]$ carries a natural measure $\mu = f_*(dx)$, and the map $\phi \mapsto \phi \circ f^{-1}$ gives a change of coordinates isomorphism

$$\mathcal{H} = L^2[0,1] \to L^2[a,b],$$

   sending T_f to multiplication by λ. Thus $[\mu]$ is the spectral measure for T, and $m(\lambda) = 1$. Note that μ need not be absolutely continuous.

3. Now suppose $f(x) = x^2$ on $\mathcal{H} = L^2[-1,1]$. In this case $T = T_f$ satisfies $(\sigma(T), \mu) = ([0,1], d\lambda)$ but $m(\lambda) = 2$. In fact we can write $\mathcal{H} = L^2[-1,0] \oplus L^2[0,1]$, and then T_f behaves like the preceding result on each factor.

4. If $f : [0,1] \to \mathbb{R}$ is a smooth Morse function, then f has finite fibers and we have a similar result with $(\sigma(f), \mu) = ([a,b], dx)$ and $m(\lambda) = |f^{-1}(\lambda)|$.

5. If $f : [0,1] \to [0,1]$ is the Cantor function, then $\sigma(f) = [0,1]$ and $[\mu]$ consists of δ-masses at the dyadic rationals in $[0,1]$, each with $m(\lambda) = \infty$. This corresponds to the decomposition $L^2[0,1] = \oplus L^2(I_i)$, where (I_i) are the complementary intervals to the Cantor set.

6. Suppose T is a self–adjoint operator with a *cyclic* unit vector $x \in \mathcal{H}$, i.e. a vector such that $\overline{\mathbb{C}[T] \cdot x} = \mathcal{H}$. Then, as we will see below, there is a measure μ and an isomorphism

$$\mathcal{H} \cong L^2(\sigma(T), \mu)$$

   such that T is sent to multiplication by λ, and the cyclic vector corresponds to $f(\lambda) = 1$.

   In fact T has a cyclic vector if and only if $m = 1$ almost everywhere.

Proof of (B): The Borel calculus. It is now easy to extend the continuous functional calculus to *pointwise* limits of continuous functions. Namely we consider L^2 of a general measure space, $\mathcal{H} = L^2(X, \mu)$. Then we have a commutative operator algebra $B = L^\infty(X, \mu)$ in $\mathcal{B}(\mathcal{H})$. We say $T_n \to T$ *strongly* in $\mathcal{B}(\mathcal{H})$ if $T_n f \to T f$ for each $f \in \mathcal{H}$.

Theorem 10.14 *Suppose $h_n \in L^\infty(X, \mu)$ is a bounded sequence and $h_n \to h$ pointwise. Then $T_n \to T$ strongly, where $T_n(f) = h_n f$.*

Proof. Assuming $\|h_n\|_\infty \le M$, we have

$$\|(T_n - T)f\|^2 = \int |h_n - h|^2 |f|^2 \, d\mu \le \int M|f|^2 \, d\mu < \infty,$$

so $T_n f \to Tf$ by the dominated convergence theorem. ∎

In the case of the spectral theorem, we don't know the spectral measures μ, but in any case the *Borel* measurable functions are in $L^\infty(\mathbb{R}, \mu)$ for any measure μ. Furthermore, $C(\sigma(T))$ is *dense* in $L^\infty_b(\sigma(T))$ in the 'strong topology', that is with respect to the topology of bounded pointwise convergence. (To see this, first note that χ_U is in the closure for any open set U; then that the E with χ_E in the closure form a σ-algebra, so we get all Borel sets; and finally that any L^∞ function is a strong limit of simple functions.)

Thus the functional calculus extends to $L^\infty_b(\sigma(T))$. ∎

Proof of (A): Integration with respect to a projection-valued measure. Let $\pi_t \in \mathcal{B}(\mathcal{H})$ be the image of $f_t(\lambda) = \chi_{(-\infty,t]}(\lambda)$ under the Borel functional calculus. Since $f_t(\lambda)$ is real and $f_t^2 = f_t$, we see f_t is a **projection** onto a closed subspace

$$H_t = \pi_t(\mathcal{H}).$$

Note that if $\sigma(T) \subset [a, b]$, then $H_t = \{0\}$ for $t < a$ and $H_t = \mathcal{H}$ for $t > b$.

Similarly, $\pi_{(s,t]} = (1 - \pi_s)\pi_t$ is projection onto the subspace

$$H_{(s,t]} = H_t \ominus H_s.$$

Heuristically, $H_{(s,t]}$ is the subspace of $\mathcal{H}$ on which T has eigenvalues in the interval $(s, t]$.

Using the Borel functional calculus, we can more generally send any Borel set E to the projection π_E which is the image of χ_E. This function from Borel sets to measures is a 'projection-valued measure', in the sense that for any sequence of disjoint Borel sets E_i we have

$$\pi_{\bigcup E_i} = \sum \pi_{E_i},$$

where the sum converges in the strong topology.

We can now interpret the integral representation of T in various ways; the simplest being:

$$T = \int \lambda \, d\pi_\lambda = \lim \sum_1^n \lambda_i \pi_{(\lambda_i, \lambda_i+1]}$$

over finer and finer subdivision $\lambda_1 < \lambda_2 < \cdots < \lambda_n$ of an interval $[a, b]$ containing $\sigma(T)$.

By the Borel functional calculus, we see the limit of these sums converges in *norm* to T. Indeed,

$$\left| \lambda - \sum_1^n \lambda_i \chi_{(\lambda_i, \lambda_i+1]} \right|_\infty \to 0,$$

and T is the image of λ while the approximations to the integral are the images of weighted sums of indicator functions above. ∎

Spectral theorem for unitary and normal operators. An operator is *unitary* if $T^{-1} = T^*$; it is *normal* if $[T, T^*] = 0$. Clearly self-adjoint and unitary operators are special cases of normal operators.

The spectral theorem extends in a straightforward way to normal operators, since T and T^* generate a *commutative* C^*-subalgebra $A \subset \mathcal{B}(\mathcal{H})$.

Theorem 10.15 *Let $\mathcal{H}$ be a separable Hilbert space and let $T \in \mathcal{B}(\mathcal{H})$ be a normal operator. Then:*

(A) There is a projection-valued Borel measure such that $T = \int_\sigma (T) \lambda \, d\pi_\lambda$;

(B) We have a Borel functional calculus $L_b^\infty(\sigma(T)) \to \mathcal{B}(\mathcal{H})$, and

(C) There are spectral measures μ_i on $\sigma(T)$ and an isomorphism $H \to \oplus L^2(\sigma(T), \mu_i$ sending the action of T on $\mathcal{H}$ to the operator $T(f(\lambda)) = \lambda f(\lambda)$.

Complete reducibility. The proof of the spectral theorem for a normal operator proceeds as before, with a little point to be addressed at the inductive step.

Namely, suppose we have a closed A-invariant subspace $\mathcal{H}_0 \subset \mathcal{H}$. We claim $\mathcal{H}_0^\perp$ is also A-invariant. Indeed, given $x \in \mathcal{H}_0^\perp$ we have $\langle f(T)y, x \rangle = 0$

for all $y \in \mathcal{H}_0$ and $f(T) \in A$. Thus $\langle y, f(T)^* x\rangle = 0$ for all $y \in \mathcal{H}_0$, and thus $f(T)^* x \in \mathcal{H}_0^\perp$. Since A is a $*$-algebra, this shows $A\mathcal{H}_0^\perp = \mathcal{H}_0^\perp$.

Application: Unitary representations. The space of *irreducible* unitary representations of $G = \mathbb{Z}$ is canonically identified with $\widehat{G} = S^1$. Thus the spectral theorem for a unitary operator can be interpreted as saying that any unitary action of G on $\mathcal{H}$ can decomposed into irreducibles by writing

$$\mathcal{H} = \int_{\widehat{G}} \mathbb{C}^{m(\lambda)}\, d\mu(\lambda).$$

Extension: Unbounded self–adjoint operators. Differential operators provide a major source of self-adjoint operators, motivating the general theory.

On $\mathbb{R}^n$ the operators $D_j = i\, d/dx_j$ are commuting and self-adjoint, by integration by parts. Note that the product of any two *commuting* self-adjoint operators is self-adjoint, because $(ST)^* = (TS)^* = S^*T^* = ST$. Thus *any real polynomial* $P(D)$ in the operators D_j is self-adjoint.

These operators are simultaneously diagonalized by the Fourier transform: they act by multiplication by $P(t)$ on $L^2(\mathbb{R}^n)$. They are, however, *unbounded* operators. The spectrum is $P(\mathbb{R}^n)$.

For example, the Laplacian $\Delta = -\sum D_j^2$ has spectrum $(-\infty, 0]$.

It is also interesting to consider these operators on the torus $\mathbb{R}^n/\mathbb{Z}^n$, where the spectrum becomes discrete. Finally the eigenfunction decomposition of even an operator as simple as $T = d^2/dx^2 + f(x)$ on S^1 is nontrivial (difficult to give explicitly) but covered by the general (unbounded) theory.

11 Ergodic theory: a brief introduction

We now turn to an important source of *unitary operators*, namely *measure-preserving* dynamical systems.

References for this section: [CFS], [Me].

1. Let (X, μ) be a measure space, and let $T : X \to X$ be a measure-preserving transformation: that is, $\mu(T^{-1}(E)) = E$.

   The main problem of *ergodic theory* is the classification of such mappings T up to *isomorphism*, i. e. measure-preserving conjugacy.

   We say T is *ergodic* if for any splitting of X into T-invariant sets A and B, one of them has zero measure.

2. Examples:

   (a) $T : S^1 \to S^1$, a rotation of the circle with linear measure.

   (b) $T : S^1 \to S^1$ with $T(z) = z^n$. This is a measure-preserving map which is *not* bijective.

   (c) $T : T^2 \to T^2$, a toral automorphism specified by a matrix $A \in GL_2(\mathbb{Z})$. Since $|\det A| = 1$, T preserves area measure.

   (d) $T : \Sigma_n \to \Sigma_n$, the n-shift, with the measure corresponding to independent event with probabilities $p_1 + \ldots + p_n = 1$.

   (e) The solar system: $X = (T\mathbb{R}^3)^N$, T is the 1-year map for the Hamiltonian flow generated by

$$H(p,q) = \sum m_i q_i^2 + \sum m_i m_j |p_i - p_j|^{-1}.$$

3. *Ergodicity.* Theorem. An irrational rotation of S^1 is ergodic. Proof. Consider any T-invariant sets A and B of positive measure. Since the orbits of T are dense, we can move a point of density of A close to a point of density of B, contradiction. ■

4. *Spectral ergodic theory.* The dynamical system $T : X \to X$ induces an operator $U : L^2(X) \to L^2(X)$ by $U(f) = f \circ T$. This operator preserves norm, because T preserves measure. Also U is *unitary* when T is invertible.

   For almost all X we want to consider, $\mathcal{H} = L^2(X, \mu)$ is an infinite-dimensional separable Hilbert space. Thus all the Hilbert spaces are isomorphic.

   We can then apply the spectral theory to U. We say T_1 and T_2 are *spectrally equivalent* if their unitary operators U_1 and U_2 are isomorphic.

5. *Spectral classification of rotations.* Let $T(z) = \lambda z$ be a rotations of $S^1 \subset \mathbb{C}$. Then $L^2(S^1)$ decomposes as a sum of eigenspaces $\langle z^n \rangle$ with eigenvalues λ^n. Thus the spectral measure class is $\mu = \sum \delta_{\lambda^n}$.

   If T has finite order n, then μ has finite support and each eigenvalue has *infinite* multiplicity. Clearly μ is then supported on the nth roots of unity, and the operator U is determined by n.

Indeed, any two T of the same order are related by a measurable (but not continuous) automorphism of S^1.

Now assume T is an irrational rotation. Since the support of μ is a cyclic subgroup of S^1, we see $\lambda^{\pm 1}$ is uniquely determined by the unitary operator associated to T. In particular, a pair of irrational rotations are spectrally isomorphic $\iff$ $\lambda_1 = \lambda_2$ or λ_2^{-1}, in which case they are spatially isomorphic (use complex conjugation if necessary).

6. *Ergodicity is detected by the spectrum.* Indeed, T is ergodic iff $m(1) = 1$. This gives another proof that an irrational rotation is ergodic.

   (Of course T ergodic does not imply U is irreducible; indeed all irreducible representations of abelian groups are 1-dimensional.)

7. *von Neumann's Ergodic Theorem.* Suppose T is ergodic. Then for any $g \in L^2(X, \mu)$ we have

$$\frac{g + Ug + \cdots U^{n-1}g}{n} \to \left(\int g\right)$$

   in $L^2(X)$.

   Proof. By applying the spectral theorem to U, we see $(1 + U + \cdots + U^{n-1})/n$ is the same as $f(U)$ where

$$f(\lambda) = \frac{1 + \lambda + \ldots + \lambda^{n-1}}{n} = \frac{1 - \lambda^n}{1 - \lambda}.$$

   Now $f(\lambda) \to \chi_1$ pointwise on S^1, so be the Borel functional calculus $f(U) \to \pi$ strongly, where π is the projection onto the part of the spectrum over $\lambda = 1$. On this subspace, U acts by the identity, so by ergodicity π is projection onto the constant functions. ∎

   Remark. The deeper Birkhoff-Khintchine ergodic theorem says that for $f \in L^1(X)$, the averages of f along the orbit of x under T converge to $\int f$ for almost every x.

   This pointwise ergodic theorem supports Nietzsche's philosophy of eternal recurrence. Indeed, the universe itself (e.g. from the Newtonian point of view) is a Hamiltonian dynamical system, and thus measure-preserving, and presumably ergodic (the Boltzmann hypothesis).

8. *Lebesgue spectrum.* We say $U : H \to H$ has *Lebesgue spectrum* of multiplicity n if it is conjugate to multiplication by z on $\oplus_1^n L^2(S^1, d\theta)$; equivalently, if H decomposes as a direct integral over S^1 with Lebesgue measure and multiplicity n. (Infinite multiplicity is permitted).

9. *The shift operator.* Consider the shift $T : \mathbb{Z} \to \mathbb{Z}$ preserving counting measure. Then the corresponding U on $\ell^2(\mathbb{Z})$ has Lebesgue spectrum. Indeed, the Fourier transform gives an isomorphism $\ell^2(\mathbb{Z}) \cong L^2(S^1, d\theta)$ sending U to multiplication by z.

10. *Open problem.* An automorphism T of a measure space is a *Lebesgue automorphism* if the associated unitary operator has Lebesgue spectrum.

    It is remarkable that *there is no known Lebesgue automorphism of finite multiplicity* [Me, p.146].

11. Theorem. Any ergodic toral automorphism T has Lebesgue spectrum of infinite multiplicity.

    Proof: Let T be an ergodic automorphism of a torus $(S^1)^n$. Then U is given (after Fourier transform) by the action of T on the Hilbert space $H = \ell^2(\mathbb{Z}^d)$, using the characters as a basis for $L^2((S^1)^d)$. Given $\chi \neq 0$ in $\mathbb{Z}^d$, the orbit $\langle T^i(\chi)\rangle$ is countable and gives rise to a subspace of H isomorphic to $\ell^2(\mathbb{Z})$ on which U acts by the shift. Since $\mathbb{Z}^d - \{0\}$ decomposes into a countable union of such orbits, we have the theorem. ∎

    Corollary. Any two ergodic toral automorphisms are spectrally equivalent.

    Note that the equivalence to $\oplus_1^\infty \ell^2(\mathbb{Z})$ is easy to construct concretely on the level of characters.

12. *Mixing.* We say T is *mixing* if for any $f, g \in L^2(X)$ we have

    $$\int f \circ T^n(x) g(x)\, d\mu(x) \to \int f \int g.$$

    Equivalently, $\langle U^n f, g\rangle \to 0$ for any two functions f and g of mean zero.

    Any mixing transformation is ergodic. The irrational rotation is ergodic but not mixing.

13. Spectrum and mixing. Any transformation with Lebesgue spectrum is mixing, since

$$\langle U^n f, g\rangle = \int_{S^1} z^n \langle f(z), g(z)\rangle \, |dz| \to 0$$

(indeed by the Riemann-Lebesgue lemma, the Fourier coefficients of any L^1 function tend to zero). Mixing can be similarly characterized in terms of a representative spectral measure: it is necessary and sufficient that $\int z^n f \, d\mu \to 0$ for any $f \in L^2(S^1, \mu)$. (In particular mixing depends only on the measure class, not on the multiplicity).

14. Spectral theory of the baker's transformation T: it is also Lebesgue of infinite multiplicity.

    Proof. Consider the space $X = \sigma_2$ as a group, namely $X = G = \prod_{-\infty}^{\infty} \mathbb{Z}/2$, with Haar measure. Then T is a measure-preserving automorphism of G.

    The space of characters $\widehat{G} = \operatorname{Hom}(G, S^1)$ is isomorphic to the direct sum $\widehat{G} = \oplus \mathbb{Z}/2$ (meaning all but finitely many terms are zero). Here T also acts by the shift, with infinite orbits except for the trivial character. So again we have a natural decomposition of $L^2_0(G) \cong L^2_0(\widehat{G})$ into a countable direct sum of copies of $\ell^2(\mathbb{Z})$ with T acting by the shift. Thus T has Lebesgue spectrum with infinite multiplicity. ■

15. Corollary. All n-shifts and all toral automorphisms are spectrally isomorphic.

16. Here is a result subsuming our analyses of automorphisms of T^2 and σ_n.

    Theorem. For any automorphism $T : G \to G$ of a compact group, T is ergodic iff the only finite orbit of T on $\widehat{G}$ is the one containing the trivial character.

    Moreover if T is ergodic then it is mixing and Lebesgue.

    Proof. We have already proved ergodicity and mixing in the case where all nontrivial orbits on $\widehat{G}$ are infinite.

For the converse, note that if a non-trivial character $\chi : G \to S^1$ has finite orbit (say $U^n(\chi) = \chi$), then the function $f = \sum_1^n U^n \chi$ is fixed by U and non-constant, so T is not ergodic. ■

17. *Information and entropy.* Since ergodic group automorphisms are all spectrally equivalent, we need a finer *spatial* invariant to distinguish between them. Such an invariant is provided by the *entropy* $h(T)$.

    How much information is gained if you learn that a random element x of a probability space (X, μ) belongs to a subset $A \subset X$? The answer should depend only on the *measure*, $\mu(A)$, it should vanish for $A = X$, and it should be additive for *independent* observations (in the sense of probability theory). From these constraints one finds the natural measure of information is unique up to a multiplicative factor, and given by:

$$I(A) = \log(1/\mu(A)).$$

    For a *partition* $X = \sqcup A_i$, the entropy is the *expected* information acquired when we learn to which block of the partition x belongs. It is given by

$$h(\{A_i\}) = \sum \mu(A_i) I(A_i) = -\sum \mu(A_i) \log \mu(A_i).$$

    The *entropy* $h(T)$ measures the growth rate of information under the dynamics, by repeated measurement of a single observable. Starting with any partition $\mathcal{A} = \{A_1, \ldots, A_n\}$, let $\mathcal{A}_n$ be the common refinement of $\mathcal{A}, T^{-1}(\mathcal{A}), \ldots, T^{-n}(\mathcal{A})$. Then we define

$$h(T) = \sup_{\mathcal{A}} \lim \frac{1}{n} h(\mathcal{A}_n).$$

18. *Computing entropy.* Theorem. If the smallest T-invariant σ-algebra generated by $\mathcal{A}$ coincides with the algebra of all measurable sets (mod those of measure zero), then we have

$$h(T) = \lim \frac{1}{n} h(\mathcal{A}_n).$$

19. *Examples.* The entropy of an irrational rotation is zero. The entropy of (Σ_n, T) is $\log n$. The entropy of a Bernoulli shift is $h(T) = -\sum p_i \log p_i$.

    Exercise: The entropy of an ergodic toral automorphism is $\log \lambda$, where $\lambda > 1$ is the expanding eigenvalue of $T \in SL_2(\mathbb{Z})$.

20. *Kazhdan's property T.* Let G be a finitely generated group, with generators $g_1, \ldots, g_n$.

    We say G has *property T* if there exists an $\epsilon > 0$ such that, for any unitary representation $G \to \mathcal{B}(\mathcal{H})$, if there exists a unit vector v with

$$\|g_i \cdot v - v\| < \epsilon, \quad i = 1, \ldots, n,$$

    then there exists a unit vector w with $g \cdot w = w$ for all g in G.

    In other words, if G has an almost invariant vector, then G has an invariant vector.

21. *Easy examples.* Any finite group G has property T. Indeed, by averaging an almost invariant vector over G we obtain an invariant vector.

    The group $\mathbb{Z}$ does *not* have property T. For example, $\mathbb{Z}$ acts on $\ell^2(\mathbb{Z})$ with almost invariant vectors but without invariant vectors.

22. *Functoriality.* Theorem. If G has property T, then so does any quotient group $K = G/H$.

    Cor. If G has property T, then the abelian group $G/[G, G]$ is finite.

23. *Induced representations.* Let $H \subset G$ be a subgroup of finite index in a group G. Then a unitary representation $H \to \mathcal{B}(\mathcal{H})$ gives rise to a canonical *induced representation* of G.

    Namely we consider the space V of all maps $\sigma : G \to \mathcal{H}$ such that $\sigma(xh) = h^{-1} \cdot \sigma(x)$. Since h is unitary, $\|\sigma(x)\|$ depends only on the coset in G/H to which x belongs, and thus V can be given a Hilbert space structure by setting

$$\|\sigma\| = \int_{G/H} \|\sigma(x)\| \, dx = \sum_{G/H} \|\sigma(x)\|.$$

    Then G acts on this space by left translation $(g \cdot \sigma(x) = \sigma(gx))$.

Theorem. If G has property T and $H \subset G$ is a subgroup of finite index, then H has property T.

Proof. Suppose H acts on $\mathcal{H}$ with an almost invariant vector v. Then $\sigma(x) = v$ is an almost-invariant vector for the induced representation on V. Thus G has an invariant vector, and therefore H does as well. ■

Cor. The group $SL_2(\mathbb{Z})$ does *not* have property T.

Proof. $SL_2(\mathbb{Z})$ contains a free subgroup H with finite index, and since H maps surjective to $\mathbb{Z}$ it does not have property T. ■

24. Theorem. The group $SL_n(\mathbb{Z})$ has property T for $n \geq 3$.

25. *Expanding graphs.* Theorem. Fix a set of generators for $SL_3(\mathbb{Z})$. Then the Cayley graphs Γ_p for $SL_3(\mathbb{Z}/p)$ with those generators form a sequence of expanders.

    Proof. ■

12 Retrospective

What is the principal narrative of the preceding discussion as a whole? Several themes emerge, as general mathematical motifs.

1. *Duality.* The theory of distributions emerges in a natural way by considering the dual of the space of smooth functions. To discuss the dual, we must first introduce a C^∞ topology, which takes us out of the category of Banach spaces. We are thus led to study locally convex topological vector spaces, and to generalize the main principles of analysis (e.g. the Hahn Banach theorem and the closed graph theorem) to a more general setting.

2. *Convolution.* A central role in analysis, and further use of the group structure on $\mathbb{R}^n$, arises from the operator of convolution $f * g$. This can be used to smooth distributions as well as to develop Tauberian theorems and prove the prime number theorem. Partial differential equations can be solved by convolution with a fundamental solution.

3. *The Fourier transform.* This is a second kind of duality, emerging from the consideration of the characters $\exp(ixt)$ on the group $G = \mathbb{R}^n$. We get an isomorphic $L^2(G) \cong L^2(\widehat{G})$, with the important property that it turns both convolution and translation into multiplication by functions.

   Now differentiation is just infinitesimal translation, so PDEs become polynomial equations, and compactly supported distributions become analytic functions. The Fourier transform also renders compactly supported distributions *visible* by converting them into analytic functions.

4. *Algebras and their ideals.* The Fourier transform emerges naturally by studying $(L^1, *)$ as a commutative Banach algebra. In the setting of an algebra we revisit duality, with the focus on describing the multiplicative linear functionals.

   The continuous functional calculus and the spectral theorem also emerge from algebraic considerations, and show that every self–adjoint operator is equivalent the a multiplication operator. In this way the notion of composition of operators and multiplication of functions become profoundly related.

A Appendix: Problems

1. Let us say $f : [0, 1] \to \mathbb{R}$ is *pretty continuous* if it is continuous at a dense set of points in $[0, 1]$. Show that the sum of two pretty continuous functions is pretty continuous.

2. What are the extreme points for the closed unit ball in $L^p[0, 1]$, $1 \le p \le \infty$?

3. Given an example of a positive measurable function $f : [0, 1] \to \mathbb{R}$ such that $\int_a^b f$ is infinite for all $0 \le a < b \le 1$.

4. Is the product of two measurable functions always measurable?

5. Given an example of a continuous function $f : [0, 1] \to [0, 1]$ and a (Lebesgue) measurable set $E \subset [0, 1]$ such that $f^{-1}(E)$ is not measurable.

6. Let $A \subset [0,1]$ be a set of positive measure. Show that the set of differences $x-y$, $x, y \in A$ contains a nonempty open interval. Can this happen if A has measure zero?

7. Let X and Y be compact Hausdorff spaces. Show that the set of functions of the form $\sum_1^n f_i(x)g_i(y)$, $f_i \in C(X)$, $g_i \in C(Y)$, is dense in $C(X \times Y)$. (Here $C(X)$ is the Banach space of continuous, real-valued functions on X, with the sup-norm.)

8. Let $A : X \to Y$ be a linear map between Banach spaces that is continuous from the weak topology on X to the norm topology on Y. Show that $A(X)$ is finite-dimensional. (Hint: show every open set in X that contains zero also contains a subspace of finite codimension.)

9. Let $X = \ell^\infty(\mathbb{N})$ be the Banach space of bounded real sequences $x = (x_0, x_1, \ldots)$ and let $T(x)_i = x_{i+1}$. Let X^* be the dual of X with the weak* topology. Define $\phi_n \in X^*$ by $\phi_n(x) = (1/n)\sum_0^{n-1} x_i$.

   (a) Show that ϕ_n has an accumulation point in $\psi \in X^*$. (b) Show that $\psi(T(x)) = \psi(x)$, and $\psi(x) = \lim x_i$ if this limit exists. (c) Show there is no subsequence of ϕ_n that converges to ψ.

10. Give an example of a sequence of continuous functions $f_n \in C[0,1]$ with $\|f_n\| = 1$ but $f_n \to 0$ weakly.

11. Let X be a complete, locally convex topological vector space. Show the following are equivalent: (a) The topology on X is induced by a translation invariant metric (X is a *Fréchet space*); (b) The topology on X is induced by a sequence of seminorms $p_n(x)$.

12. Show that $X = C(\Omega)$ is an example of a Fréchet space which *cannot* be metrized so that $nB(0, 1/n) \supset B(0, 1)$.

13. Let X be the Banach space $C[0,1]$, with dual space $X^* = M[0,1]$ the space of measures. (a) Show that the measures of the form $f(x)\,dx$, $f \in L^1[0,1]$, form a closed subspace of $M[0,1]$ in the norm topology. (b) Show that $X^{**} \neq X$.

14. Let $K \subset C[0,1]$ be the set of all continuous functions satisfying $|f(x)| \leq 1$ and $|f(x) - f(y)| \leq |x - y|$ for all $x, y \in [0,1]$. (a) Prove that K is a compact, convex set. (b) What are its extreme points?

15. Give an example of a compact convex set $K \subset \mathbb{R}^3$ whose set of extreme points is not closed.

16. Show that the group G of affine automorphisms $g : \mathbb{R} \to \mathbb{R}$ has a right invariant measure and a left invariant Haar measure, but they are not proportional. (Here G consists of the maps $g(x) = ax + b$ with $a \in \mathbb{R}^*$ and $b \in \mathbb{R}$.)

17. Show that $\ell^1(\mathbb{N})$ is a dual space. That is, exhibit a Banach space X such that X^* is isometric to $\ell^1(\mathbb{N})$.

18. Let $C(\mathbb{R})$ denote the space of all continuous functions $f : \mathbb{R} \to \mathbb{R}$, with the topology of uniform convergence on compact sets. Describe its dual space.

19. Let $X \cong \mathbb{R}^{\mathbb{N}}$ be the vector space of all real sequences $a = (a_0, a_1, a_2, \ldots)$ with the product topology.

    (a) Show that X^* can be identified with the subspace of X where $a_i = 0$ for all i sufficiently large, using the pairing

    $$\langle a, b \rangle = \sum a_i b_i.$$

    (b) Show that if $a(n)$ is a sequence in X^*, and $a(n) \to 0$ in the weak* topology, then there is a single N such that $a_i(n) = 0$ for all $i > N$.

20. Let $X = C_c(\mathbb{R})$ be the space of compactly supported continuous functions with the inductive topology. (a) Show the topology on X does not have a countable base at $x = 0$. (b) Show that X is not metrizable. **(c) Is a set $F \subset X$ closed iff it is sequentially closed?

21. Describe explicitly a family $\{B_\alpha\}$ of sets in $C^\infty(\mathbb{R})$, such that (i) every B_α is bounded and (ii) if B is bounded, then we have $B \subset B_\alpha$ for some α.

22. Prove that a continuous function $f : \mathbb{R} \to \mathbb{R}$ is *Lipschitz* (it satisfies a bound of the form $|f(x) - f(y)| \le M|x - y|$ for all $x \neq y$ in $\mathbb{R}^n$) if and only its distributional derivative df/dx lies in $L^\infty(\mathbb{R})$.

23. Consider the locally integrable function defined by $f(x) = -2x^{-1/2}$ for $x > 0$, and $f(x) = 0$ otherwise, as a distribution on $\mathbb{R}$. Prove that, if we regard $f'(x) = Df$ as a distribution, then

$$\int_{\mathbb{R}} \phi(x) f'(x) dx = \int_0^\infty (\phi(x) - \phi(0)) x^{-3/2}\, dx$$

for all $\phi \in C_c^\infty(\mathbb{R})$.

24. Recall that $df/d\overline{z} = (1/2)(df/dx + i df/dx)$ vanishes if f is holomorphic. Show the function $f(z) = 1/z$ on $\mathbb{C}$ satisfies, as a distribution,

$$df/d\overline{z} = C\delta_0,$$

and compute the value of C.

25. Let μ be a finite measure on $\mathbb{R}^n$, and let $F(x_1, \ldots, x_n)$ be the measure of the set of y such that $y_i < x_i$. Prove that $D^1 D^2 \cdots D^n F = \mu$ as a distribution, where $D^i = d/dx_i$.

26. Define $T_n \in L^1(\mathbb{R})$ with $\int T_n = 1$ by $T_n(x) = n/2$ on $[-1/n, 1/n]$ and $T_n(x) = 0$ elsewhere. Prove that if $\sum 1/n_i < \infty$, then $f_i(x) = T_{n_1} * \cdots * T_{n_i}$ converges uniformly to a compactly supported C^∞ function $\phi \geq 0$ with $\int \phi = 1$.

27. Let $X = C^{-\infty}(\Omega)$ be the space of distributions on $\Omega \subset \mathbb{R}^n$. (a) Show that compactly supported distributions are dense in X. (b) Show that the real vector space V spanned by the distributions δ_p, with $p \in \Omega$, is dense in X.

28. *Also:* Rudin, Chapter 6, problems 11, 13, 14, 18, 20, 21, 24.

29. Find the Fourier transform of $f(x) = 1/(1 + x^2)$. (Hint: use complex analysis.)

30. Given an arbitrary sequence of real numbers $a_n \to \infty$, define Λ on test functions by $\Lambda(\phi) = \sum a_n \phi(n)$.

    (a) Show that Λ is a continuous linear functional on $C_c^\infty(\mathbb{R})$, and hence a distribution. (b) Show that a_n can be chosen so that Λ has no continuous extension of $\mathcal{S}(\mathbb{R})$ (i.e. Λ is not a *tempered* distribution). (c) Why does this not contradict the Hahn–Banach theorem?

31. Give an explicit distribution Λ on $\mathbb{R}$ such that $x\Lambda = 1$ as distributions. (Note: $1/x$ is not in $L^1(\mathbb{R})$, so it does not define a distribution as it stands.)

32. Prove that $\mathcal{S}(\mathbb{R}^n)$ is the largest subspace $V \subset C_0(\mathbb{R}^n) \cap C^\infty(\mathbb{R}^n)$ that is closed under differentiation and multiplication by polynomials. (This means $f \in V \implies D^\alpha f, x^\alpha f \in V$ for all α, where $x^\alpha = x_1^{\alpha_1} \cdots x^{\alpha_n}$.)

33. Fix $a > 0$ and let $f(x) = \exp(-a|x|)$. (a) What is $\widehat{f}(t)$? (b) Using Poisson summation, obtain a formula for $\coth(x)$ as an infinite sum of rational functions with denominators $x^2 + \pi^2 n^2$.

34. Let $\Lambda \subset \mathbb{R}^n$ be a lattice (a discrete subgroup whose quotient $\mathbb{R}^n/\Lambda$ is compact), and let V be the volume of $\mathbb{R}^n/\Lambda$. Let

$$\Lambda' = \{x \in \mathbb{R}^n \ : \ \langle x, y\rangle \in \mathbb{Z} \text{ for all } y \in \Lambda\}$$

be the dual lattice, let $f \in \mathcal{S}_n$ be a Schwartz function and normalize the Fourier transform (at variance with Rudin) by:

$$\widehat{f}(t) = \int f(x) \exp(-2\pi\langle x, y\rangle)\, dx.$$

Prove that:

$$\sum_{\Lambda} f(x) = V^{-1} \sum_{\Lambda'} \widehat{f}(t).$$

35. Let

$$H(\mathbb{Z}) = \langle a, b, c \ : \ [a, b] = c, [a, c] = [b, c] = 1\rangle$$

be the Heisenberg group, and let $W_n \subset H(\mathbb{Z})$ denote all elements that can be expressed as words of length at most n in a, b, c and their inverses. Show that $|W_n| \geq cn^4 > 0$.

36. Let $f(t)$ be an analytic function on $\mathbb{C}$ such that for each $N \geq 0$ there is a constant C_N such that

$$|f(t)| \leq C_N \exp(R|\operatorname{Im} t|)/(1 + |t|^2)^N$$

for all $t \in \mathbb{C}$. Show that $f|\mathbb{R}$ belongs to the Schwartz space $\mathcal{S}(\mathbb{R})$.

37. Show there is a smooth function $f(x)$ on $\mathbb{R}$ that belongs to every Sobolev space H^s but which is not a Schwartz function.

38. Let μ be a finite positive measure on $\mathbb{R}^n$. Show that μ is a tempered distribution, and that its Fourier transform $\widehat{\mu}(t)$ is bounded and uniformly continuous.

39. Show that $f(x) = e^x \cos(e^x)$ is a tempered distribution on $\mathbb{R}$, but $g(x) = e^x$ is not.

40. Let $f(x) = C\exp(-x^2/2)$ on $\mathbb{R}$, where C is chosen so that $\|f\|_2 = 1$. (a) Show that the uncertainty principle is sharp for f, in the sense that $1/2 = (\Delta P)(\Delta Q)$. (b) Show that if $f \in \mathcal{S}(\mathbb{R})$, $\|f\|_2 = 1$, and $\langle P\rangle = \langle Q\rangle = 0$, and $1/2 = (\Delta P)(\Delta Q)$, then $f = C\exp(-ax^2/2)$ for some C and a. (Hint: when is the Cauchy–Schwarz inequality an equality?)

41. (The quantum harmonic oscillator.) Define operators on $f(x) \in \mathcal{S}(\mathbb{R})$ by

$$Hf = x^2 f - D^2 f, \quad Rf = xf - Df, \quad \text{and} \quad Lf = xf + Df,$$

where $D = d/dx$. (These are the Hamiltonian, raising and lower operators. The eigenvalues of H are the energy levels of the system, which are quantized.)

    (a) Prove that $H = (LR + RL)/2$.
    (b) Prove that $I = (LR - RL)/2$.
    (c) Show that $f_0(x) = \exp(-x^2/2)$ satisfies $Hf_0 = f_0$ and $Lf_0 = 0$.
    (d) Define $f_n(x) = R^n(f_0)$ for $n > 0$, and show that $f_n(x) = P_n(x)f_0(x)$ where $P_n(x)$ is a polynomial of degree n.
    (e) Show that $Hf_n = (1 + 2n)f_n$.
    (f) Show that $\widehat{f_n}(x) = (-i)^n f_n(x)$.

42. Let H^s denote the Sobolev space on $\mathbb{R}^n$. (a) Show that a tempered distribution f belongs to H^N, $N \geq 0$ an integer, if and only if $D^\alpha f \in H^0$ for all α with $|\alpha| \leq N$. (b) Show that $H^0 = (I - \Delta)H^2$. (c) Show there is no linear differential operator such that $H^0 = P(D)H^1$, when $n > 1$.

43. *Also:* Rudin, Chapter 7, problems 2, 4, 8, 10, 21, 24.

44. Prove that a continuous function $f : \mathbb{R}^n \to \mathbb{R}$ is *Lipschitz* (it satisfies a bound of the form $|f(x) - f(y)| \le M|x - y|$ for all $x \neq y$ in $\mathbb{R}^n$) if and only its distributional derivatives df/dx_i lie in $L^\infty(\mathbb{R})$ for $i = 1, \ldots, n$. (This generalizes Problem 22.)

45. Evaluate $\int_{-\infty}^{\infty} e^{ixt}/(1+t^2)\, dm(t)$. (Cf. Problem 29.) Then use your answer to guess a fundamental solution to $(I - \Delta)E = \delta_0$ on $\mathbb{R}$. Finally verify directly that your chosen E is a fundamental solution.

46. Write a computer program to solve $\Delta f = 0$ on a square with boundary values $f = 1$ on a one edge and $f = 0$ on the other three edges. Display your results graphically (e.g. by 'coloring' the square with letters indicating the size of f).

47. Find a fundamental solution for the operator $P(D) = (I + D)^n$ on $\mathbb{R}$, $n > 0$ an integer. (Hint: first find a tempered fundamental solution for $n = 1$, then convolve.)

48. Let $P(t) = \sum_1^n t_i^2$ on $\mathbb{R}^n$, so $P(iD) = -\Delta$. For which values of n is $1/P(t)$ a tempered distribution? Compute the Fourier transform of $1/P(t)$ for $n = 3$ and use the result to give a (fundamental) solution E to the equation $\Delta E = \delta$.

49. Let $f(z)$ be a holomorphic function on the unit disk $\Delta \subset \mathbb{C}$. Show there exists a sequence of polynomials $p_n(z)$ such that $p_n|\Delta \to f$ in $C(\Delta)$.

50. (Continuation.) Show that for any $g \in \mathcal{D}'(\Delta)$, there exists an $f \in \mathcal{D}'(\Delta)$ such that $\overline{\partial} f = g$.

    (Hint: write $g = \sum g_n$ as a sum of compactly supported distributions whose supports tend to $\partial\Delta$; solve $\overline{\partial} f_n = g_n$; and use the preceding problem to find polynomials such that $\sum(f_n - p_n)$ converges on Δ.)

51. (Continuation.) Show that for any $g \in \mathcal{D}'(\Delta)$, there exists an $f \in \mathcal{D}'(\Delta)$ such that $\Delta f = g$.

52. Let f be a compactly supported distribution on $\mathbb{R}^n$, and ϕ a test function such that $\phi * f = f$. Prove that $f = 0$.

53. Fix $s \in \mathbb{R}$, and let H^s denote the Sobolev space on $\mathbb{R}^n$, $n > 0$. Show that no matter how large we take $N > 0$, the natural inclusion map $T : H^{s+N} \to H^s$ is not a compact operator.

54. (Harmonic polynomials.) Let P_d denote the space of real homogeneous polynomials p of degree d on $\mathbb{R}^n$, and let $H_d \subset P_d$ be the subspace of *harmonic polynomials*, where $\Delta p = 0$.

    (a) Show that on $\mathbb{R}^2 \cong \mathbb{C}$, for $d > 0$ the space H_d is spanned by $\operatorname{Re} z^d$ and $\operatorname{Im} z^d$.

    (b) Show that on $\mathbb{R}^2$, we have $P_d = H_d \oplus |x|^2 P_{d-2}$ for all d.

    (c) Show the same is true on $\mathbb{R}^3$, and use it to compute $\dim H_d(\mathbb{R}^3)$.

    (d) Let f be a tempered distribution on $\mathbb{R}^n$ such that $\Delta f = 0$. Show that $\widehat{f}$ is supported at the origin, and hence f is a harmonic polynomial.

55. Show there exist compact supported smooth functions f_i on $\mathbb{R}^n$ such that $\|\Delta f_i\|_{L^2} = 1$ but $\|f_i\|_{L^2} \to \infty$. Why does this not contradict Theorem 7.10?

56. Let $P(x, D)$ be an elliptic differential operator of order N on $\mathbb{R}^n$. Suppose N is *odd*. (i) Prove that $n = 1$ or $n = 2$. (ii) Prove that if $n = 2$, then the symbol $P(x, t)$ cannot be a real–valued function.

57. Given a sequence of weights $w_n > 0$, let $\ell^2_w(\mathbb{Z})$ be the Banach space of sequences such that

$$\|a\|_w^2 = \sum_{\mathbb{Z}} w_n |a_n|^2 < \infty.$$

    (i) Show that if $x_n/y_n \to \infty$ as $|n| \to \infty$, then the natural inclusion $T : \ell_x(\mathbb{Z}) \to \ell_y(\mathbb{Z})$ is a compact operator.

    (ii) Make up a reasonable definition of Sobolev spaces $H^s(S^1)$ for functions on the unit circle in $\mathbb{C}$ (with Fourier series $f(z) = \sum a_n z^n$), and prove that the inclusion $H^r(S^1) \to H^s(S^1)$ is a compact operator whenever $r > s$.

58. (The mean value property.) Prove, using the following outline, that if f is harmonic on $\mathbb{R}^n$, then $f(x)$ is equal to the average of f over $B(x, r)$.

(Outline: let $g_r(x) = \chi_{B(0,r)}(x)/\operatorname{vol}(B(0,r))$, so $\int g_r = 1$. There is a compactly supported continuous function $h_{r,s}$ such that $\Delta h_{r,s} = g_r - g_s$. Then $(\Delta f) * h_{r,s} = f * (\Delta h_{r,s}) = f * g_r - f * g_s = 0$. Letting $s \to 0$, we have $(f * g_s)(x) \to f(x)$ while $(f * g_r)(x)$ gives the average of f over $B(x, r)$.)

59. *Also:* Rudin, Chapter 8, problems 4, 5, 8, 10, 12.

60. Let f_α and f_β denote the indicator functions of $[0, \alpha]$ and $[0, \beta]$ respectively, with $\alpha, \beta > 0$. Show that the ideal $J \subset A = (L^1(\mathbb{R}), *)$ generated by these two functions is dense iff α/β is irrational.

61. *Also:* Rudin, Chapter 9, problems 1, 2, 15.

62. Let $\phi(x) = \cos(x^2) \in L^\infty(\mathbb{R})$. Show that $\lim_{x\to\infty}(\phi * f) = 0$ for all $f \in L^1(\mathbb{R})$, even though $\phi(x)$ does not tend to zero.

63. Compute $\int_{\mathbb{R}} f(x) - f(x+1)\,dx$ when (i) $f(x) = 1/x$ for $x > 1$ and 0 elsewhere; (ii) $f(x) = e^{-x}[e^x]$, where $[y]$ is the largest integer not exceeding y.

64. Prove or disprove: (i) Every measurable function $f(x)$ on $\mathbb{R}$ can be written in the form $f(x) = g(x+1) - g(x)$ for some other measurable function g; (ii) If $f \in L^1(\mathbb{R})$ and $\int f = 0$, then $f(x) = g(x+1) - g(x)$ for some $g \in L^1(\mathbb{R})$.

65. Let $g(x) \geq 0$ be an increasing function such that $h(x) = e^{-x}g(x) \leq 1$ for all x. (i) Can we conclude that $h(x)$ is slowly oscillating? (ii) Suppose $\lim_{x\to\infty}(f * h)(x) = L$ for all $f \in L^1(\mathbb{R})$ with $\int f = 1$. Prove that $\lim_{x\to\infty} h(x) = L$.

66. Prove that $\sum 1/p = \infty$, where the sum is over all primes. (Suggestion: use the behavior of $\zeta(s)$ at $s = 1$.)

67. Prove that the nth prime number satisfies $p_n \sim n \log n$. You may assume the prime number theorem.

68. Let $\mu(n)$ and $\Lambda(n)$ be the Möbius and von Mangoldt functions. Prove that $\sum \mu(n)n^{-s} = 1/\zeta(s)$ and $\sum \Lambda(n)n^{-s} = -\zeta'(s)/\zeta(s)$, for $\operatorname{Re}(s) > 1$.

69. Prove that $\Gamma(1+is) \neq 0$ for all $s \in \mathbb{R}$.

70. Recall that $s_n \to S(A)$ means $\lim_{r\to 1}(1-r)\sum s_n r^n = S$, and $s_n \to S(C)$ means $\lim (1/N)\sum_1^N s_n = S$.

    Show that if $s_n \to S(C)$ then $s_n \to S(A)$. We do *not* assume $s_n = O(1)$. (Hint: summation by parts.)

71. Give an example to show that $s_n \to S(A)$ does not imply $s_n \to S(C)$. (Of course the s_n must be unbounded.)

72. Let $k(t) = e^{-t}$, let s_n be a bounded sequence, and let $f(t) = s_{[t]}$. Prove that $(1/L)\int f(t)k(t/L)\,dt \to S$ as $L \to \infty$ if and only if $s_n \to S(A)$.

73. Evaluate $1^2 - 2^2 + 3^2 - 4^2 + \cdots$ and $\sin(\theta) + \sin(2\theta) + \sin(3\theta) + \cdots$.

74. Which $f \in A = (L^1(\mathbb{R}), *)$ satisfy $f * A = A$?

75. *Also:* Rudin, Chapter 9, problems 1, 2, 6, 7, 15.

76. (i) Prove that if a Banach space A is also an algebra, and multiplication is continuous, then there exists a C such that $\|ab\| \leq C\|a\| \cdot \|b\|$ for all $a, b \in A$. (ii) Show that we can find an equivalent norm $|a|$ on A such that $|ab| \leq |a| \cdot |b|$. (Equivalent means that $\|a\|$ and $|a|$ induce the same topology on A.)

77. Let $A = C^k[0,1]$ with $\|f\| = \sup_{i\leq k}\sup_{[0,1]}|f^{(i)}(x)|$. Show by example that for $k \geq 1$, A is not a Banach algebra as it stands, and say explicitly how to modify the norm so it becomes one for all k.

78. (Functional calculus.) Let A be a Banach algebra (over $\mathbb{C}$), and let $f(t) = \sum a_n t^n$ be an entire function on $\mathbb{C}$. (i) Prove that $F(x) = \sum a_n x^n$ exists for all x and defines an analytic map $F : A \to A$, in the sense that $\phi(F(a+bt))$ is an analytic function of t for all $a, b \in A$ and $\phi \in A^*$. (ii) Prove that $\exp(A) \subset A^\times$ (the invertible elements). (iii) Give examples of algebras of the form $A = C(K)$ to show that sometimes equality holds, and sometimes it does not. (iv) What topological property of K characterizes surjectivity of the exponential map?) (Hint: use an appropriate cohomology theory.)

79. Show that every maximal ideal in $C^k[0,1]$ has the form $M_p = \{f : f(p) = 0\}$ for some $p \in [0,1]$.

80. Let $T = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$, considered as an operator on the Hilbert space $\mathbb{C}^2$. Compute $\|T^n\|$ for all n, and verify that $\rho(T) = \lim \|T^n\|^{1/n}$.

81. Show that every continuous homomorphism $\phi : \mathbb{R}^n \to S^1 \subset \mathbb{C}^*$ has the form $\phi(x) = \exp(it \cdot x)$ for some $t \in \mathbb{R}^n$.

    (Hint: First prove for result for $n = 1$, and then use the result along each coordinate axis in $\mathbb{R}^n$.)

82. Let A be a Banach algebra. Prove in detail that for any $x \in A$ and $\phi \in A^*$, the function $f(\lambda) = \phi((\lambda - x)^{-1})$ is an analytic function on the open region $\Omega = \mathbb{C} - \sigma(x)$.

83. Let A be the matrix algebra $M_n(\mathbb{C})$, with $n \geq 2$. Describe the space of multiplicative linear functionals, $\mathrm{Spec}(A)$.

84. Give an example of a finite–dimensional, commutative Banach algebra with nontrivial radical.

85. Given an example of an operator T on a Hilbert space such that $\rho(T) = 0$ but $T^n \neq 0$ for all $n > 0$.

86. Let $A = C^k[0, 1]$ with $\|f\| = \sum_{i=0}^{k} \sup |f^{(i)}(x)|$. Prove directly that $\sup |f(x)| = \lim \|f(x)^n\|^{1/n}$.

87. Let a_n be a real sequence such that $a_{i+j} \leq a_i + a_j$ for all i, j. Show that $\lim a_n/n$ exists (provided we allow the limit to assume the value $-\infty$).

88. (a) Show that if the Fourier series of $f \in L^1(S^1)$ vanishes, then f vanishes.

    (b) Show that for each $L > 0$ we have a bounded operator

    $$T_L : L^1(\mathbb{R}) \to L^1(S^1)$$

    sending f to $\sum_n f(x/L + n)$, regarded as a function on $\mathbb{R}/\mathbb{Z}$. How is the Fourier series for $T_L(f)$ related to $\widehat{f}(t)$?

    (c) Show that $T_L(f) = 0$ for all L if and only if $f = 0$.

    (d) Combine these observations to show that the Fourier transform

    $$\mathcal{F} : L^1(\mathbb{R}) \to C_0(\mathbb{R})$$

    is injective.

    (e) How would you generalize this argument to $\mathbb{R}^n$?

89. Let $\phi : \ell^\infty(\mathbb{Z}) \to \mathbb{C}$ be a multiplicative linear functional, and let $\mathcal{F} = \{S \subset \mathbb{Z} : \phi(\chi_S) = 1\}$. Prove that $\mathcal{F}$ is an ultrafilter on $\mathbb{Z}$. Then show that $\phi(a) = L$ if and only if, for all $\epsilon > 0$, $\{n : |a_n - L| < \epsilon\} \in \mathcal{F}$. Finally show there exists a nonprincipal ultrafilter on $\mathbb{Z}$, i.e. one satisfying $|A| = \infty$ for all $A \in \mathcal{F}$.

90. Let A denote the algebra of all complex–valued continuous function $f : \mathbb{R} \to \mathbb{C}$. Prove that A is not a Banach algebra. That is, there is no way to introduce a norm on A so it becomes a Banach space and $\|fg\| \le \|f\| \cdot \|g\|$ for all $f, g \in A$.

91. Let X be a compact Hausdorff space. Prove that $C(X)$ is separable (has a countable dense set) if and only if X is metrizable.

92. Prove that X is connected if and only if $C(X)$ has no idempotents (solutions to $a^2 = a$ other than $a = 0$ and $a = 1$).

93. Let $\phi_n(a) = a_n$, $n \in \mathbb{N}$, be the point evaluations in the compact space $\operatorname{Spec} \ell^\infty(\mathbb{N})$. Prove that ϕ_n has no (weak*) convergent subsequences.

94. Prove that the radical of $(\ell^1(\mathbb{Z}), *)$ is trivial.

95. Given $f \in L^2(S^1)$ with Fourier series $f = \sum a_n \lambda^n$, let $S(f) = \sum |a_n|$. (i) Prove there exists a bounded function such that $S(f) = \infty$. (ii) Prove that if $f_n \to f$ in $L^2(S^1)$, then $S(f) \le \limsup S(f_n)$. (iii) Prove that the closure in L^2 of the unit ball in $C(S^1)$ is the unit ball in $L^\infty[0,1]$. (iv) Prove that $S(f)$ is unbounded on the unit ball in $C(S^1)$. (Otherwise using (ii) and (iii) we contradict (i)). (v) Prove there exists an $f \in C(S^1)$ with $S(f) = \infty$. (Use Baire category.)

96. Let $K \subset \mathbb{C}^n$ be a compact set. We say $z \in K$ is an *analytic extreme point* if any analytic map $f : \Delta \to K$ with $f(0) = z$ is constant.

    (i) Prove that a polynomially convex set is the polynomial convex hull of its analytic extreme points. (Use the maximum principle).

    (ii) Prove that the set of analytic extreme points of $\overline{\Delta}^2 \subset \mathbb{C}^2$ is the torus $K = S^1 \times S^1$.

    (iii) Prove that the polynomial convex hull of $S^1 \times S^1$ is $\overline{\Delta}^2$.

97. Prove that the union K of two disjoint closed disks in $\mathbb{C}$ is polynomially convex.

98. Let T_n and T be bounded operators on the Hilbert space $\mathcal{H}$, and suppose $T_n \to T$ weakly. Prove that $\|T_n\|$ is bounded.

99. Let $H = \ell^2(\mathbb{N})$ and let $A = \ell^\infty(\mathbb{Z}) \subset \mathcal{B}(H)$, where the sequence $a \in A$ corresponds to the operator $T_a(b) = a_n b_n$. (i) Prove that A is closed in the weak operator topology. (ii) Describe all $a \in A$ such that T_a is a compact operator.

100. Let $T \in \mathcal{B}(\mathcal{H})$ be a compact operator. Show that if $x_n \to y$ weakly in $\mathcal{H}$, then $T(x_n) \to T(y)$ strongly (i.e. $\|T(x_n) - T(y)\| \to 0$).

101. Show that if $T \in \mathcal{B}(\mathcal{H})$ and $\langle Tx, x\rangle \in \mathbb{R}$ for all $x \in \mathcal{H}$, then T is self–adjoint.

102. Let $\mathcal{H} = L^2[-1,1]$. Given $f \in L^\infty[-1,1]$, define an operator in $\mathcal{B}(\mathcal{H})$ by $T_f(g) = fg$. Describe the spectrum of T_f, its spectral measure class and its multiplicity function for (a) $f(x) = x^3$, (b) $f(x) = \sin(1/x)$, and (c) $f(x) = \max(0, x)$.

103. Let $\mathcal{H} = L^2(\mathbb{R})$ and let $T \in \mathcal{B}(\mathcal{H})$ be the operator $T = (I - \Delta)^{-1}$. Describe the spectrum of f, its spectral measure class and its multiplicity function. (Hint: use the Fourier transform to turn T into a more recognizable operator.)

104. Let $A_0 \subset A = (\ell^1(\mathbb{Z}), *)$ be the set of sequences $a = (a_n)$ such that $\sum a_n = 0$. (i) Show that A_0 is a maximal ideal in A. (ii) Given $a \in A_0$, show that $a * A_0$ is dense in A_0 if and only if $f(t) = \sum a_n t^n \neq 0$ for all $t \in S^1$ other than $t = 1$.

References

[CFS] I. P. Cornfeld, S. V. Fomin, and Ya. G. Sinai. *Ergodic Theory.* Springer-Verlag, 1982.

[Fol] G. B. Folland. *Harmonic Analysis on Phase Space.* Princeton University Press, 1989.

[GH] P. Griffiths and J. Harris. *Principles of Algebraic Geometry.* Wiley Interscience, 1978.

[Har] G. H. Hardy. *Divergent Series.* Chelsea, 1991.

[Me] R. Mañé. *Ergodic Theory and Differentiable Dynamics.* Springer-Verlag, 1987.

[Nai] M. A. Naimark. *Normed Rings.* P. Noordhoff N. V., 1964.

[Ru] W. Rudin. *Functional Analysis.* McGraw-Hill, 1991.

[Ser] J. P. Serre. *A Course in Arithmetic.* Springer-Verlag, 1973.

[St1] E. M. Stein. *Singular Integrals and Differentiability Properties of Functions.* Princeton University Press, 1970.

[St2] E. M. Stein. *Harmonic Analysis.* Princeton University Press, 1993.

[Tr] F. Trèves. *Linear Partial Differential Equations with Constant Coefficients.* Gordon and Breach Science Publishers, 1966.

[Wie] N. Wiener. Tauberian theorems. *Annals of Math.* **33**(1932), 1–100.