Advanced Real Analysis
Harvard University — Math 212b
Course Notes

Contents

1 Introduction

1. Some basic references:

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

   See [Ru], [RS], [Be] and [RN].

2. 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 $i\xi$).

3. 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 Convexity and locally convex topological vector spaces

1. Note that $X = L^p[0,1]$, $0 < p < 1$, is complete, metrizable, but not locally convex. In fact that convex hull of any open neighborhood of the origin is the entire space. This can be seen by writing $f \in L^p[0,1]$ as $\sum_1^n f_i$ where $\int |f_i|^p = (1/n) \int |f|^p$.

   Corollary: X^* is trivial.

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

2. Krein-Milman: A compact convex set in a LCTVS is the closed convex hull of its extreme points. Milman: if L and $K = \overline{\text{hull}(L)}$ are both compact, then $\text{ex}(K) \subset L$.

3. Extreme points: $X = L^1[0,1]$ has no extreme points, and thus it is not a dual space. The unit balls in the other L^p spaces, $p > 1$, do have

extreme points; this follows directly from Krein-Milman.

4. An example of a compact set $L \subset X$ 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.

5. Applications: Stone-Weierstrass, existence of Haar measure on compact groups.

3 Distributions

1. Topology on $C_c^\infty(\Omega)$. Two points. First, for every compact set, we make $C_c^\infty(K)$ into a Fréchet space by using

   $$U_{k,\epsilon} = \{\phi \,:\, \|\phi\|_{C^k} < \epsilon\}$$

   as a base at the origin. That is, $C_c^\infty(K)$ is locally convex, Hausdorff and metrizable.

   Then, we give $C_c^\infty(\Omega) = \bigcup C_c^\infty(K_i)$ the *inductive* topology: a base at the origin consists of *convex* sets U that meet each $C_c^\infty(K)$ in a convex, open set.

2. *Direct limit topology.* This topology has the property that a linear map $A : C_c^\infty(\Omega) \to X$, where X is a LCTVS, is continuous iff A is continuous on each $C_c^\infty(K)$. It is the weakest topology making this assertion true.

   More generally, if $V_1 \subset V_2 \subset \cdots$ is a sequence of LCTV's with continuous inclusions, $V = \varinjlim V_i$ has a natural topology define by taking as a base at the origin the *convex* sets that meet each V_i in a convex, balanced open set.

3. *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 $Y = \bigwedge_1^\infty S^1$. 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.

4. *Convexity.* If we drop the requirement of *convexity* of U we obtain a different topology on $C_c^\infty(\Omega)$ — one that is not locally convex. For example, the set

$$U = \bigcup_1^\infty \{f \ : \ \operatorname{supp} f \subset [-k,k], \|f\|_{C^0} < 1/k\}$$

would be open, but it contains no open convex set. (Proof: *any* open neighborhood V of the origin contains a function f with $f(0) > 1/k$ for some k. It also contains a function $g(x) \neq 0$, where $x \gg k$ and $f(x) = 0$. Then $(1-\epsilon)f + \epsilon g$ has large support and large C^0-norm, so it does not belong to U.)

   Reference: [Tr, p.18].

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

6. Topologies on distributions. There are two: the *weak* (or weak*) topology, the topology of pointwise convergence; and the *strong* topology, the topology of uniform convergence on bounded sets.

   These two topologies agree on sequences!

   For K compact, we have $C_c^\infty(K) = \bigcap C_c^i(K)$; similarly, we have

$$C^{-\infty}(K) = C^\infty(K)^* = \bigcup C_c^i(K)^*.$$

   The *strong* topology on the space of distributions is the *inductive* topology on this union of Banach spaces.

   In this topology, a sequence of distributions satisfies $\phi_n \to \phi$ iff there exists a single i such that $\phi_n, \phi \in C_c^i(K)^*$ for all n, and convergence takes place there; in particular, there is an M such that for all n,

$$|\phi_n(f)| \leq M\|f\|_{C^i}.$$

7. Example: given a compactly suppport smooth ψ on $\mathbb{R}^n$, with $\int \psi = 1$, let $\psi_r(x) = r^{-n}\psi(x/r)$. Then as $r \to \infty$, $\psi_r \to \delta$ as a distribution.

Example: let $f_n(x) = \sin(nx)$ on $\mathbb{R}$. Then as $n \to \infty$, $f_n \to 0$ as a distribution.

Example: let $f_n(x) = n^2 \sin(n\pi x)$ on the interval $[-1/n, 1/n]$ (and 0 elsewhere). Then for any smooth $\phi(x)$, we have

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

8. Theorem. If $\Lambda f = \lim \Lambda_i f$ exists for all $f \in C_c^\infty(\Omega)$, then Λ is a distribution.

   The proof is a generalization of the uniform boundedness principle: review the case of Banach spaces. (You can't construct an unbounded linear functional by hand.)

9. Theorem: A sequence of distributions in $C^{-\infty}(K)$ converges weakly iff it converges strongly.

   Preliminary: Compact operators. If $A : X \to Y$ is an operator between Banach spaces, we say A is *compact* if $\overline{A(B)}$ is compact for any bounded set $B \subset X$.

   Prime example: the inclusion $C_c^{k+1}(K) \to C_c^k(K)$ is compact.

   Proof. Suppose the sequence $\Lambda_n \to 0$ weakly. We claim there is a k and an M such that
$$|\Lambda_n(f)| \le M \|f\|_{C^k}$$
   for all smooth f. If not, then
$$U_M = \{f \; : \; \sup |\Lambda_n(f)| > M\}$$
   is a dense G_δ in $C_c^\infty(K)$, and thus $\bigcap U_M$ is nonempty; it contains say f. But then $\Lambda_n(f)$ is unbounded, so it cannot converge.

   Thus Λ_n is bounded in $C_c^k(K)^*$, for some k. But the unit ball in $C_c^{(k+1)}(K)$ is compact in $C^k(K)_c$, so from pointwise convergence of Λ_n plus equicontinuity (boundedness) on the unit ball in $C_c^k(K)$, we obtain uniform convergence on the unit ball in $C_c^{k+1}(K)$, and hence norm convergence in $C_c^{k+1}(K)^*$. ∎

Reference: [Tr, p.22].

10. Corollary: 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 $g \in C_c^\infty(\Omega)$. Then $f_i g \to gf$ in every C^k; since the Λ_i are *uniformly* bounded, we have

$$|\Lambda_i(f_i g - fg)| \to 0.$$

By weak convergence, $\Lambda_i(fg) \to \Lambda(fg)$, and thus $\Lambda_i(f_i g) \to \Lambda(fg)$, which shows $\Lambda_i \to \Lambda$. ∎

11. Remark: the weak* and strong topologies 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.

12. *The sheaf of distributions.* Given an open set $U \subset V$, we have $C_c^\infty(U) \subset C_c^\infty(V)$ (extend by zero) and hence $C^{-\infty}(U) \to C^{-\infty}(V)$.

Using these restriction maps, the distributions become a *sheaf.* (The sheaf axioms follow using a partition of unity.) We can define the sheaf of distributions on a *smooth manifold* M^n by setting $C^{-\infty}(U)$ equal to the dual to the smooth, compactly supported *n-forms* on U.

The cohomology of the sheaf of distributions is trivial.

13. *Supports.* We say $\operatorname{supp}\Lambda = F$ if F is the smallest closed set such that $\Lambda|\Omega - F = 0$.

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

$$\Lambda(f) = \Lambda(\psi f)$$

for any smooth function f with $f = 1$ on $\operatorname{supp}\Lambda$. It follows that there exist k and M such that

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

for all $f \in C_c^\infty(\Omega)$.

14. *Skyscrapers.* Theorem. 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 f vanishes to order N at 0, meaning $D^\alpha f = 0$ for $|\alpha| \le N$. Then for $x \in B(0,r)$ we have:

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

and more generally

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

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

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

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

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

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

15. Theorem. A positive distribution is a measure.

16. Theorem. Any compactly supported distribution is a finite sum of derivatives of 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. ■

17. Theorem. Any compactly supported distribution is a finite sum $\sum D^\alpha f_\alpha$ 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 $F(x) = \mu[x, \infty)$ we find

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

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

$$\int f(x_1, \ldots, x_n)\, d\mu(x) = \int (D^1 \cdots D^n f)(x) F(x)\, dx.$$

Doing it twice we again get up to a continuous function. ∎

18. *Convolutions.* We define, for $f, g \in L^1(\mathbb{R}^n)$,

$$(f * g)(y) = \int f(x) g(y-x) dy \int_{a+b=y} f(a) g(b) da = (g * f)(x).$$

19. *Group rings.* Let G be a multiplicative group (not necessarily commutative), and let A be a ring (often $A = \mathbb{Z}$). To make G into an *algebra* over A we consider the *group ring* $A[G]$. Its elements are formal finite sums $f = \sum a_g \cdot g$; they can be thought of as maps $f : G \to A$ 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 representations of G correspond to modules over $\mathbb{Z}[G]$. 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.

20. *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, $f * f * f * \cdots * f$.

21. Convolution is associative:

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

    By Fubini's theorem we find $\|f * g\|_1 \leq \|f\|_1 \|g\|_1$. Thus $(L^1, *)$ is a *Banach algebra*.

    Note that $f * g$ is generally not continous; for example, if $f \in L^1(\mathbb{R}^n) - L^2(\mathbb{R})$ and $g(x) = f(-x)$, then $f * g(y)$ blows up at $y = 0$.

22. We can regard $f * g$ as a limit of convex combinations of translates of f. Thus if $f \in B$, a Banach space with translation acting continuously, preserving norm, then $f * g \in B$ as well, and

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

    This means $f * g$ tends to inherit the good properties of both f and g.

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

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

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

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

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

24. $A-A$. 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.

25. *Convolutions with distributions.* By analogy with $\Lambda f = \Lambda(f)$, for a distribution Λ and an $f \in C_c^\infty(\mathbb{R}^n)$ we define

$$(\Lambda * f)(x) = \int \Lambda(y) f(x-y)\, dy = \Lambda_y(f(x-y)).$$

Note that $\Lambda * f$, by definition, is a function. In fact, since translation is a continuous operation on $C_c^\infty(\mathbb{R}^n)$, the convolution $\Lambda * f$ is a continuous function. And then one easily sees that it is a smooth function, in fact:

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

26. We can also define $f * \Lambda$, by analogy with $f\Lambda$, to be another distribution:

$$(f * \Lambda)(g) = \Lambda(\check{f} * g),$$

where $\check{f}(x) = f(-x)$. This definition is motivated by the formal calculation when Λ is a function.

Theorem: $f * \Lambda = \Lambda * f$; i.e. the distribution $f * \Lambda$ is represented by the smooth function $\Lambda * f$.

Proof. We have:

$$\begin{aligned}(f * \Lambda)(g) &= \Lambda_x\left(\int f(y-x)g(y)dy\right) = \int (\Lambda_x f(y-x))g(y)dy \\ &= \int (\Lambda * f)(y)g(y)dy = (\Lambda * f)(g).\end{aligned}$$

Here we use continuity and linearity of Λ on the compact set of translates $f(x-y)$ of f, $y \in \operatorname{supp} g$.

27. *Smooth functions are dense.* Corollary: $C^\infty(\mathbb{R}^n)$ is dense in $C^{-\infty}(\mathbb{R}^n)$.

Proof. Letting $\psi_r(x) = r^{-n}\psi(r^{-1}x)$ as $r \to 0$, where $\int \psi = 1$, we have $\psi_r * f \to f$ for all smooth functions. Therefore $\psi_r * \Lambda \to \Lambda$. But $\Lambda * \psi_r$ is smooth, so by the result above Λ is a limit of smooth functions. ∎

Remark: Also $C_c^\infty(\Omega)$ is dense in $C^{-\infty}(\Omega)$.

28. *Translation invariant operators.* Theorem. Any continuous translation invariant operator
$$L : C_0^\infty(\mathbb{R}^n) \to C_0^\infty(\mathbb{R}^n)$$
is given by $Lf = \Lambda * f$, some distribution Λ.

Proof. Set $\Lambda(f) = (Lf)(0)$. ∎

29. *Convolution of distributions.* Using the last result, we can define $\Lambda_1 * \Lambda_2$ to be the unique distribution such that
$$(\Lambda_1 * \Lambda_2) * f = \Lambda_1 * (\Lambda_2 * f).$$
Fact: this operation is associative so long as we work with compactly supported distributions; otherwise it need not be!

4 Fourier transforms

1. First motivation for the Fourier transform: was to express the unitary action of translation on $L^2(\mathbb{R}^n)$ as multiplication by a function of modulus 1.

2. Second motivation: How to make the Banach algebra $L^1(\mathbb{R}^n, *)$ look like a commutative algebra of functions?

   First note that the *point evaluations* in $C(X)$, X a compact Hausdorff space, correspond to the multiplicative linear functions $\phi : C(X) \to \mathbb{C}$: that is, those satisfying $\phi(fg) = \phi(f)\phi(g)$.

   Next note that if $\chi : \mathbb{R}^n \to \mathbb{C}^*$ is a group homomorphism, then:
$$\begin{aligned}\int (f * g)(x)\chi(x)\,dx &= \int f(x-y)g(y)\chi(x)\,dx\,dy = \int f(x)g(y)\chi(x+y)\,dx\,dy \\ &= \left(\int f(x)\chi(x)\,dx\right)\left(\int g(y)\chi(y)\,dy\right).\end{aligned}$$
   Finally note that $\chi(x)$ must be bounded to define a continuous linear functional on L^1, and so $\chi(x) = \exp(ixt)$.

3. We are thus lead to define the *Fourier transform* by

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

where $dm(x) = (2\pi)^{-n/2}\, dx$ is normalized volume measure on $\mathbb{R}^n$. Then by what we have just observed,

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

moreover $\widehat{f}(t) \in C(\mathbb{R}^n)$ (it inherits continuity from that of the characters), and $\widehat{f}(t) \to 0$ at infinity.

Summing up, the Fourier transform gives an algebra map

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

Also note that this satisfies

$$\|\widehat{f}\|_\infty \leq \|f\|_1,$$

where the L^1-norm is measured using $dm(x)$.

4. Theorem. Any multiplicative linear functional $\phi : L^1(\mathbb{R}^n) \to \mathbb{C}$ is given by integration against a unitary character $\chi : \mathbb{R}^n \to \mathbb{C}^*$; where

$$\chi(x) = \exp(ixt)$$

for some $t \in \mathbb{R}^n$.

Proof. Pick f such that $\phi(f) = 1$ and observe that for any p, $(\delta_p * f)(x) = f(x+p)$. Define

$$g(p) = \phi(f(x+p)) = \phi(\delta_p * f).$$

Observe that g is a *continuous* function of p. On the other hand, approximating δ_p by L^1 functions we conclude that

$$g(p) = \phi(\delta_p)\phi(f)$$

and thus $\phi(f) = \int fg$. Then from the fact that $\delta_p * \delta_q = \delta_{p+q}$ we see that g is a *continuous homomorphism* of $\mathbb{R}^n$ into $\mathbb{C}^*$. Finally since g is bounded it is a unitary character as above. ∎

Exercise: Show that if $\chi : \mathbb{R} \to \mathbb{C}^*$ is simply a *measurable* homomorphism, then $\chi(x) = \exp(xt)$, some $t \in \mathbb{C}$.

5. *Basic properties.*

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

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

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

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

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

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

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

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

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

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

   N.B. Here we use the traditional D^α, not Rudin's special D_α.

6. We now seek a class of functions as small as possible, containing $C_c^\infty(\mathbb{R}^n)$, and closed under $\mathcal{F}$.

   First notice 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.

   On the other hand, $\widehat{f}$ is smooth since its derivatives are related to $\widehat{x^\alpha f}$. Similarly, all derivatives of f are in L^1, and so

$$P(t)\widehat{f}(t) \to 0$$

   at infinity for any polynomial t. (This is a typical manifestation of the reversal of large and small scales under $\mathcal{F}$.) Thus $\widehat{f}$ is a smooth function vanishing rapidly at infinity, and the same is true for all its derivatives.

7. *Schwartz functions.* We are thus lead to introduce the class $\mathcal{S}(\mathbb{R}^n)$ of Schwartz functions, a Fréchet space with the family of norms:

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

This rapid decay implies all the derivatives of f are integrable, so by the preceding discussion we have:

Theorem. $\mathcal{F} : \mathcal{S} \to \mathcal{S}$ is continuous.

Proof. To bound the sup-norm of $(1 + |t|^2)^N D^\alpha \widehat{f}$, it suffices to bound the L^1-norm of $|x^\alpha (1 - \Delta)^N f|$, and this in turn is controlled by the rapid decrease of f and all its derivatives of order up to $2N$. Thus $p_N(f)$ controls $p_N(\widehat{f})$. ■

8. *The normal distribution.* Can we find a function which is its own Fourier transform? What we notice is that the Fourier transform interchanges the operators d/dx and $x\cdot$. So if we have $df/dx + xf = 0$, then taking the Fourier transform we get $(-itf + id/dt)\widehat{f}(t) = 0$. Since f and $\widehat{f}$ satisfy the same 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 $f(x) = \exp(-x^2/2)$ 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.

It remains only to compare values at zero: but 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 $\int f(x)\, dx = \sqrt{2\pi}$ and 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.

9. *Probability motivation.* Note that if X and Y are independent random variables with distribution functions f and g, then $f * g$ is the distribution function of $X+Y$. The *central limit theorem* guarantees that for independent, identically distributed random variables with mean zero and bounded variance, the limit distribution of $(1/\sqrt{n})(X_1 + \cdots X_n)$ is Gaussian.

   If we call this limit random variable Z, and its distribution f, then $\sqrt{2}Z$ should have the same distribution as $Z_1 + Z_2$. This means:

$$2^{-1/2} f(2^{-1/2}x) = (f * f)(x).$$

   Taking Fourier transforms, we obtain:

$$\widehat{f}(\sqrt{2}t) = \widehat{f}(t)^2,$$

   which is satisfies by $\widehat{f}(t) = Ae^{-Bt^2}$.

10. *Uncertainty principle.* Note that by the homogeneity condition, if we take a more concentrated Gaussian then its Fourier transform becomes broader, and the δ-function (formally) has Fourier transform equal to 1 everywhere. This is a manifestation of the uncertainty principle.

    Note also that there is a *natural* volume form, indeed symplectic form, on $T^*\mathbb{R}^n$.

11. *The Inversion Formula.* Theorem. For any $f \in \mathcal{S}$, we have $\mathcal{F}^2(f) = \check{f}$, where $\check{f}(x) = f(-x)$.

    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 linear span of the Gaussians. By continuity we're done. ∎

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:

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

■

12. *The Plancherel Theorem.* The Fourier transform extends to an isometry $\mathcal{F} : L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$.

Proof. For $f \in \mathcal{S}$, we have

$$\|\mathcal{F}(f)\|_2^2 = (\mathcal{F}(f), \overline{\mathcal{F}(f)})(f, \mathcal{F}(\overline{\mathcal{F}(f)})).$$

Now notice that

$$\overline{\mathcal{F}(f)} = \overline{\int f(x)\exp(-ixt)\,dx} = \int \overline{f}(-x)\exp(-ixt)\,dx = \mathcal{F}(\check{\overline{f}}).$$

Thus

$$(f, \mathcal{F}(\overline{\mathcal{F}(f)})) = (f, \mathcal{F}(\mathcal{F}(\check{\overline{f}}))) = (f, \overline{f}) = \|f\|_2.$$

■

13. *The Plancherel Theorem.* 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 $\operatorname{supp} f \subset [0, M]$. Extend f to a periodic function with period M, and note that the functions

$$e_n(x) = \frac{\exp(2\pi i n x/M)}{\sqrt{M}}$$

form an orthonormal basis for such functions. Writing $f = \sum a_n e_n$, we have

$$a_n = \langle f_n, e_n \rangle = \frac{1}{\sqrt{M}} \widehat{f}(2\pi n/M),$$

where we use dx instead of $dm(x)$ to define the Fourier transform. Then

$$\|f\|_2^2 = \sum_{-\infty}^{\infty} |a_n|^2 = \frac{1}{M} \sum |\widehat{f}(2\pi n/M)|^2 \to 2\pi \int |\widehat{f}|^2$$

as $M \to \infty$. We obtain an isometry if we absorb the 2π into the factor of integration.

The Fourier coefficients a_n come abstractly from the map $L^2(S^1) \to L^2(\mathbb{Z})$, and the π enters because of the length of the circle.

14. *Additional properties of the Fourier transform.*

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

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

    (c) $\|\widehat{f}\|_2 = \|f\|_2$.

    (d) $\int \widehat{f}(x) g(x)\, dm(x) = \int f(x) \widehat{g}(x)\, dm(x)$.

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

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

15. *Poisson summation.* Theorem. For $f \in \mathcal{S}$ (and often for more general f), we have

$$\sum_{\mathbb{Z}} f(n) = \sqrt{2\pi} \sum_{\mathbb{Z}} \widehat{f}(2\pi n). \tag{4.1}$$

Remark. A more common and equivalent formulation is

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

with the normalization

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

Proof. The function $F(x) = \sum f(x+n)$ has period 1; it arises from the pushforward $\mathbb{R} \to S^1$. Similarly there is an adjoint pullback on Fourier transforms, $\mathbb{Z} \to \mathbb{R}$, and this formula comes from evaluating F at zero.

More concretely, $F(x) = \sum a_n e_n$, where $e_n = \exp(2\pi i n x)$ are an orthonormal basis (with respect to dx-measure). We have $a_n = \sqrt{2\pi}\widehat{f}(2\pi n)$ (because of our normalized measure), and thus $F(0) = \sqrt{2\pi}\sum \widehat{f}(2\pi n)$.

∎

16. *Jacobi's theta function.* Theorem. The function

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

satisfies the remarkable identity:

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

Proof. Setting

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

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

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

∎

17. *Automorphic forms.* 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:

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

Thus $\vartheta(z)$ 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

$$\vartheta(z)^k = \sum a_n(k) q^{n/2}$$

where $a_n(k)$ is the number of ways to express n as a sum of squares of k integers.

Compare [Ser, Chapter VII.6].

18. *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* $A : H \to A$; the expected value of A is $\langle Af, f\rangle$.

Two of the most important observables are:

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

If we work in coordinates where the reduced Planck's constant $\hbar = 1.0546 \times 10^{-34} m^2 kg/s = 1$, then we have $\widehat{P(f)} = t\widehat{f}(t)$, so it looks just like Q in these coordinates.

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 $P_2(x)$.

19. *Gaussians.* The behavior of f and $\widehat{f}$ is especially easy to see when $f_a(x) = 2^{1/4}\sqrt{a}\exp(-(ax)^2/2)$ is a Gaussian distribution normalized to have $\|f_a\|_2 = 1$. Noting that $\widehat{f_1} = f_1$, we have

$$\widehat{f_a}(t) = a^{-1/2}\widehat{f_1(x/a)}(t) = a^{-1/2} a f_1(at) = f_{1/a}(t).$$

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

Note well! The probability distribution of a particle in a state given by a Gaussian is *still* itself Gaussian! It is given by Gaussian: $|f(x)|^2\, dm(x)$, and $|e^{-x^2/2}|^2 = e^{-x^2}$.

20. *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 no momentum! That is, $\langle P(f), f\rangle = 0$. 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.

21. *Uncertainty principle.* Theorem. 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. ∎

22. *Uncertainty principle: variation.* For a second version of 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. 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:

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

■

23. *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:

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

24. Temperment is a growth condition at infinity. Thus we have the following examples of tempered distributions.

    (a) Any compactly supported distribution. In particular, differential operators at a point, such that $u = D^\alpha \delta$.

    (b) Any finite positive measure, or more generally a measure that $\int (1 + |x|^2)^{-N}\, d\mu < \infty$ for some N.

    (c) Any function $g \in L^p(\mathbb{R}^n)$ for some p, $1 \leq p \leq \infty$.

    (d) More generally, any g with $(1 + |x|^2)^{-N} g$ in L^p.

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

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

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

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

Thus we define, for $u \in \mathcal{S}'$ and $f \in \mathcal{S}$,

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

Clearly $\widehat{u}$ is *also* a tempered distribution.

27. *Warning!* One can consider ordinary functions $f(x)$ as distributions via

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

We must use normalized measure, not ordinary dx.

28. Examples.

(a) 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, $\widehat{1} = \delta_0$.

(b) 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.

(c) Similarly, if $u = P(x)$, then $\widehat{u} = P(it)$.
Theorem. u is a polynomial iff $\widehat{u}$ is supported at one point, and vice-versa.

29. As before,we define $(u * f)(y) = u_x(f(y - x))$. For u tempered and f Schwartz, $u * f$ is a C^∞ function with polynomial growth. The usual algebraic relations extend, including $\widehat{u * f} = \widehat{u}\widehat{f}$.

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

Its inverse is given by

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

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

32. 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.)

33. The next Lemma shows we can speak unambiguously about the analytic extension of f when one exists.

    Lemma. 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. ■

34. Theorem. (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

$$|\widehat{f}(t)| \leq C_N \frac{e^{R|\operatorname{Im} t|}}{(1+|t|^2)^N} \tag{4.2}$$

    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 (4.2).

Now suppose $\widehat{f}(t)$ satisfies (4.2). 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 $x < 0$.

■

35. Theorem. 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,

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

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

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

36. *Sobolev spaces.* The 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. An operator $A : \mathcal{S} \to \mathcal{S}$ is of *order* t if it maps H^s to H^{s-t} continuously, for every s.

(a) $L^2(\mathbb{R}^n) = H^0$.

(b) For $f \in \mathcal{S}$, the operator $A_f(u) = fu$ is of order 0. (Since $\widehat{f} \in L^1$, we have $\|\widehat{f} * \widehat{u}\|_{L^2_s} \leq \|\widehat{f}\|_1 \|\widehat{u}\|_{L^2_s}$.)

(c) The operator $A(u) = D^\alpha(u)$ is of order $|\alpha|$.

(d) Every compactly supported distribution is in H^s for some s (often negative). (Proof: compactly supported continuous functions are in H^0, and every compactly supported distribution is a finite sum of derivatives of continuous functions.)

(e) 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.

37. Sobolev theorem: if f is in $H^{n/2+\epsilon}$, then f is continuous.

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 $s > n/2$.

38. Sobolev theorem, smooth version: if f is in $H^{p+n/2+\epsilon}$, then f is in $C^p(\mathbb{R}^n)$.

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

39. *Consistency of derivatives* Suppose $D^\alpha f$ is continuous, as a distribution, for all $|\alpha| \leq N$. Then we can write $f = \lim_{r \to 0} f * \psi_r$. The convolutions are smooth and their first N derivatives converge uniformly on compact sets, so $f * \psi_r \to f$ in the C^N topology. It follows that f is N-times differentiable and its distribution derivatives agree with its 'ordinary' derivatives.

40. *Example: L^2-derivatives with f not continuous.* Consider the function $f(z) = |\log|z||^\alpha$ in $\mathbb{C}$. Then the distributional derivative $|\nabla f|$ is proportional to $|\log|z||^{\alpha-1}/|z|$. This derivative is in $L^2(\mathbb{C})$ for $0 < \alpha < 1/2$, even though $f(z) \to 0$ at infinity (and hence is not continuous).

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

41. *The Heisenberg group $H(\mathbb{Z})$.* 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 $H(\mathbb{R})$.

42. *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 (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].

5 Elliptic equations

1. Fundamental solutions. Theorem. Any linear differential equation $P(D)u = f$ with constant coefficients has a (distributional) fundamental solution E, such that $P(D)E = \delta$.

2. Examples:

   (a) On $\mathbb{R}$, the fundamental solution to $DE = \delta$ is given by the Heaviside function $E(x) = \chi_{[0,\infty)}$.

   More generally, $(D - \alpha)E = \delta$ has as solution $E_\alpha = e^{\alpha x}\chi_{[0,\infty)}$.

   For $\operatorname{Re}\alpha < 0$, this *is* a tempered distribution. Its Fourier transform is just what a formal calculation would suggest: $\widehat{E_\alpha} = 1/(it - \alpha)$.

   For $\operatorname{Re}\alpha > 0$, we get a tempered solution by setting $E(x) = -H(-x)e^{\alpha x}$.

   (b) A general constant-coefficient equation $P(D)u = v$ on $\mathbb{R}$ can be solved as follows:

   Write $P(D) = \prod(D - \alpha_i)$; then
   A fundamental solution is $E = E_{\alpha_1} * \cdots * E_{\alpha_n}$.

   In fact:

$$P(D)E = (D - \alpha_1)E_1 * \cdots * (D - \alpha_n) * E_n = \delta * \cdots * \delta = \delta.$$

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

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

   (e) 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 $\Box u = 0$ is $f(x - t) + g(x + t)$.

   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 solutioin, change coordinates so

$\Box = D_x D_y$, and $E(x,y) = H(-x)H(-y)$. Then we have

$$\begin{aligned} \int f\Box E &= \int E\Box f \\ &= \int_{-\infty}^{0}\int_{-\infty}^{0} \frac{\partial f}{\partial x \partial y} \\ &= f(0). \end{aligned}$$

(f) *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.

(g) Exercise: 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.)

3. Not all linear PDE have solutions! The famous example of Hans Lewy in $\mathbb{C} \times \mathbb{R}$, namely

$$\frac{\partial u}{\partial \bar{z}} + iz\frac{\partial u}{\partial t} = f$$

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.

4. *Existence of fundamental solutions.* (Malgrange–Ehrenpreis.) We now show there exists a distribution u satisfying

$$P(D)u = \delta$$

for any constant coefficient linear PDE on $\mathbb{R}^n$.

A fundamental solution u must satisfy, for all $f \in C_0^\infty$,

$$f(0) = \int (P(D)u)f = \int (P(-D)f)u = u(P(-D)f).$$

To show u exists, we just need to show the map

$$P(-D)f \mapsto f(0)$$

is continuous on (the image of $P(-D)$) in $C_0^\infty(\mathbb{R}^n)$. If so, then by the Hahn-Banach theorem, this map will extend to a linear functional and hence to a distribution u.

5. *Convergence of entire functions.* To apply the Fourier transform, we complement the Paley-Wiener theorem as follows.

Theorem. If $f_i \to 0$ in $C_c^\infty(\mathbb{R}^n)$, then for any compact set K in $\mathbb{R}^n$ and any $N > 0$ we have

$$\sup_{t \in \mathbb{R}^n + iK} (1 + |t|^2)^N |\widehat{f_i}(t)| \to 0$$

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. ■

When this condition is satisfied, we say $\widehat{f_i} \to 0$ *rapidly near* $\mathbb{R}^n$.

6. *Passing to frequency space.* We now return to the continuity of $P(-D)f \mapsto f(0)$. Applying the Fourier transform, it suffices to prove the following:

> *Fix a polynomial $P(t) \neq 0$. Then if $P(t)\widehat{f_i}(t) \to 0$ rapidly near $\mathbb{R}^n$, then $f_i(0) = \int_{\mathbb{R}^n} \widehat{f_i}(t)\, dt \to 0$.*

We are thus reduced to a problem in complex variables.

7. *The one-dimensional case.* To prove this theorem in the one-dimensional case (on $\mathbb{R}$), we apply the maximum principle.

Choose a compact ball $K \subset \mathbb{C}$ containing all the zeros of P. Then for any entire function $g(t)$, we have

$$\sup_K |g| \le (\inf_{\partial K} |P|)^{-1} \sup_{\partial K} |Pg|.$$

In particular, $\widehat{f_i} \to 0$ uniformly on K. On the other hand, $|P|$ is bounded below outside K, and thus $\widehat{f_i}(t) \to 0$ rapidly near $\mathbb{R}$. In particular $\int_{\mathbb{R}} \widehat{f_i} \to 0$ (since $\int (1 + |t|^2)^{-1} < \infty$). ■

8. *The n-dimensional case.* For the case of $\mathbb{R}^n$, we will establish that for any entire function $g(t)$, we have

$$|g(t)| \le C_P \int_{|z|=1} |(Pg)(t+z)|\,|dz|.$$

(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. 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

$$\begin{aligned} |g(0)| &= \frac{|Q(0)g(0)|}{|c|} \le \frac{1}{2\pi|c|} \int_{|z|=1} |Q(z)g(z)|\,|dz| \\ &= C_P \int_{|z|=1} |P(z)g(z)|\,|dz| \end{aligned}$$

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 $n = 1$.

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

$$2\pi|c(L)g(t)| \le \int_{S^1(L)} |(Pg)(z+t)|\,|dz|;$$

integrating over projective space, we get

$$|g(t)| \int_{\mathbb{P}^{n-1}} |c(L)| \le C_n \int_{|z|=1} |(Pg)(z+t)|\,|dz|,$$

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.

9. *Elliptic equations.* Consider a linear partial differential operator $P(D)$ of order $N \geq 0$ on an open set $\Omega \subset \mathbb{R}^n$. This means

$$P(D, x) = \sum_{|\alpha| \leq N} a_\alpha(x) D^\alpha$$

with $a_\alpha(x) \in C^\infty(\Omega)$, and the highest-order part

$$P_N(D, x) = \sum_{|\alpha| = N} a_\alpha(x) D^\alpha \neq 0.$$

We say $P(D, x)$ is *elliptic* if for every x_0, the homogeneous polynomial

$$P_N(it, x_0) = \sum_{|\alpha| = N} a_\alpha(x_0)(it)^\alpha$$

has no zeros for $t \in \mathbb{R}^n - \{0\}$. In other words, $P_N(it, x_0) : \mathbb{R}^n \to \mathbb{R}$ is *proper*.

10. *Elliptic regularity.* Theorem. Let $P = P(D, x)$ be an elliptic operator of order N on Ω. Then if $u \in C^{-\infty}(\Omega)$ satisfies

$$Pu = 0,$$

then in fact u is smooth ($u \in C^\infty(\Omega)$).

More generally, if v is locally in H^s and

$$Pu = v,$$

then u is locally in H^{s+N}.

We will prove the elliptic regularity theorem under the simplifying assumption that the *principal symbol* $P_N(D, x)$ has constant coefficients; that is, $a_\alpha(x)$ is a constant for $|\alpha| = N$.

11. *Example: the wave operator.* The operator $P(D) = \Box = d^2/dx_2^2 - d^2/dx_1^2$ has principal symbol $P(it) = t_2^2 - t_1^2$, 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.

12. *Example: the Laplacian.* To treat a simple but important case, consider the Laplacian

$$P(D) = \Delta = \sum \frac{\partial^2}{\partial x_i^2}.$$

Then $P(it) = -\sum t_i^2 < 0$ for $t \neq 0$, so $P(D)$ is elliptic.

Now suppose u and v are distributions with compact support, $u \in H^t$ and $v \in H^s$, and

$$\Delta u = v.$$

Then $(I - \Delta)u = u - v \in H^{\min(s,t)}$. On the other hand, applying the Fourier transform we have

$$\widehat{(I - \Delta)u} = (1 + \sum t_i^2)\widehat{u} = \widehat{u} - \widehat{v}.$$

Therefore:

$$u = (1 - \Delta)^{-1}(u - v)$$

is in $H^{\min(s+2,t+2)}$, because the operator $(I - \Delta)^{-1}$ is smoothing of order 2. It follows then that $u \in H^{s+2}$, i.e. u is two derivatives smoother than v.

In particular, if v is smooth then so is u.

13. *Smoothing operators.* Quite generally we note that any $\widehat{Q}(t) \in L^\infty(\mathbb{R}^n)$ defines an operator of order zero on all the Sobolev spaces, characterized by:

$$\widehat{Qf} = \widehat{Q}(t)\widehat{f}(t).$$

Moreover, if

$$|t|^N \widehat{Q}(t) \in L^\infty(\mathbb{R}^n),$$

then Q is smoothing of order N.

The idea of elliptic regularity is to find an operator Q of order zero such that $(P + Q)^{-1}$ is smoothing. For $P = \Delta$ we can take $Q = -I$ as we have just seen.

14. *Example: The $\overline{\partial}$ operator.* The case of the Laplacian is particularly simple because $P_N(t)$ is *real*. For complex operators, some more work is required to obtain an operator whose inverse is smoothing.

The prime example of a complex operator is

$$P(D)f = \frac{\partial f}{\partial \overline{z}} = \frac{1}{2}\left(\frac{\partial f}{\partial x_1} + i\frac{\partial f}{\partial x_2}\right),$$

where $z = x_1 + ix_2$. This operator annihilates holomorphic functions.

If we identify $\mathbb{R}^2$ with $\mathbb{C}$, then $\widehat{P}(t) = it/2$. Clearly $\widehat{P}(t) + C$ has a zero in $\mathbb{C}$ for any constant C. To obtain an invertible function (or at least a function whose inverse is bounded), we set

$$\widehat{Q}(t) = \frac{\widehat{P}(t)}{|\widehat{P}(t)|}.$$

Then $\widehat{Q}(t)$ and its inverse are operators of order 0, corresponding to operators Q and Q^{-1} on the Sobolev spaces H^s.

On the other hand,

$$(\widehat{P}(t) + \widehat{Q}(t))^{-1} = \frac{|\widehat{P}(t)|}{\widehat{P}(t)(1 + |\widehat{P}(t)|)}$$

is a bounded function, behaving like $|t|^{-1}$ as $t \to \infty$. Thus $(P+Q)^{-1}$ is smoothing of order 1.

(If $\widehat{P}(t)$ is homogeneous and elliptic of order N, then $(P+Q)^{-1}$ is smoothing of order N.)

Finally suppose we have compactly supported distributions u and v in H^t and H^s such that

$$\overline{\partial} u = v.$$

Then we have $(P+Q)u = v + Qu$, and thus

$$u = (P+Q)^{-1}(v + Qu).$$

Since Q has order 0, while $(P+Q)^{-1}$ is smoothing of order 1, we find $u \in H^{s+1}$. That is, $t = \min(s+1, t+1)$ and thus u is one derivative smoother than v.

15. *Question:* where did the preceding argument use the fact that P is elliptic!?

    Answer: to know that $1/|P|$ behaves like $|t|^{-N}$ at infinity. Otherwise we would just conclude that $(P+Q)^{-1}$ is an operator of order 0, yielding $t = \min(s+t, t)$ which gives no information on t.

16. Theorem. Let $P_N(D)$ be homogeneous elliptic operator of degree N, with constant coefficients, and suppose u and v are compactly supported distributions satisfying

$$P_N(D)u = v.$$

Then if $v \in H^s$, we have $u \in H^{s+N}$.

(Established by the argument above.)

17. *Commutators.* Theorem. If $P(D,x)$ has order N, and f is smooth, then the operator

$$[P(D,x), f]u = P(D,x)(fu) - fP(D,x)u$$

has order $N-1$.

Proof. By Leibniz's rule, the terms in $P(D,x)(fu)$ other than $fP(D,x)u$ all involve lower order derivatives of u, multiplied by smooth functions. ∎

18. *Localization.* We now remove the hypothesis of compact support, and homogeneity of $P(D,x)$. (But we continue to assume the principal symbol $P_N(D)$ has constant coefficients and is elliptic.)

Let $u, v \in C^{-\infty}(\Omega)$ satisfy

$$P(D,x)u = v,$$

where v is locally in H^s. (This means $fv \in H^s$ for any compactly supported smooth f.) Shrinking Ω slightly, we can also assume that u is locally in H^t for some t.

Given $f \in C_c^\infty(\Omega)$, let us now estimate the smoothness of fu. Writing $P(D,x) = P_N(D) + R(D,x)$, where $R(D,x)$ has order $N-1$, we have

$$\begin{aligned} P_N(D)fu &= (P(D,x) - R(D,x))(fu) \\ &= fP(D,x)u + [P(D,x), f]u - R(D,x)(fu) \\ &= fv + Q(D,x)u, \end{aligned}$$

where $Q(D,x)$ is a compactly supported operator with order $N-1$.

Now we have an equation where both sides are compactly supported. The term $fv + Q(D,x)u$ has order $\min(s, t-N+1)$, so fu has order $\min(s+N, t+N)$, by the compact case of elliptic regularity. It follows that $t = \min(s+N, t+1)$, and thus u is locally in H^{s+N}. ∎

19. *Elliptic regularity.* Theorem. 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

$$\sup_K |D^\alpha f| \le C(K, |\alpha|) \sup_\Omega |f|.$$

Proof. Let $S \subset C^0(\Omega)$ be the set of bounded, continuous solutions to $Pf = 0$. Then S is closed (any C^0 limit is a distributional solutions) and all elements of f are smooth. Thus we have a well-defined operator $D^\alpha : S \to C(K)$. Since all elements of S are smooth, D^α has a closed graph, and hence D^α is bounded. ∎

20. *Finiteness.* For a vector bundle $E \to M$ on a *compact* manifold, the sections of E satisfying an elliptic differential equation (locally) span a finite-dimensional space. Here is a concrete example:

    Theorem. 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.

21. *Serre duality and Riemann-Roch.* The only step in the proof of Riemann-Roch 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. $(H^{0,1}(X))^* = \Omega(X)$. ■

22. *Further Sobolev Theorems.* Theorem. Suppose $\nabla f \in L^{n+\epsilon}(\mathbb{R}^n)$. Then f is (Hölder) continuous.

Proof. Given $f \in C_0^\infty(\mathbb{R}^n)$, we can recover $f(x)$ from ∇f by integrating along radial lines. That is,

$$f(x) = (\nabla f) * c_n \frac{x}{|x|^n},$$

where $c_n = (\mathrm{vol}(S^{n-1}))^{-1}$ (e.g. $c_1 = 1/2$). For example, at $x = 0$ we have:

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

Now $x/|x|^n$ behaves like r^{n-1}, so 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 $p > n$, then p is dual to such a q, and $g_t = \nabla f(x+t)$ moves continuously in L^p. So by Hölder's inequality we find that $f(x) = c_n \nabla f * x/|x|^n$ is continuous (in fact Hölder continuous) if ∇f is in $L^p(\mathbb{R}^n)$, $p > n$. ■

23. *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.

24. *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.

$$(Tf)(y) = \lim_{r\to 0} \int |x| > rf(y-x)K(x)\,dx.$$

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

$$T : \mathcal{S}(\mathbb{R}^n) \to \mathcal{S}(\mathbb{R}^n).$$

25. *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 $0 < \alpha < 1$.

26. *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:
    $$\int \frac{e^{itx}}{x}\, dx = \pi i \operatorname{Res}_{x=0} \frac{e^{itx}}{x} = \pi i.$$
    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 $\operatorname{Im} z \to \infty$, 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 $\widehat{K}(t) = i$.

27. *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 $\mathbb{H} = \{t : \operatorname{Im}(t) > 0\}$. 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.

28. *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 $H(f) = -if$.

29. *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

$$F(z) = \log(z) : \mathbb{H} \to \{z \,:\, 0 < \operatorname{Im} z < \pi\}.$$

    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.

30. *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.)

31. *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 $i\overline{t}/2$ upon Fourier transform, we have

$$\widehat{Tf} = \frac{\overline{t}}{t}\widehat{f}.$$

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:

$$\|\partial f\|_p \leq C_p \|\overline{\partial} f\|_p.$$

32. *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.

33. *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.

34. *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

$$\left|\frac{\partial f}{\partial \overline{z}}\right| \le k \left|\frac{\partial f}{\partial z}\right|.$$

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,

$$\text{area}(f(U)) = \int_U |Jf| \ge C\int_U |Df|^2.$$

This is why it is naturally to require derivatives in L^2.

35. *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.

36. *The Uniformization Theorem.* Cor. For any smooth metric g on S^2, or even measurable metric with bounded eccentricity, there exists a *conformal* homeomorphsim $f : (S^2, g) \to \widehat{\mathbb{C}}$; in the sense that the derivatives of f are in L^2, and Df sends g-circles to standard tangent circles for $\widehat{\mathbb{C}}$.

6 The prime number theorem

References: Rudin, Chapter 9; Hardy [Har, Ch. 12]; Wiener [Wie].

1. Warmup. Let $A \subset C(S^1)$ be a closed sub-space, invariant under multiplication by z and z^{-1}. Then $A \subset A_t$ for some $t \in S^1$, where $A_t = \{f(z) : f(t) = 0\}$.

   Proof. A 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. ■

2. Wiener's Theorem. Let $A \subset L^1(\mathbb{R}^n)$ be a closed, translation-invariant subspace. Then either $A = L^1$ or $A \subset A_t$, the set of L^1 functions with $\widehat{f}(t) = 0$.

3. Proof by approximation. Suppose $A \neq L^1$, then (by Hahn-Banach) there is a function $h \in L^\infty$ such that $\int hf = 0$ for all $f \in A$. Let t belong to the support of the tempered distribution $\widehat{h}$. We will show $A \subset A_t$.

   If not, there is an $f \in A$ with $\widehat{f}(t) = 1$. By multiplying A by e^{-ixt} (which preserves translation invariance) we can assume $t = 0$. We wish to construct a function in or near A whose Fourier transform is supported close to t and pairs nontrivially with $\widehat{h}$, to obtain a contradiction.

   STEP 0. The naive strategy is this: $\widehat{f}(0) = 1$ so $\widehat{f}$ is non-vanishing on a neighborhood of $t = 0$. Pick a smooth function $\widehat{\psi}$ supported in that neighborhood such that $\widehat{h}(\widehat{\psi}) = 1$. Then $\widehat{\psi} = \widehat{f} \cdot (\widehat{\psi}/\widehat{f})$.

   Since A is closed under translation, it is closed and convolution, and thus $\widehat{A}$ is closed under multiplication, so we have a function in A on which h is nonzero.

Unfortunately, $\widehat{A}$ is only closed under multiplication by functions $\widehat{g}$ with $g \in L^1$. This is a tricky condition to verify — in particular there is no reason it should hold for $\widehat{\psi}/\widehat{f}$ (which, while compactly supported, is not smooth). An additional finesse is required.

STEP 1. Suppose $f \in L^1$ and $\widehat{f}(0) = 1$. Then there is a g close to f in L^1 such that $\widehat{g} = 1$ identically on a neighborhood of $t = 0$.

Proof. Let h be a Schwartz function such that $\widehat{h} = 1$ on a neighborhood of $t = 0$. Then the same holds true for the functions

$$h_a(x) = a^n h(ax).$$

As $a \to 0$, these functions spread out on $\mathbb{R}^n$, while $\widehat{h}_a$ focuses near zero (in fact $\widehat{h}_a(x) = \widehat{h}(x/a)$.

Now for a near 0, we set

$$g = f - f * h_a + h_a$$

so that

$$\widehat{g} = \widehat{f}(1 - \widehat{h}_a) + \widehat{h}_a = 1$$

near $t = 0$. Then we have

$$\|g - f\|_1 = \|h_a - f * h_a\|_1.$$

If we rescale, setting $f_a(x) = a^{-n} f(x/a)$, we find

$$\|g - f\|_1 = \|h - f_a * h\|_1.$$

Recalling that $\int f = 1$, we see $f_a * h \to h$ in L^1, and thus for a small we get g close to f. ∎

STEP 2. Pick g as above with $\|g - f\|_1 < \epsilon$. Pick a function $\widehat{\psi}(t) \in C_0^\infty(\mathbb{R}^n)$ supported on a neighborhood of $t = 0$ where $\widehat{g}(t)$ is identically 1, and normalized so $\widehat{h}(\widehat{\psi}) = 1$.

Since A is closed under translation, it is closed under convolution by L^1 functions. Thus for any $k \in L^1$ we have $h(k * f) = 0$. Since the L^1-norm of $(g - f)$ is bounded by ϵ, its n-fold convolution has norm bounded by ϵ^n, and thus $\psi * (g - f)^{*n} \to 0$ in L^1. Therefore:

$$h(\psi * (g - f)^{*n}) = h(\psi * g^{*n}) \to 0$$

as $n \to \infty$. On the other hand, taking Fourier transforms we have

$$h(\psi * g^{*n}) = \widehat{h}(\widehat{\psi} \cdot g^n) = \widehat{h}(\widehat{\psi}) = 1$$

for all n.

By contradiction we conclude that $A \subset A_0$. ■

4. Proof by Banach algebras. Since A is translation-invariant, it is closed under convolution; thus it is a closed subalgebra of L^1. Now the space of maximal ideals for $(L^1, *)$ is $\mathbb{R}^n$, so if $A \neq L^1$ then A is contained in a maximal ideal, A_y, for some y. ■

5. *Wiener's Tauberian Theorem.* For $f \in L^\infty(\mathbb{R})$, let $K_0 \in L^1$ be a kernel with $\int K_0 = 1$ and $\widehat{K_0} \neq 0$ everywhere, and suppose the smoothed function $f * K_0$ converges to A at infinity. Then $f * K$ also converges to A for any other kernel $K \in L^1$, $\int K = 1$.

   Proof. The set of kernels which work is closed, translation invariant and contains K_0.

6. *Convergence of $f(x)$ as $x \to \infty$.* Theorem. Let $f : \mathbb{R} \to \mathbb{R}$, let $K_0 \in L^1(\mathbb{R})$ be a kernel with $\int K_0 = 1$, and suppose:

   $f(x)$ is uniformly continuous (e.g. $f'(x) = O(1)$),
   $f * K_0(x) \to A$ as $x \to \infty$, and
   $\widehat{K_0}(t) \neq 0$ for all $t \in \mathbb{R}$.

   Then $f(x) \to A$ as $x \to \infty$.

   Proof: For K a bump function with support of size ϵ and $\int K = 1$, it is clear that $|f * K(x) - f(x)| = O(\epsilon)$, while $f * K(x) \to A$ by Wiener's theorem. Thus $f(x) \to A$. ■

7. Non-example. Let $f(x) = \sin(2\pi x)$, $K(x) = \chi_{[0,1]}(x)$. Then f is slowly varying and $f * K = 0$. But $f(x)$ does not tend to zero. The source of the problem is that $\widehat{K}(t)$ is equal to $C(e^{it} - 1)/t$, which has periodic zeros.

   Wiener's Theorem is a converse: if K 'detects all periodic oscillations', then convergence of f follows form that of $f * K$.

8. Classical Tauberian Theorems. Euler proved the famous 'formula' $1 + 2 + 3 + \cdots = -1/12$ by analytically continuing $\zeta(s)$ to $s = -1$. Similar (but easier) ideas allow us to try to evaluate $\lim s_n$, $s_n = \sum_1^n a_i$, even when this does not converge.

   Say $s_n \to S$ (A) (for Abel) if

$$\sum a_n r^n = (1-r) \sum s_n r^n \to S$$

   as $r \to 1$. Say $s_n \to S$ (C) (for Césaro) if

$$\frac{1}{N} \sum_1^N s_n \to S$$

   as $N \to \infty$. Here are two results which hold for any sequence s_n.

   (a) If $s_n \to S$ in the ordinary sense, then we also have $s_n \to S(A)$ and $s_n \to S(C)$.

   (b) 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.*

   Theorem. If $s_n \to S$ (A) and $a_n = O(1/n)$, then $s_n \to S$.

   Theorem. If $s_n \to S$ (A) and $s_n = O(1)$, then $s_n \to S$ (C).

   We will prove the second result using Wiener's theorem, after making a multiplicative change of coordinates (replacing the Fourier transform by the Mellin transform).

9. *The Mellin transform.* Replace $L^1(\mathbb{R})$ with $L^1(\mathbb{R}_+)$ using the map $t = e^x$. The translation-invariant measure dx then goes over to the multiplicatively invariant measure dt/t.

   We will be interested in convolutions $f * k$ where $k \in L^1(\mathbb{R}_+, dt/t)$ and $f \in L^\infty(\mathbb{R}_+)$. (We think of k as a smoothing kernel.) Such a convolution is given by

$$(f * k)(u) = \int_0^\infty f(t) k(ut^{-1})\, dt/t.$$

Now for the Fourier transform, we will *keep* the dual $\mathbb{R}$ as an additive group. Then the character $e^{-ixs} = (e^x)^{-is}$ goes over to t^{-is}. Thus the Fourier transform of $k(t)$ becomes the *Mellin transform*:

$$\widehat{k}(s) = \int k(t)\, t^{-is}\, dt/t.$$

Finally it is useful to consider the function $K(t) = t^{-1}k(t^{-1})$. Then $K(t)\, dt$ is the measure with which we are convolving:

$$\|k\|_1 = \int_0^\infty |K(t)|\, dt,$$

$$(f * k)(u) = \int_0^\infty f(t)K(ut)\, d(ut),$$

and

$$MK(s) = \widehat{k}(s) = \int_0^\infty K(t)t^{is}\, dt.$$

Example: if $K(t) = \chi_{[0,1]}$, then:

$$(f * k)(1/N) = (1/N) \int_0^N f(t)\, dt.$$

10. *The Gamma function.* The kernel $K(t) = e^{-t}$ 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 given by:

$$MK(s) = \int_0^\infty e^{-t}t^{is+1}\, dt/t = \Gamma(1 + is).$$

For later use, recall that $\Gamma(z)$ has no zeros (although it has poles at $z = 0, -1, -2, \ldots$). Also note that $\int K(t)\, dt = MK(0) = \Gamma(1) = 1$.

11. *The multiplicative Tauberian theorem.* Wiener's Tauberian theorem now yields:

Theorem. Suppose $\int_0^\infty |K_0(t)|\, dt$ is finite, $\int K_0 = 1$ and

$$MK_0(s) = \int K_0(t)t^{is}\, dt \neq 0$$

for all $s \in \mathbb{R}$.

Then, if $f(n) \in L^\infty(\mathbb{R}_+)$ satisfies

$$t \int_0^\infty f(n) K_0(nt)\, dn \to A$$

as $u \to 0$, then we also have

$$\frac{1}{N} \int_0^N f(n)\, dn \to A$$

as $N \to \infty$.

(The variable n is used to suggest sequences.)

Proof. This is just Wiener's theorem: convergence for the kernel $K_0(t)$ implies convergence for the kernel $K(t) = \chi_{[0,1]}(t)$. ■

12. *Proof of a classical Tauberian theorem.* Take $K(t) = e^{-t}$. Then as we have seen, the Mellin transform $MK(s)$ is nowhere vanishing.

    Set $f(n) = s_{[n]}$, and let $t = \log(1/r)$ where $0 < r < 1$. Then as $r \to 1$ we have $t \sim 1 - r \to 0$, and $K(nt) = r^n$; thus:

$$t \int_0^\infty f(n) K(nt)\, dn \sim (1 - r) \sum s_n r^n.$$

    Hence $s_n \to S$ (A) implies $N^{-1} \sum_1^N s_n \to A$, i.e. $s_n \to S(C)$.

13. *Generalization.* The same reasoning shows: if $s_n = O(1)$, $\int K = 1$, $K'(t)$ is bounded, the Mellin transform of K is nonvanishing, and

$$t \sum s_n K(nt) \to A$$

    as $t \to 0$, then

$$\frac{1}{N} \sum_1^N s_n \to A$$

    as well. The bound on $K'(t)$ insures that $K'(nt)$ varies slowly over intervals of length one, so $\int f(n) t K(tn)\, dn$ can be approximated by a sum.

14. *The zeta-function.* Notice that if $K_n(t) = K(nt)$, then

$$\begin{aligned} MK_n(s) &= \int K(nt)t^{is}\,dt \\ &= \int \frac{K(nt)(nt)^{is}\,d(nt)}{n^{1+is}} \\ &= \frac{(MK)(s)}{n^{1+is}}. \end{aligned}$$

Thus if $L(x) = \sum_1^\infty K(nx)$, we get

$$ML(s) = \sum_1^\infty \frac{1}{n^{1+is}} MK(s) = \zeta(1+is)MK(s).$$

As a formal example, if $K(t) = e^{-t}$, then

$$L(t) = \frac{1}{e^t - 1} = e^{-t} + e^{-2t} + e^{-3t} + \cdots$$

and, since $MK(s) = \Gamma(1+is)$, we get

$$ML(s) = \Gamma(1+is)\zeta(1+is).$$

Note however that $L(t)$ is not quite in L^1, since it behaves like $1/t$ near $t = 0$. Similarly $ML(s)$ has a pole at $s = 1$.

15. *Prime number theorem.* Let $\pi(n)$ denote the number of primes $p \leq n$. The prime number theorem, proved by Hadamard and de la Vallé Poussin, asserts that

$$\pi(n) \sim \frac{n}{\log n}.$$

Heuristically, the probability that n is prime is asymptotic to the reciprocal of the number of digits of n in base e. Since $log(10) = 2.30259\ldots$, this says a number near $n = 10,000$ should have probability about $1/(2.3 * 4) = 0.11$ of being prime. Indeed, the 1,230th prime is 10,007 and the the 1,336th is 11,003, giving a 'probability' of $106/1000 = 0.106$.

16. *The von Mangoldt function* $\Lambda(n)$. Now set $\Lambda(p^k) = \log p$ for prime powers, and $\Lambda(n) = 0$ otherwise. Then to prove the prime number

theorem, $\pi(n) \sim n/\log n$, it suffices to prove

$$\sum_1^N \Lambda(n) \sim N.$$

Proof. Indeed, suppose we have the result above. Pick a small $\epsilon > 0$. Then the same result holds for the sum from $n \in [N^{1-\epsilon}, N]$, since $N^{1-\epsilon}\log N \ll N$. On the other hand, $\log(n) = \log N$ to within a factor of $1+\epsilon$ if $n \in [N^{1-\epsilon}, N]$. Thus we have, for all N sufficiently large,

$$\pi(N) \leq (1+\epsilon)\frac{N}{\log N} + \pi(N^{1-\epsilon}) \sim (1+\epsilon)\frac{N}{\log N}$$

since $\pi(x) \leq x$.

Now the sequence $\Lambda(n)/\log(n)$, $n \leq N$, slightly overcounts primes, because we also get contributions from p^k, $k \geq 2$. However, the number of k which work is on the order of $\log N$, and the number of p is at most $\sqrt{N}$, so in fact these terms make a negligible contribution $O(N^{1/2}\log N)$ to the sum. Thus we have

$$\pi(x) \geq \sum_{N^{1-\epsilon}}^{N} \frac{\Lambda(n)}{\log N} - O(N^{1/2}\log N) \sim \frac{N}{\log N}.$$

∎

17. Use of factorization. Recalling that each $n \geq 1$ has a unique factorization into primes, we find

$$f(r) = \sum_1^\infty \Lambda(n)\frac{r^n}{1-r^n} = \sum_{p,k}(\log p)(r^{p^k}+r^{2p^k}+r^{3p^k}+\cdots) = \sum(\log n)r^n.$$

The sums converge for $|r| < 1$. We will be interested in letting $r \to 1$.

18. Asymptotics of $f(r)$: differentiation and passage to the multiplicative group.

To get an idea of the size of $f(r) = \sum_1^\infty(\log n)r^n$, we note that

$$\log n = \int_1^n \frac{dx}{x} = \sum_1^n 1/k - \gamma + O(1/n),$$

where $\gamma = 0.577216\ldots$ is Euler's constant. Replacing $\log n$ with $\sum_1^n (1/k)$ yields:

$$f(r) \sim (1 + r + r^2 + \cdots)(r + r^2/2 + r^3/3 + \cdots) = \frac{1}{1-r} \log \frac{1}{1-r}.$$

Setting $r = \exp(-t)$, we have $1 - r \sim t$ as $t \to 0$, and thus

$$f(e^{-t}) \sim \frac{1}{t} \log \frac{1}{t}.$$

This suggest that as $t \to 0$ we have:

$$-t \frac{d}{dt} t f(e^{-t}) \sim 1.$$

To verify this, we must show

$$-t \frac{d}{dt} t \left(\sum (-\gamma + O(1/n)) e^{-nt} \right) \sim 0.$$

To check this, first note that $t \sum e^{-nt} = t/(1 - e^{-t})$ is analytic at $t = 0$, so the term involving γ can be ignored. For the term involving $O(1/n)$, note that

$$-t(d/dt) t e^{-nt} = nt(t - 1/n) e^{-nt};$$

multipling by $O(1/n)$, taking absolute values and summing over n, we obtain:

$$O(t^2 \sum e^{-nt} + t \sum e^{-nt}/n).$$

The first term vanishes as $t \to 0$ as before, and the second behaves like $t \int_0^t ds/(1 - e^{-s}) \sim |t \log t|$, so it also vanishes. Thus $-t(d/dt) t f(e^{-t}) \to 1$ as $t \to 0$.

19. Applying the operator $-t(d/dt)t$ to the other side of the equation, we get

$$\sum_1^\infty \Lambda(n) t \frac{d}{d(nt)} \frac{-(nt)e^{-nt}}{1 - e^{-nt}} \sim 1$$

as $t \to 0$. Setting

$$K(t) = \frac{d}{dt} \frac{-te^{-t}}{1 - e^{-t}} = \frac{d}{dt} \frac{t}{1 - e^{t}}$$

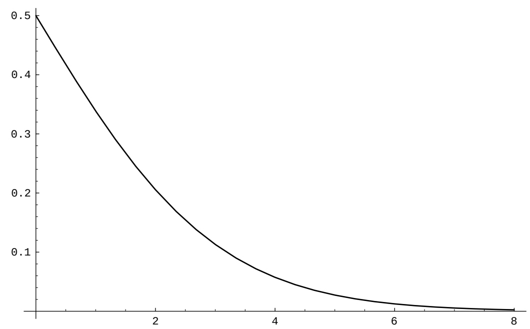

Figure 1. The kernel $K(t) = (d/dt)(t/(1-e^t))$.

we find:

$$\sum \Lambda(n) t K(nt) \to 1$$

as $t \to 0$. See Figure 1 for the graph of $K(t)$. Note also that $L(t) = t/(1-e^t)$ satisfies $L(0) = -1$, $L(\infty) = 0$, so $\int L'(t)\,dt = \int K(t)\,dt = 1$.

20. *The zeta function enters.* Next we look at the Mellin transform $MK(s)$. We first look at functorial properties of $MK(s) = \int_0^\infty K(t) t^{is}\,dt$. We get:

$$M(tK(t))(s) = \int K(t) t^{1+is}\,dt = (MK)(s-i);$$

and

$$M(K'(t))(s) = -\int isK(t) t^{is-1}\,dt = -is(MK)(s+i).$$

Thus starting with $F(t) = 1/(e^t - 1)$, for which we computed

$$MF(s) = \Gamma(is+1)\zeta(1+is),$$

we find $K(t) = -(d/dt)(tF(t))$ satisfies

$$MK(s) = is\Gamma(is+1)\zeta(1+is).$$

21. *Zeros of zeta with* $\operatorname{Re} s = 1$. Now we finally appeal to some well-known properties of Γ and ζ:

(a) $\Gamma(s)$ has no zeros, and poles only at the negative integers;
(b) $\zeta(1+s) = 1/s + O(1)$ near $s = 0$; and most importantly,
(c) $\zeta(1+is) \neq 0$ for $s \in \mathbb{R}$.

Combining these facts we conclude:

$MK(s) \neq 0$ for $s \in \mathbb{R}$.

(Note that $s\zeta(1+is) = 1$ at $s = 0$.)

22. *Diffusing the primes.* At this point we see $K(t)$ satisfies the hypothesis of the Tauberian theorem, and

$$\sum \Lambda(n) t K(tn) \to 1$$

as $t \to 0$. We wish to conclude that

$$\frac{1}{N} \sum_{1}^{N} \Lambda(n) \to 1.$$

to obtain the prime number theorem.

Unfortunately the sums above are not quite convolutions, and in any case $\Lambda(n)$ is not bounded. However $\Lambda(n)$ and $K(t)$ are both positive. Given $\epsilon > 0$, let

$$L(t) = \frac{1}{\epsilon)} \chi_{[1,1+\epsilon]}(t).$$

Define:

$$\Lambda_\epsilon(N) = \frac{1}{\epsilon N} \sum_{N}^{N+\epsilon N} \Lambda(n) = \sum \Lambda(n) t L(nt)$$

where $t = 1/N$. Since $L(t) = O(K(t))$, and since the above sum tends to one if we replace $L(t)$ with $K(t)$, we conclude that $\Lambda_\epsilon(N) = O(1)$.

But the bounded sequence $\Lambda_\epsilon(n)$ is just a smoothing of $\Lambda(n)$, so it also satisfies

$$\sum \Lambda_\epsilon(n) t K(nt) \to 1$$

as $t \to 0$. Consequently we have, by Wiener's multiplicative Tauberian theorem,

$$\frac{1}{N} \sum_{1}^{N} \Lambda_\epsilon(n) \to 1.$$

But if we unfold the sum above, we see $\Lambda(n)$ occurs about $n\epsilon$ times, each weighted by $(1/n)\epsilon$, so the sum above is well-approximated by $(1/N)\sum_1^N \Lambda(n)$ and we are done.

23. *Zeros of* ζ. To make 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

$$\sum \Lambda(n) n^{-\sigma} = \frac{1}{\epsilon} + O(1). \tag{6.1}$$

On the other hand, if ζ has a zero at $1 + it$, then

$$\sum \Lambda(n) n^{-\sigma} n^{-it} = \frac{-1}{\epsilon} + O(1). \tag{6.2}$$

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.

To see this last bit, the idea is just to think of (6.2) dividing by the normalizing factor (6.1) as the integral of $f(n) = n^{-it}$ against a probability measure dm on $\mathbb{N}$. Then $|f| = 1$ and $\int f\, dm = -1 + O(\epsilon)$ so $\|f+1\|_2^2 = O(\epsilon)$. That is, f is nearly the constant function -1. But then we expect $\int f^2$ to be nearly 1, and indeed

$$\int f^2\, dm = \langle f, f\rangle = \langle -1, -1\rangle + O(\|f+1\|_2) = 1 + O(\sqrt{\epsilon}).$$

This shows $(\zeta'/\zeta)(\sigma + 2it)$ grows like $1/\epsilon + O(1/\sqrt{\epsilon})$, and since ζ'/ζ is analytic at $1 + 2it$ we obtain a contradiction.

24. *Appendix: Weyl's law, the heat kernel and Tauberian theorems.* Theorem. Let M^n be a closed Riemannian manifold, and let $0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \cdots$ be the eigenvalues of $-\Delta$ on M. Then

$$N(\Lambda) = |\{k : \lambda_k \leq \Lambda\}| \sim C_n \operatorname{vol}(M^n)\Lambda^{n/2}.$$

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 $E(f) \leq \Lambda \|f\|^2$.

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

$$N(\Lambda) \geq c_n \operatorname{vol}(M)\Lambda^{n/2}.$$

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

$$\mathrm{Tr}(K_t) = \int_M K_t(x, x)\, dx = \sum e^{-\lambda_k t} \sim c_n t^{-n/2} \operatorname{vol}(M).$$

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:

$$\int_0^\infty f(t) u K(tu)\, dt = \sum u K(\alpha_k u) = t^{n/2} \sum e^{-t\lambda_k} \to c_n \operatorname{vol}(M).$$

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

$$MK(s) = \int \exp(-t^\alpha) t^i s\, dt = \int \exp(-u) u^{i\beta s} \beta u^\beta\, (du/u) = \beta\Gamma(\beta(1+is)) \neq 0$$

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

$$\int_0^N f(t)\, dt = |\{k : \alpha_k \leq N\}| \sim c_n' \operatorname{vol}(M) N,$$

which implies Weyl's law by taking $N = \Lambda^{n/2}$. ∎

7 Banach algebras

1. 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
$$\|ab\| \leq \|a\| \cdot \|b\|.$$

2. Examples:

   (a) $C(X)$, where X is a compact Hausdorff space. For X discrete we obtain $\mathbb{C}^n$.

   (b) $\ell^p(\mathbb{Z})$, $1 \leq p \leq \infty$, with pointwise multiplication.

   (c) $P(K)$, 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 $\operatorname{int}(K)$.

   (d) 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.

   (e) $L^1(\mathbb{R}^n)$ and $\ell^1(\mathbb{Z})$ with convolution.

   (f) $M(\mathbb{R}^n)$, the algebra of signed measures (with bounded total variation), under convolution.

   (g) The bounded operators on a Hilbert space, $\mathcal{B}(\mathcal{H})$. For a finite-dimensional Hilbert space we obtain the matrices $M_n(\mathbb{C})$.

   (h) $L^\infty(E)$, where $E \subset \mathbb{R}^n$ is measurable.

(i) $H^\infty(\Delta)$ — the functions on the unit disk with finite L^∞-norm.

(j) Non-examples: $L^p[0,1]$ with pointwise multiplication, $1 \le p < \infty$.

3. *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:

   (a) 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.*

   (b) For a compact set $K \subset \mathbb{C}^n$, the space X for $P(K)$ will be the *polynomial convex hull* of K.

   (c) 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 $X = \sigma(T)$.

4. *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 $\|T_a\| \ge \|a\|/\|1\|$, 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.

5. *Examples.* 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.

6. *Adjunction of 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.

7. *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 $S(T(ST)^{-1}) = I$.

8. *Topology of units.*

   Theorem. 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. 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. 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. ■

9. *The spectrum.* For any element x in an algebra $A/\mathbb{C}$, we define the *spectrum* of x (relative to A) as:

$$\sigma(x) = \{\lambda \in \mathbb{C} \,:\, (\lambda - x) \notin A^\times\}.$$

   The *spectral radius* is given by

$$\rho(x) = \sup\{|\lambda| \,:\, \lambda \in \sigma(x)\}.$$

Clearly $(\lambda - x) = \lambda(1 - x/\lambda)$ is invertible if $|\lambda| > \|x\|$. More generally, the power series converges if $\|\lambda^{-n}x^n\| < 1$ for $n \gg 0$. Thus we have the elementary:

Theorem: The spectrum $\sigma(x)$ is compact, and the spectral radius satisfies

$$\rho(x) \le \limsup \|x^n\|^{1/n} \le \|x\|.$$

We will later sharpen this result with an equality and show the limsup is a limit.

10. *Examples of the spectrum* . For $f \in A = C(X)$, the spectrum $\sigma(f) = f(X)$ detects the *range* of f.

    For $T \in M_n(\mathbb{C})$, the spectrum $\sigma(T)$ is the set of *eigenvalues* of T. Note that for $T = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ we have $\rho(T) = 1$ even though $\|T^n\| \asymp n$; this operator still satisfies $\lim \|T^n\|^{1/n} = 1$.

    The spectrum of an operator on Hilbert space is like a continuous version of the eigenvalues.

11. *Spectrum of the shift.* Example: 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 $\sigma(S) = \overline{\Delta}$.

12. Theorem. For any element x in a Banach algebra A, the spectrum $\sigma(x)$ is nonempty.

    This is the first main result in the subject of Banach algebras.

    Proof. For any $\phi \in A^*$, the dual space to A, consider the function

    $$f(\lambda) = \phi\left(\frac{1}{\lambda - x}\right).$$

    Then for λ, λ_0 in the open domain $\Omega = f^{-1}(\mathbb{C} - \sigma(x))$ we have

    $$\frac{f(\lambda) - f(\lambda_0)}{\lambda - \lambda_0} = \phi\left(\frac{-1}{(\lambda - x)(\lambda_0 - x)}\right) \to -\phi((\lambda - x)^{-2})$$

by continuity of inversion. Thus f is analytic. Moreover f tends to zero at infinity. Thus $f(\lambda) = 0$ identically; in particular, $\phi(-x^{-1}) = 0$ for all ϕ. But then $x^{-1} = 0$ by the Hahn-Banach theorem, which is obviously impossible. ■

13. *Quotient fields.* Now suppose A is a division algebra, i.e. suppose $A^\times = A - \{0\}$. Then for every $x \in A$, there is a $\lambda \in \mathbb{C}$ such that $\lambda - x$ is non-invertible, and hence $x = \lambda$. Therefore $A = \mathbb{C}$.

    Problem: Are the quaternions a Banach algebra (over $\mathbb{C}$)? (No! Because the center is $\mathbb{R}$, not $\mathbb{C}$.)

14. *Multiplicative linear functionals.* An *algebra* homomorphism $\phi : A \to \mathbb{C}$ is called a multiplicative linear functional. (By definition, $\phi(1) = 1$.)

    Such a map is *automatically* continuous, since it is nonzero on the open set $A^\times$.

    We can reformulate the previous discussion as:

    Theorem. Let A be a commutative Banach algebra. Then $x \in A$ is invertible iff $\phi(x) \neq 0$ for all multiplicative linear functionals.

    Proof. Clearly if x is invertible then $\phi(x)\phi(x^{-1}) = 1$ and hence $\phi(x) \neq 0$.

    Conversely, suppose x is not invertible. Then $xA \subset A$ is an ideal (since M is commutative), and hence $xA \subset M$ for some maximal ideal $M \subset A$ (whose existence follows from the axiom of choice). Then M is closed (since it is disjoint from $A^\times$, which is open), and $A/M \cong \mathbb{C}$ since it is a Banach division algebra. It follows that $M = \operatorname{Ker} \phi$ for some multiplicative linear functional ϕ, and $\phi(x) = 0$. ■

15. *Gelfand transform.* Let A be a commutative Banach algebra (with identity). Let $X = \operatorname{Spec} A \subset A^*$ be the set of multiplicative linear functionals on A. Then there is a natural, continuous algebra homomorphism, the *Gelfand transform*

    $$A \to \widehat{A} \subset C(X),$$

    given by $a \mapsto \widehat{a}$ where $\widehat{a}(\phi) = \phi(a)$.

We have

$$\widehat{a}(X) = \sigma(a) \quad \text{and}$$
$$\sup_X |\widehat{a}| = \rho(a) \leq \|a\|.$$

Proof. Immediate. ■

Example: $C^k[0,1] \subset C[0,1]$ is an example of the inclusion $A \subset \widehat{A}$. Note that the norm is strictly decreased at many points. (Of course we are assuming a sub-multiplicative norm on $C^k[0,1]$.)

16. *Properties of* $\widehat{A}$. Theorem. The algebra $\widehat{A}$ separates points, and $1/\widehat{a} \in \widehat{A}$ so long as $\widehat{a}$ is nonvanishing.

    Proof. If $\widehat{a}(\phi) = \widehat{a}(\psi)$ for all $a \in A$ then, by the definition of A^*, we have $\phi = \psi$. Thus $\widehat{A}$ separates points. Similarly, if $\widehat{a}(\phi)$ is never zero, then aA belongs to no maximal ideal, hence $aA = A$; thus a is invertible and therefore so is $\widehat{a}$. ■

    Note: we do *not* in general have $\widehat{A}$ closed under complex conjugation (consider algebras of analytic functions), and thus $\widehat{A}$ might not be dense in $C(X)$.

17. Example: Absolutely convergent Fourier series.

    Let $A = \ell^1(\mathbb{Z})$ under convolution, and let $e_n \in A$ be the sequence equal to 1 iat n and 0 elsewhere on $\mathbb{Z}$. Then $e_0 = \text{id}$ and $e_1^n = e_n$.

    Let $\phi : A \to \mathbb{C}$ be a continuous multiplicative linear functional with $\phi(e_1) = \lambda$. Then $\phi(e_n) = \lambda^n$, and since ϕ is bounded we have $|\lambda| = 1$. Thus the space of multiplicative linear functions on A is $X = S^1$, and the Gelfand transform

    $$\ell^1(\mathbb{Z}) \to \widehat{A} \subset C(S^1)$$

    is given by $a_n \mapsto \sum a_n \lambda^n$.

18. Theorem (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.

Note: it is very difficult to tell directly if a given continuous function $f \in C(S^1)$ has an absolutely convergent Fourier series, and thus this result stumped direct attempts.

Proof. Since $f = \widehat{a}$ is nowhere vanishing, its inverse $1/f$ also belongs to $\widehat{A}$. ■

Example. 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 $\sum |a_n| \leq C\|f\|_\infty$. 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.

19. Theorem (Wiener) Every maximal ideal in $A = L^1(\mathbb{R}^n) \oplus \mathbb{C}\delta_0$ is of the form $A_t = \{f : \widehat{f}(t) = 0\}$ or $A_\infty = L^1(\mathbb{R}^n)$.

    Proof. Every maximal ideal is the kernel of a multiplicative linear functional $\phi : A \to \mathbb{C}$. This $\phi | L^1(\mathbb{R}^n)$ is given by integration against an L^∞ function $g(x)$. If $g = 0$ then the ideal is $L^1(\mathbb{R}^n)$, so suppose $g \neq 0$.

    Now if $f_n \to \delta_x$, $h_n \to \delta_y$, then $f_n * g_n \to \delta_{x+y}$ while

$$g(x)g(y) = \lim \phi(f_n)\phi(g_n) = \lim \phi(f_n * g_n) = g(x+y)$$

    for a.e. x and y. Thus g defines a measurable multiplicative homomorphism from $\mathbb{R}^n$ to $\mathbb{C}^*$.

    Now take a neighborhood U of 1 in $\mathbb{C}^*$ and a smaller neighborhood V with $V \cdot V^{-1} \subset U$. Then $E = g^{-1}(V)$ has positive measure, since $\mathbb{C}^*$ is covered by countably many translates of V. On the other hand, $E - E$ contains an open neighborhood W of $x = 0$, by the Lebesgue density theorem. Thus $g(W) \subset g(E)g(E)^{-1} = V \cdot V^{-1} \subset U$. Thus g is continuous, from which it follows easily that $g(x) = \exp(ixt)$. ■

    Thus Wiener's theorem on translation-invariant subspaces can be restated as: any ideal $I \neq L^1(\mathbb{R}^n)$ is contained in a maximal ideal $M \neq L^1(\mathbb{R}^n)$. This would be immediate from the theorem above if we knew I were the intersection of all maximal ideals that contain it.

20. *Several complex variables.* Example. Let $K = S^1 \times S^1 \subset \mathbb{C}^2$. Let $A = P(K)$ be the uniform closure of the polynomial functions. Then the space of maximal ideals is $X = \overline{\Delta \times \Delta}$.

Indeed, every element of A extends holomorphically to X, and hence point evaluations in X give multiplicative linear functions on A. (By the maximum principle they are bounded.)

On the other hand, if $\phi : A \to \mathbb{C}$ is a multiplicative linear functional, then it restricts to a multiplicative linear functional on the polynomials ring $R = \mathbb{C}[z_1, \ldots, z_n]$. Once we know the values $\phi(z_i) = w_i$, we know ϕ; in other words, ϕ is given by a point evaluation.

But for every $w \notin X$ we can find a polynomial $p(z)$ (indeed a linear function) such that

$$\|p\| = \sup_{\Omega} |p| < |p(w)| = |\phi(w)|.$$

Thus ϕ is not a multiplicative linear functional on A. (In fact we can choose a linear function with $L(w) = 1$ such that $f(z) = (1 - L(z))^{-1}$ is in A, i.e. we can invert certain polynomials vanishing at w.)

21. *Polynomial hull.* More generally, for a compact set $K \subset \mathbb{C}^n$ we define its *polynomial hull* by

$$\widehat{K} = \{z \ : \ |p(z)| \leq \sup_{K} |p(w)| \text{ for all polynomials } p\};$$

then $A(K)$ has $\widehat{K}$ as its space of maximal ideals.

22. *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 $\widehat{K}$.

Open Question: if K is a union of disjoint balls in $\mathbb{C}^n$, does $\widehat{K} = K$?

23. *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. 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 $1 \in I$. ∎

24. *The Stone-Čech compactification.* 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. The sequence $x_n = n \in \mathbb{Z}$ has convergent subnets, but *no* convergent subsequence, in X.

    For an even more exotic space, try to visualize $\operatorname{Spec} L^\infty[0,1]$.

25. Calculation the spectral radius: *I*. How to calculate $\rho(a)$ for $a \in M_n(\mathbb{C})$? If a is diagonal, the entries of a^n grow like $\rho(a)^n$. For the same reason, if a is diagonalizable we have $\rho(a) = \lim \|a^n\|^{1/n}$.

    This is actually a reasonable algorithm! Although the convergence is slow; once should expect to acquire about 1 digit of accuracy per iteration, while Newton's method doubles the number of digits of accuracy per iteration.

26. *Spectral radius theorem.* Theorem. In any Banach algebra A, we have

$$\rho(a) = \lim \|a^n\|^{1/n}.$$

    (In particular the limit exists).

    Proof. We will show

$$\limsup \|a^n\|^{1/n} \le \rho(a) \le \liminf \|a^n\|^{1/n}.$$

    For the upper bound, observe that $\rho(a^n) = \rho(a)^n$ since the spectral radius is the sup-norm of the Gelfand transform, which satisfies $\hat{a^n} = \widehat{a}^n$. Thus $\rho(a) = \rho(a^n)^{1/n} \le \|a^n\|^{1/n}$.

    The lower bound is trickier; like the proof of non-emptiness of the spectrum, it involves the analytic function:

$$f(\lambda) = \phi((\lambda - x)^{-1}) = \frac{1}{\lambda}\sum_0^\infty \frac{\phi(x^n)}{\lambda^n}$$

    defined for any $\phi \in A^*$. By assumption, this function is analytic in the region $|\lambda| > \rho(x)$. Consequently the radius of convergence of the power series defining f is at least this large; this implies:

$$\limsup |\phi(x^n)|^{1/n} \le \rho(x).$$

Now we fix $\epsilon > 0$; then we have

$$\sup_n \left| \phi\left(\frac{a^n}{(\rho(a)+\epsilon)^n} \right) \right| \le M_\phi.$$

By the uniform boundedness principle, this implies:

$$\frac{\|a^n\|}{(\rho(a)^+\epsilon)^n} \le M_\epsilon$$

and hence $\limsup \|a^n\|^{1/n} \le \rho(a)+\epsilon$. Since $\epsilon > 0$ was arbitrary, the desired inequality follows. ■

27. *Spectral radius and subalgebras.* If $x \in B \subset A$, where B is a Banach subalgebra of A, then B has fewer units and thus $\sigma_B(x) \subset \sigma_A(x)$.

The inclusion can be strict. The function $f(z) = z$ has spectrum S^1 in $C(S^1)$ but spectrum $\overline{\Delta}$ in $A(\overline{\Delta}) \subset C(S^1)$.

Theorem. For $B \subset A$, $\sigma_B(x)$ is obtained from $\sigma_A(x)$ by filling in some bounded holes. More precisely, $\partial\sigma_B(x) \subset \partial\sigma_A(x)$.

Cor. The spectral radius of x is independent of the algebra B containing x.

Lemma. If $x_n \in B^\times$ converges to a non-invertible x, then $\|x_n^{-1}\| \to \infty$.

Proof. Otherwise, we would have $\|x_n^{-1}\| \le M$ along a subsequence; then for large n, $(x_n - x)$ is close to zero, and hence $x_n^{-1}x$ is close to $x_n^{-1}x_n = 1$, and thus x is invertible.

Proof of the Theorem. Obviously $B^\times \subset A^\times$; by the divergence property above, $\partial B^\times \subset \partial A^\times$, i.e. points in the boundary of the units cannot become invertible under passage to a larger algebra.

Since points in $\partial\sigma_B(x)$ correspond to such robustly non-invertible elements, we have $\partial\sigma_B(x) \subset \partial\sigma_A(x)$. ■

28. *The radical.* The (Jacobsen) *radical* of a Banach algebra A can be defined equivalently as:

    (a) $\mathrm{rad}(A) = \bigcap M$ over all maximal ideals $M \subset A$ (a purely algebraic definition);

(b) $x \in \mathrm{rad}(A)$ iff $\rho(x) = 0$ (i.e. the spectrum of x is trivial); or

(c) $x \in \mathrm{rad}(A)$ iff $\|x^n\|^{1/n} \to 0$.

Since $\rho(f) = \sup |\widehat{f}|$, we have:

Theorem. The kernel of the Gelfand representation $A \mapsto \widehat{A}$ is the radical of A.

29. 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,

$$|I^n(f)(x)| \le \|f\| x^n/n! \le \|f\|/n!.$$

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'.

30. *The algebra determines the topology.* A Banach algebra is *semisimple* if its radical is trivial.

Theorem. 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 $\rho(x) \le \|x\|$, we have $|\phi(x)| \le \|x\|$ 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. ∎

31. *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. 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.

32. *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.

33. *Evaluation of integrals.* Theorem. If $f(z) = P(z)/Q(z)$ is a rational function with poles outside of $\sigma(a)$, then $f(a) = P(a)/Q(a)$. That is, the definition of $f(a)$ by algebra agrees with the definition via the Cauchy integral formula.

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 $f(a) = 1$.

34. *Density of rational functions.* Finally we extend the calculus by continuity to holomorphic f. To this end we note:

Theorem. 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. ■

35. *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. ■

36. *Motivation of C^* algebras.* Norm computation in $\mathcal{B}(H)$. How can you compute the operator norm of a matrix $T \in A = M_n(\mathbb{C})$? For example, $T = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$?

One way is to observe that if T is *self-adjoint*, that is $T = T^*$, then there is a basis of *orthogonal* eigenvectors, and therefore $\|T\| = \sup |\lambda_i|$ over the eigenvalues.

For a general operator we have

$$\|T\|^2 = \sup_{|x|=|y|=1} |\langle Tx, Ty \rangle| = \sup |\langle T^*Tx, y\rangle| = \|T^*T\|.$$

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^*).

37. *C^* algebras.* A *C^*-algebra* is a Banach algebra equipped with a map $a \mapsto a^*$ satisfying:

(a) $a^{**} = a$;

(b) $(ab)^* = b^* a^*$;

(c) $(a+b)^* = a^* + b^*$;

(d) $(\lambda a)^* = \overline{\lambda} a^*$; and

(e) $\|aa^*\| = \|a\|^2$.

Prime examples: $A = C(X)$ with complex conjugation as $*$; $A = \mathcal{B}(\mathbb{H})$ with adjoint as $*$ (the last property is verified above).

An *involution* on A is an operator $*$ satisfying all the hypotheses above except the last.

Warning Example: 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.

38. *Commutative C^* algebras.* Theorem. Let A be a commutative C^* algebra. Then Gelfand transform gives an isomorphism

$$A \cong \widehat{A} = C(X)$$

between A and the algebra of *all* continuous functions on a compact Hausdorff space X. The involution $*$ is sent to complex conjugation.

Proof. The trickiest part of the proof is to show $*$ goes to complex conjugation. To do this, it suffices to show that the spectrum of any self-adjoint element is real. (Any $a \in A$ can be written uniquely as $a = x + iy$ where x and y are self-adjoint.)

So suppose $a = a^*$, $\phi \in \operatorname{Spec} A$ and $\phi(a) = z = x + iy \in \mathbb{C}$, say with $y \geq 0$. We will show $y = 0$. To this end, let a real parameter Y tend to $+\infty$ and consider the growth of $\phi(a + iY)$: we have

$$|\phi(a + iY)|^2 = x^2 + (y+Y)^2 = (y+Y)^2 + O(1).$$

On the other hand, we expect iY and a to be 'orthogonal' in the algebra A, and this is borne out by:

$$\begin{aligned}\|a+iY\|^2 &= \|(a+iY)(a+iY)^*\| = \|(a+iY)(a-iY)\| = \|a^2+Y^2\| \\ &\leq Y^2+O(1).\end{aligned}$$

Since $|\phi(a-iY)|^2 \leq \|a-iY\|^2$, we conclude $y=0$ and thus $\phi(a) \in \mathbb{R}$.

For a general element $a = x+iy$ in A we then have

$$\phi(a^*) = \phi(x-iy) = \phi(x) - i\phi(y) = \overline{\phi(a)}$$

since $\phi(x)$ and $\phi(y)$ are real.

To complete the proof, note that for any self-adjoint element x we have $\|x^2\| = \|x\|^2$ and thus

$$\rho(x) = \lim \|x^{2^n}\|^{1/2^n} = \|x\|.$$

For an arbitrary element $a \in A$ we then have

$$\rho(a)^2 = \rho(aa^*) = \|aa^*\| = \|a\|^2.$$

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. The image is a subalgebra closed under conjugation and separating points; thus $\widehat{A} = C(X)$. ∎

39. Theorem. The functor $A \mapsto \operatorname{Spec} A$ 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 $\widehat{f} : \operatorname{Spec} B \to \operatorname{Spec} A$ defined by

$$\widehat{f}(\phi : B \to \mathbb{C}) \quad = \quad (\phi \circ f : A \to \mathbb{C}).$$

40. *Robustness of the C^*-spectrum.* Now consider possibly *non-commutative* C^* algebras again.

Theorem. If $x \in A$ is self-adjoint, then $\sigma(x) \subset \mathbb{R}$.

Proof. This holds in the commutative algebra generated by x, where the spectrum can only be larger. ■

Theorem. If $B \subset A$ is an inclusion of C^*-algbras, 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's self-adjoint, it's invertible in B. That is, $yx^*x = 1$ for some $y \in B$; but then $yx^* = x^{-1}$. ■

41. *Non-commutative spaces.* 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 $\mathcal{B}(H)$.

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 $C(S^1) \ltimes \mathbb{Z}$.

8 Operator algebras and the spectral theorem

42. Let $\mathcal{H}$ be a Hilbert space (usually separable). Then $\mathcal{B}(\mathcal{H})$ is a C^*-algebra, as we saw above.

    We will make an analysis of *self-adjoint operators*, i.e. those $T \in \mathcal{B}(\mathcal{H})$ satisfying $T = T^*$.

43. *Continuous functional calculus.* Let T be self-adjoint. Then $\sigma(T) \subset \mathbb{R}$ and 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.

    Proof. Since T is self-adjoint, A is a commutative C^*-subalgebra of $\mathcal{B}(\mathcal{H})$. Thus $A \cong C(X)$ where $X = \operatorname{Spec} A$.

    We claim $\operatorname{Spec} A = \sigma(T)$. (Recall the spectrum of T is the same in A and in $\mathcal{B}(\mathcal{H})$.) Indeed, for any $\lambda \in \sigma(T)$ we can consider a maximal ideal containing $\lambda - T$, and thus obtain a multiplicative linear functional $\phi : A \to \mathbb{C}$ with $\phi(T) = \lambda$. Conversely, if $\phi(T) = \lambda$, then $\phi(P(T)) = P(\lambda)$ for any polynomial P. Since $\mathbb{C}[P]$ is dense in A, ϕ is *determined* by the condition $\phi(T) = \lambda$. Thus $\operatorname{Spec} A = \sigma(T)$. ■

44. *Positive operators.* If $T = T^*$ and $\langle Tx, x\rangle \geq 0$ for all x we say T is a *positive* operator. A typical example is the operator $T = S^2$ where S is self-adjoint.

    Theorem. For $T \geq 0$, $\langle Tx, x\rangle$ defines an inner-product on $\mathcal{H}/\operatorname{Ker} T$, and hence we have the Cauchy-Schwarz inequality:

    $$|\langle Tx, y\rangle|^2 \leq \langle Tx, x\rangle\langle Ty, y\rangle.$$

45. *The norm of a self-adjoint operator.* Theorem. If $T = T^*$ then

$$\|T\| = \sup_{\|x\|=1} |\langle Tx, x\rangle|.$$

Proof. First assume $T \geq 0$. Clearly $|\langle Tx, x\rangle| \leq \|T\|$. For the reverse inequality, use the fact that $\mathcal{B}(\mathcal{H})$ is a C^*-algebra, and Cauchy-Schwarz:

$$\begin{aligned} \|T\|^2 &= \|T^2\| = \sup_{\|x\|=\|y\|=1} |\langle T^2 x, y\rangle| \\ &= \sup |\langle Tx, Ty\rangle| \leq \sup \langle Tx, x\rangle \langle Ty, y\rangle \leq \|T\| \sup \langle Tx, x\rangle. \end{aligned}$$

We leave the general case to the reader. ■

46. *Finite-dimensional case.* Theorem. Let T be a self-adjoint operator on a finite-dimensional Hilbert space. Then $\mathcal{H}$ admits an *orthonormal* basis such that T is *diagonal* with *real eigenvalues.*

Proof. First, if $\dim \mathcal{H} > 0$, then there is a unit vector x such that $Tx = \lambda x$, some λ. But $\overline{\lambda} = \langle x, Tx\rangle = \langle Tx, x\rangle = \lambda$, so λ is real.

Set $\mathcal{H}' = x^\perp$. If $y \in \mathcal{H}'$ then

$$\langle Ty, x\rangle = \langle y, Tx\rangle = \lambda \langle y, x\rangle = 0,$$

so T preserves $\mathcal{H}'$. The Theorem now follows by induction on $\dim \mathcal{H}$. ■

47. *The compact case.* Recall that T is *compact* if $\overline{T(B)}$ is compact, where $B \subset \mathcal{H}$ is the unit ball.

Theorem. Let T be a compact self-adjoint operator on a separable Hilbert space. Then $\mathcal{H}$ admits an orthonormal basis e_i such that

$$T(e_i) = \lambda_i e_i,$$

with $\lambda_i \in \mathbb{R}$ and $\lambda_i \to 0$. Conversely, such a basis and λ_i define a compact, self-adjoint operator.

Proof. For simplicity assume T is positive. Choose $x_n \in \mathcal{H}$ of norm one with $\langle Tx_n, x_n\rangle \to \|T\|$. Pass to a subsequence so $x_n \to y$ weakly (that is, $\langle x_n, z\rangle \to \langle y, z\rangle$ for all $z \in \mathcal{H}$).

Then $\|y\| \le 1$. By compactness of T, we have $Tx_n \to Ty$ in the norm topology. Therefore $\langle Ty, y\rangle = \|T\|$, because

$$\begin{aligned} |\langle Ty, y\rangle - \langle Tx_n, x_n\rangle| &\le |\langle T(y-x_n), y\rangle| + |\langle Tx_n, y-x_n\rangle| \\ &\le |\langle T(y-x_n), y\rangle| + |\langle x_n, T(y-x_n)\rangle| \\ &\le 2\|T(y-x_n)\| \to 0. \end{aligned}$$

Since $\|y\| = 1$ and $\langle Ty, y\rangle = \|T\|$, we see Ty and y must be parallel and thus $Ty = \lambda y$ for some $\lambda \in \mathbb{R}$.

Now set $e_1 = y$, replace $\mathcal{H}$ with $y^\perp$ and continue. By compactness of T, the process terminates in countably many steps. ■

48. *Proof via functional calculus.* The spectrum X of a compact, self-adjoint operator is a closed, countable subset $\lambda_i \to 0$ in $\mathbb{R}$. For each $\lambda_i \neq 0$, the functional calculus provides a projection $\P_i = f_i(T)$ where $f(\lambda_i) = 1$ and $f(\lambda) = 0$ otherwise. Clearly $T(\lambda) = \lambda = \sum \lambda_i f_i(\lambda)$, and thus
$$T = \sum \lambda_i P_i.$$
In other words, the spectral theorem expresses T as a weighted sum of commuting projections. Since $TP_i(x) = \lambda_i(x)$, P_i is nothing but projection onto the λ_i eigenspace.

49. *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 $\lambda_n = 1/(2\pi n)$.

50. *The spectral theorem.* In the infinite-dimensional case we will achieve a similar theorem, except the space of eigenvectors may be continuous instead of discrete.

Theorem. Let $T \in \mathcal{B}(\mathcal{H})$ be a self-adjoint operator. 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 $C(\sigma(T)) \to \mathcal{B}(H)$ extends to the bounded Borel functions $L^\infty_b(\sigma(T))$ 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 collection of 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.

51. *Multiplicity.* As an example, let $T \in M_n(\mathbb{C})$ be self-adjoint. If T has distinct eigenvalues λ_i, we can take μ to be a sum of δ-functions concentrated at the λ_i. However, if T has multiple eigenvalues, then we need to take more than one measure to realize T as in (C) above.

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 a basis for $\mathcal{H} = \mathbb{C}^n$. Thus the map from $\mathcal{H}$ to $\oplus L^2(\sigma(T), \mu_i)$ is not unique.

52. *Spectral measures.* The crucial and beautiful idea behind the proof is that each unit vector $x \in \mathcal{H}$ defines a probability measure on $\sigma(T)$ by

$$\int_{\sigma(T)} f(\lambda) \, d\mu_x(\lambda) = \langle f(T)x, x \rangle.$$

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.

53. *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 $|\phi(x)|^2\,dx$.

54. *Proof of (C) of the spectral theorem.* Recalling the isomorphism $C(\sigma(T)) \cong A$, we now consider the subspace $Ax \subset \mathcal{H}$ and define a map

$$\pi : A \cdot x \to L^2(\sigma(T), \mu_x)$$

by $\pi(f(T)x) = f(T)$. 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 $\mathcal{H}_x = \overline{Ax}$ 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 preseved 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. ■

55. *The Borel calculus.* (B) 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 $T_n f \to Tf$ for each $f \in \mathcal{H}$.

Theorem. 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 $L^\infty_b(\sigma(T))$. ∎

56. *Integration with respect to a projection-valued measure.* (A) 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.

57. *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. 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 $T(f(\lambda)) = \lambda f(\lambda)$.

58. *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 $A\mathcal{H}_0^\perp = \mathcal{H}_0^\perp$.

59. *Measure classes.* Suppose T is a normal operator with a 'cyclic vector' v, i.e. one such that $\overline{A \cdot v} = \mathcal{H}$. Then we have seen there is a measure μ and an isomorphism

$$\mathcal{H} \cong L^2(\sigma(T), \mu).$$

To what extent are μ and this isomorphism determined by T?

We say Borel measures μ and ν are in the same *measure class*, $\mu \asymp \nu$, if they have the same sets of measure zero. Then there is a natural isometry between $L^2(\mu)$ and $L^2(\nu)$ given by $f \mapsto f(d\mu/d\nu)^{1/2}$, since

$$\|f\|_\mu^2 = \int f^2\, d\mu = \int f^2 \frac{d\nu}{d\mu}\, d\nu = \int \left(f \sqrt{\frac{d\nu}{d\mu}} \right)^2 d\nu.$$

One would like to attach all of these isomorphic Hilbert spaces to a single *measure class* $[\mu]$. This can done by consider *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 f such that $[|f|^2] \le [\mu]$ (that is, $|f|^2$ is absolutely continuous with respect to μ), with $\langle f, g\rangle = \int fg$.

Then *canonically* associated to the operator T is the *measure class* $[\mu] = [\sum \mu_i]$, where μ_i are the spectral measures for any vectors x_i spanning a dense subspace of $\mathcal{H}$.

60. *Multiplicity.* A normal operator $T \in \mathcal{B}(\mathcal{H})$ also canonically determines a μ-measurable *multiplicity* function

$$m : \sigma(T) \to \{1, 2, 3, \ldots, \infty\}.$$

(Here m is defined μ-a.e.) In terms of the spectral decomposition $\mathcal{H} \cong \oplus L^2(\sigma(T), \mu_i)$ we can express the multiplicity as $m = \sum \chi_{E_i}$, where $\mu_i = f_i \mu$ and $E_i = \{\lambda \,:\, f_i(\lambda) > 0\}$.

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 $f(\lambda)$ with values in $H(\lambda)$, and inner product

$$\langle f, g\rangle = \int_{\sigma(T)} \langle f(\lambda), g(\lambda)\rangle_{H(\lambda)}.$$

Theorem. The data $(\sigma(T), \mu, m)$ determines T up to an isometry of $\mathcal{H}$; that is, there is an isometery of $\mathcal{H}$ to $L^2([\mu], m)$ sending T to multiplication by λ.

61. *Examples.* Let $\mathcal{H} = L^2[0,1]$, let $h \in L^\infty[0,1]$ be a continuous map sending $[0,1]$ to $[a,b]$, and let $T(g) = hg$.

Then g determines a measure μ_g by

$$\int_a^b f(\lambda)\, d\mu_g(\lambda) = \langle f(T)g, g\rangle = \int f(h(x))|g(x)|^2\, dx.$$

In other words,

$$\mu_g = h_*(|g(x)|^2\, dx).$$

All of these measures are absolutely continuous with respect to $\mu = h_*(dx)$, and so $[\mu]$ is *the* spectral measure associated to h.

Now if h is a homeomorphism, then it provides an *isomorphism* between $L^2[0,1]$ and $L^2([a,b],[\mu])$, and so the multiplicity function is equal to one.

If h is a *smooth function* with *isolated critical points*, then $m(\lambda) = \#h^{-1}(\lambda)$.

If h is the Cantor function, then $[\mu]$ consists of δ-masses at the dyadic rationals in $[0,1]$, each with $m(\lambda) = \infty$.

62. *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.

63. *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

$$\mathcal{H} = \int_{\widehat{G}} \mathbb{C}^{m(\lambda)}\, d\mu(\lambda).$$

9 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], [Man].

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 Botzmann 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* [Man, 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 $\widehat{G} = \mathrm{Hom}(G, S^1)$ 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 Bernouilli 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. ∎

10 Summary

1. Distributions: consider $(C^\infty(\Omega))^*$.

2. Fourier transforms, quantum mechanics: diagonalize translation and multiplication. $[d/dx, x] \neq 0$.

3. PDE: linearize differentiation.

4. Prime number theory: the maximal ideals in $A = (L^1(\mathbb{R}), *)$ correspond to points in $\widehat{\mathbb{R}}$, so one is lead to consider $\widehat{f}$ and thus $\zeta(s)$.

5. Banach algebras: to analyze T it is useful to analyze the full algebra A generated by T.

6. C^*-algebras: when commutative, $A \cong C(X)$.

7. Ergodic theory: to study T, study the unitary operator U.

References

[Be] S. K. Berberian. *Lectures on Functional Analysis and Operator Theory.* Springer-Verlag, 1974.

[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.

[Har] G. H. Hardy. *Divergent Series*. Chelsea, 1991.

[Me] R. Mañé. *Ergodic Theory and Differentiable Dynamics*. Springer-Verlag, 1987.

[RS] M. Reed and B. Simon. *Functional Analysis*. Academic Press, 1980.

[RN] F. Riesz and B. Sz.-Nagy. *Functional Analysis*. Frederick Ungar Publishing Co., 1955.

[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.