Real Analysis

Course Notes

Contents

1 Measure, integration and differentiation on $\mathbb{R}$

Motivation. Suppose $f : [0, \pi] \to \mathbb{R}$ is a reasonable function. We define the Fourier coefficients of f by

$$a_n = \frac{2}{\pi} \int_0^\pi f(x) \sin(nx)\, dx.$$

Here the factor of $2/\pi$ is chosen so that

$$\frac{2}{\pi} \int_0^\pi \sin(nx) \sin(mx)\, dx = \delta_{nm}.$$

We observe that if

$$f(x) = \sum_1^\infty b_n \sin(nx),$$

then at least formally $a_n = b_n$ (this is true, for example, for a finite sum).

This representation of $f(x)$ as a superposition of sines is very useful for applications. For example, $f(x)$ can be thought of as a sound wave, where a_n measures the strength of the frequency n.

Now what coefficients a_n can occur? The orthogonality relation implies that
$$\frac{2}{\pi}\int_0^\pi |f(x)|^2\,dx = \sum_{-\infty}^{\infty} |a_n|^2.$$
This makes it natural to ask if, conversely, for any a_n such that $\sum |a_n|^2 < \infty$, there exists a function f with these Fourier coefficients. The natural function to try is $f(x) = \sum a_n \sin(nx)$.

But why should this sum even exist? The functions $\sin(nx)$ are only bounded by one, and $\sum |a_n|^2 < \infty$ is much weaker than $\sum |a_n| < \infty$.

One of the original motivations for the theory of Lebesgue measure and integration was to refine the notion of function so that this sum really does exist. The resulting function $f(x)$ however need to be Riemann integrable! To get a reasonable theory that includes such Fourier series, Cantor, Dedekind, Fourier, Lebesgue, etc. were led inexorably to a re-examination of the foundations of real analysis and of mathematics itself. The theory that emerged will be the subject of this course.

Here are a few additional points about this example.

First, we could try to define the required space of functions — called $L^2[0,\pi]$ — to simply be the metric completion of, say $C[0,\pi]$ with respect to $d(f,g) = \int |f-g|^2$. The reals are defined from the rationals in a similar fashion. But the question would still remain, can the limiting objects be thought of as functions?

Second, the set of point $E \subset \mathbb{R}$ where $\sum a_n \sin(nx)$ actually converges is liable to be a very complicated set — not closed or open, or even a countable union or intersection of sets of this form. Thus to even begin, we must have a good understanding of subsets of $\mathbb{R}$.

Finally, even if the limiting function $f(x)$ exists, it will generally not be Riemann integrable. Thus we must broaden our theory of integration to deal with such functions. It turns out this is related to the second point — we must find a good notion for the length or *measure* $m(E)$ of a fairly general subset $E \subset \mathbb{R}$, since $m(E) = \int \chi_E$.

1.1 Real numbers, topology, logic

The real numbers. Conway: Construction of the real numbers [Con, p.25].

Dedekind: just as a prime is characterized by the ideal of things it divides, so a number is characterized by the things less than it.

Brouwer and Euclid: the continuum is not a union of points!

Hilbert: the axiomatic approach to the reals. Formalism versus intuitionism.

Other completions of $\mathbb{Q}$: e.g. $\mathbb{Q}_2, \mathbb{Q}_{10}$. (In the latter case the completion is a ring but not a field! If 5^n accumulates on x and 2^n accumulates on y, then $|x|_{10} = |y|_{10} = 1$ but $xy = 0$. One can make the solution canonical by asking that $x = (0,1)$ and $y = (1,0)$ in $\mathbb{Z}_{10} \cong \mathbb{Z}_2 \times \mathbb{Z}_5$; then $y = x + 1 = \ldots 4106619977392256259918212890625$.)

Basic topological property of the reals (not shared by the other completions: connectedness).

The irrationals in $[0,1]$ are isomorphic to $\mathbb{N}^{\mathbb{N}}$ by

$$(a_1, a_2, \ldots) \mapsto 1/(a_1 + 1/(a_2 + \cdots)).$$

(Here $\mathbb{N} = \{1, 2, 3, \ldots\}$.)

Proof. draw the Farey tree.

Cardinality. Russell's paradox: if $E = \{X \ : \ X \notin X\}$, then is $E \in E$? Make this into a proof that $|\mathcal{P}(X)| > |X|$. Corollary: $\mathbb{R}$ is uncountable, since $2^{\mathbb{N}}$ is isomorphic to the Cantor set.

Helpful tool: Schröder-Bernstein.

Question: How many rational numbers are there? How many algebraic numbers? Are most numbers transcendental?

Answer: in terms of counting, yes.

Answer: in terms of measure: the probability that $x \in [0,1]$ is equal to a given algebraic number $a \in A$ is zero. Thus the probability that $x \in A$ is zero.

Paradox: why doesn't the same argument show $x \neq y$ for every $y \in [0,1]$? Equivalent, what is wrong with the following equation:

$$1 = m([0,1]) = m\left(\bigcup_x \{x\}\right) = \sum m(\{x\}) = 0?$$

If uncountable additivity is suspect, what about countable additivity?

The Borel hierarchy. Induction, over the natural numbers and over an ordinal. Example: any $\mathcal{C} \subset \mathcal{P}(X)$ generates a unique smallest algebra. (Use induction over $\mathbb{N}$). Similar, generates a unique σ-algebra. (Use induction over Ω, whose existence comes from well-ordering of the reals.)

Ex: $\langle f_n \rangle$ continuous $\implies$ the set of points where $f_n(x)$ is bounded is an F_σ. E.g.

$$f_n(x) = \sum_{k=1}^{n} |\sin(\pi k! x)|^{1/n}$$

is bounded iff $x \in \mathbb{Q}$.

A *condensation point* of $E \subset \mathbb{R}$ is a point $x \in \mathbb{R}$ such that every neighborhood of x meets E in an uncountable set. In other words, its the set of points where E is 'locally uncountable'.

Theorem 1.1 *Any uncountable set contains an uncountable collection of condensation points.*

The same holds true in any complete, separable metric space. Thus only countably many Y's can be embedded disjointly in $\mathbb{R}^2$, and only countably many Möbius bands in $\mathbb{R}^3$.

Any closed uncountable set F has the order of the continuum. In fact it contains a copy of the Cantor set. (Proof: pick two condensation points, and then two disjoint closed intervals around them. Within each interval, pick two disjoint subintervals containing condensation points, and continue. By insuring that the lengths of the intervals tend to zero we get a Cantor set.)

How many open sets? Theorem. The set of all open subsets of $\mathbb{R}$ is of the same cardinality as $\mathbb{R}$ itself. Indeed, the same is true of the set of all Borel sets.

1.2 Lebesgue measurable sets and functions

On $\mathbb{R}$ we will construct a σ-algebra $\mathcal{M}$ containing the Borel sets, and a **measure** $m : \mathcal{M} \to [0, \infty]$, such that $m(a, b) = b - a$, m is translation-invariant, and m is countably additive.

Definition: the *outer measure* $m^*(E)$ is the infimum of $\sum \ell(I_i)$ over all coverings $E \subset \bigcup I_i$ by countable unions of intervals.

Basic fact: subadditivity. For any collection of sets A_i, $m^*(\bigcup A_i) \leq \sum m^*(A_i)$.

Basic fact: $m^*[a, b] = b - a$.

Proof. Clearly the outer measure is at most $b - a$. But if $[a, b]$ is covered by $\bigcup I_k$, by compactness we can assume the union is finite, and then

$$b - a = \int \chi_[a, b] \leq \int \sum \chi_{I_k} = \sum |I_k|.$$

Definition: $E \subset \mathbb{R}$ is **measurable** if

$$m^*(E \cap A) + m^*(\widetilde{E} \cap A) = m^*(A)$$

for all sets $A \subset \mathbb{R}$. Because of subadditivity, only one direction needs to be checked.

For example, if $E \subset [0,1]$ then $m^*(E) + m^*([0,1] - E) = 1$.

Theorem 1.2 *$E = [a, \infty)$ is measurable.*

Proof. From a good cover $\bigcup I_i$ for A we must construct good covers for $E \cap A$ and $\widetilde{E} \cap A$. This is easy because E cuts each interval I_i into two subintervals whose lengths add to that of I_i. ■

Theorem 1.3 *The measurable sets form an algebra.*

Proof. Closure under complements is by definition. Now suppose E and F are measurable, and we want to show $E \cap F$ is. By the definition of measurability, E cuts A into two sets whose measures add up. Now F cuts $E \cap A$ into two sets whose measures add up, and similarly for the complements. Thus E and F cut A into 4 sets whose measures add up to the outer measure of A. Assembling 3 of these to form $A \cap (E \cup F)$ and the remaining one to form $A \cap \widetilde{E \cup F}$, we see $E \cup F$ is measurable.

Theorem 1.4 *If E_i are disjoint and measurable, $i = 1, 2, \ldots, N$, then $\sum m^*(E_i \cap A) = m^*(A \cap \bigcup E_i)$.*

Proof. By induction, the case $N = 1$ being the definition of measurability. ■

Theorem 1.5 *The measurable sets form a σ-algebra.*

Proof. Suppose E_i is a sequence of measurable sets; we want to show $\bigcup E_i$ is measurable. Since we already have an algebra, we can assume the E_i are disjoint. By the preceding lemma, we have for any finite N,

$$\sum_1^N m^*(E_i \cap A) + m^*(A \cap \bigcap_1^N \widetilde{E_i}) = m^*(A).$$

The second term is only smaller for an infinite intersection, so letting $N \to \infty$ we get

$$\sum_1^\infty m^*(E_i \cap A) + m^*(A \cap \bigcap_1^\infty \widetilde{E_i}) \leq m^*(A).$$

Now the first term dominates $m^*(A \cap \bigcup E_i)$ so we are done.

Corollary 1.6 *All Borel sets are measurable.*

Definition. $m(E) = m^*(E)$ if E is measurable.

Theorem 1.7 *If E_i are disjoint measurable sets, then $m(\bigcup E_i) = \sum m(E_i)$. Proof: follows from the Theorem above, setting $A = \mathbb{R}$ and passing to the limit.*

Continuity of measure. If $m(E_1)$ is finite and $E_1 \supset E_2 \supset E_3 \ldots$, then $m(\bigcap E_i) = \lim m(E_i)$. **Proof.** let $F = \bigcap E_i$ and write $E_1 = F \cup (E_1 - E_2) \cup (E_2 - E_3) \cup \ldots$.

Littlewood's first principle. Let $E \subset \mathbb{R}$ be measurable. Then E is approximated from the outside (inside) by open (closed) sets, and to within a set of measure zero by a $\mathcal{G}_\delta$ ($\mathcal{F}_\sigma$).

If the measure of E is finite, then there is a finite union of intervals J such that $m(E \triangle J) < \epsilon$.

Proof. First suppose $m(E)$ is finite. Then there are open intervals such that $E \subset \bigcup I_i$ and $m(\bigcup I_i - E) \geq (\sum m(I_i)) - m(E) < \epsilon$. This shows E is approximate from the outside by an open set, and from the inside by a closed set. Also if we take J to be the union of a large finite subset of $\{I_i\}$, then what's left over has small measure, so we get $m(E \triangle J)$ small. (The difference is covered by $(\bigcup I_i - J) \cup (\bigcup I_i - E)$.)

Why not use this Theorem to define the σ-algebra of measurable sets?

Corollary: Every Borel set B can be expressed as $B = G - N = F \cup N'$, where G is a $\mathcal{G}_\delta$, F is an $\mathcal{F}_\sigma$, and N, N' are sets of measure zero. Thus measure theory 'short circuits' the Borel hierarchy. Note that the notion of zero measure is elementary (compared to the notion of measurable).

Corollary: If E has positive measure then there exists an interval I such that $m(E \cap I)/m(I) \approx 1$. **Proof.** Take a very efficient cover of E by finitely many intervals J. Then the ratio $m(J)/m(E) \approx 1$, and on the other hand $m(J)/m(E)$ is bounded below by the sup of the density of E in each subinterval.

Nonmeasurable sets. Let $G = \mathbb{R}/\mathbb{Z}$ and let $H = \mathbb{Q}/\mathbb{Z} \subset G$ be a normal subgroup. Then there exists a set of coset representatives $P \subset G$ for G/H. Since $m(G) = 1$ and $G = \bigcup_H h + P$, the measure of P cannot be defined. Thus H is nonmeasurable.

Let B be a basis for $\mathbb{R}$ over $\mathbb{Q}$. Is B measurable?

Assume the Continuum Hypothesis. Then we can well-order $[0, 1]$ such that each initial segment is countable. Set $R = \{(x, y) : x < y\}$ in this ordering. Then horizontal slices (fixing y) have measure zero, while all vertical slices (fixing x have measure one).

Measurable functions. A function $f : \mathbb{R} \to \mathbb{R}$ is *measurable* if $f^{-1}(U)$ is measurable whenever U is an open set.

First examples: continuous functions, step functions ($\sum_1^N a_i \chi_{I_i}$, I_i disjoint intervals) and simple functions ($\sum_1^N a_i \chi_{E_i}$, E_i disjoint measurable sets) are all measurable. Note that simple functions are exactly the measurable functions taking only finitely many values.

In general, if $f : A \to B$ is any map, the map $f^{-1} : \mathcal{P}(B) \to \mathcal{P}(A)$ is a σ-algebra homomorphism; indeed it preserves unions over **any** index set. Thus f is measurable is the same as: (a) $f^{-1}(x, \infty)$ is measurable for all $x \in \mathbb{R}$; or (b) $f^{-1}(B)$ is measurable for any Borel set B.

Warning. It is *not* true that $f^{-1}(M)$ is measurable whenever M is measurable! Thus measurable functions are not closed under composition.

More generally, for a topological space X we say $f : \mathbb{R} \to X$ is measurable if the preimages of open sets are measurable. Example: if f, g are measurable functions, then $h = (f, g) : \mathbb{R} \to \mathbb{R}^2$ is measurable. Indeed, for any open set $U \times V \subset \mathbb{R}^2$, the preimage $h^{-1}(U \times V) = f^{-1}(U) \cap g^{-1}(V)$ is measurable. Since every open set in $\mathbb{R}^2$ is a union of a countable number of open rectangles, h is measurable.

Similarly, if $h : \mathbb{R}^2 \to \mathbb{R}$ is continuous, then $h(f, g)$ is measurable whenever f and g are. This shows the measurable functions form an **algebra:** fg and $f + g$ are measurable if f and g are.

Moreover, the measurable functions are closed under limits. Indeed, if $f = \lim f_n$ then

$$\begin{aligned} f^{-1}(a, \infty) &= \{x \,:\, \exists k\, \exists N\, \forall n \geq N\ f_n(x) > a + 1/k\} \\ &= \bigcup_k \bigcup_N \bigcap_{n \geq N} f_n^{-1}(a + 1/k, \infty). \end{aligned}$$

Similarly for $\limsup, \liminf$ etc.

If $f = g$ a.e. and f is measurable then so is g.

Theorem 1.8 (Littlewood's second principle) *If f is measurable on $[a, b]$ then f is the limit in measure of continuous functions: there exists continuous f_n such that for all $\epsilon > 0$, $m\{|f - f_n| > \epsilon\} \to 0$.*

Proof. Let $E_M = \{|f| > M\}$; then $\bigcap E_M = \emptyset$, so after truncating f on a set of small measure we obtain f_1 bounded by M. Cutting $[-M, M]$ into finitely many disjoint intervals of length ϵ, and collecting together the values, we see f_1 is a uniform limit of simple functions. Any simple function is built from indicator functions χ_E of measurable sets. By Littlewood's first principle,

χ_E is approximated in measure by χ_J, where J is a finite union of intervals. Finally χ_J is a limit in measure of continuous functions. ∎

Theorem 1.9 (Lusin's Theorem; Littlewood's 2nd principle) *Given a measurable function f on $[0,1]$, one can find a continuous function $g : [0,1] \to \mathbb{R}$ such that $g = f$ outside a set of small measure.*

Theorem 1.10 (Egoroff; Littlewood's 3rd principle) *Let $f(x) = \lim f_n(x)$ for each $x \in [0,1]$, where f_n, f are measurable. Then $f_n \to f$ uniformly outside a set of small measure.*

Example: Recall the 'tent functions' f_n supported on $[0, 1/n]$ with a triangular graph of height n. We have $f_n \to 0$ but $\int f_n = 0$; these f_n do not converge uniformly everywhere.

Proof of the Theorem. For any $k > 0$, consider the sets

$$E_N = \{x \; : \; |f_n(x) - f(x)| > 1/k \text{ for some } n > N\}.$$

Since $f_n \to f$, we have $\bigcap E_N = \emptyset$. Since $E_N \subset [0,1]$, we have $m(E_N) \to 0$. Thus there is an $N(k)$ such that $m(E_{N(k)})$ is as small as we like, say less than $2^{-k}\epsilon$. Let $A = \bigcup_k E_{N(k)}$. Then for x outside A, we have $\sup |f_n(x) - f(x)| \le 1/k$ for all $n > N(k)$, and therefore $\sup |f_n(x) - f(x)| \to 0$. In other words, $f_n \to f$ uniformly outside the set A; and $m(A) \le \epsilon$. ∎

Finitely-additive measures on $\mathbb{N}$. The natural numbers admit a finitely-additive measure defined on *all* subsets, and vanishing on finite sets. (Such a measure is cannot be countably additive.) This construction gives a 'positive' use of the Axiom of Choice, to construct a measure rather than to construct a non-measurable set.

Filter: $\mathcal{F} \subset \mathcal{P}(X)$ such that sets in $\mathcal{F}$ are 'big':

(1) $\emptyset \notin \mathcal{F}$,
(2) $A \in \mathcal{F}, B \subset A \implies B \in \mathcal{F}$; and
(3) $A, B \in \mathcal{F} \implies A \cap B \in \mathcal{F}$.

Example: the cofinite filter (if X is infinite).

Example: the 'principal' ultrafilter $\mathcal{F}_x$ of all sets with $x \in F$. This is an *ultrafilter*: if $X = A \sqcup B$ then A or B is in $\mathcal{F}$.

Theorem 1.11 *Any filter is contained in an ultrafilter.*

Proof. Using Zorn's lemma, take a maximal filter $\mathcal{F}$ containing the given one. Suppose neither A nor $X-A$ is in $\mathcal{F}$. Adjoining to $\mathcal{F}$ all sets of the form $F \cap A$, we obtain a larger filter $\mathcal{F}'$, a contradiction. (To check $\emptyset \notin \mathcal{F}'$: if $A \cap F = \emptyset$ then $X-A$ is a superset of F, so $X-A$ *was* in $\mathcal{F}$.)

Ideals and filters. In the ring $R = (\mathbb{Z}/2)^X$, ideals $I \neq R$ and filters are in bijection: $I = \{A : \widetilde{A} \in \mathcal{F}\}$. The ideal consists of 'small' sets, those whose complements are big.

(By (2), $A \in I \implies AB \in I$. By (3), $A, B \in I \implies A \cup B \in I \implies (A \cup B)(A \triangle B) = A + B \in I$.)

Lemma: if $\mathcal{F}$ is an ultrafilter and $A \cup B = F \in \mathcal{F}$ then A or B is in $\mathcal{F}$.

Proof. We prove the contrapositive. If neither A nor B is in $\mathcal{F}$, then their complements satisfy $\widetilde{A}, \widetilde{B} \in \mathcal{F}$. Since $\mathcal{F}$ is a filter,

$$\widetilde{A} \cap \widetilde{B} = \widetilde{A \cup B} \in \mathcal{F}$$

and thus $A \cup B \notin \mathcal{F}$.

Corollary: Ultrafilters correspond to prime ideals.

By Zorn's Lemma, every ideal is contained in a maximal ideal; this gives another construction of ultrafilters.

Measures. Let $\mathcal{F}$ be an ultrafilter. Then we get a finitely-additive measure on **all** subsets of X by setting $m(F) = 1$ or 0 according to $F \in \mathcal{F}$ or not. Conversely, any 0/1-valued finitely additive measure on $\mathcal{P}(X)$ determines a filter.

Measures supported at infinity. The most interesting case is to take the cofinite filter, and extend it in some way to an ultrafilter. Then we obtain a finitely-additive measure on $\mathcal{P}(X)$ such that points have zero measure but $m(X) = 1$. When $X = \mathbb{N}$ such a measure cannot be countably additive.

1.3 Integration

We will write $f = \lim f_n$ if $f(x) = \lim f_n(x)$ for all $x \in \mathbb{R}$. Because $\lim f_n$ is measurable whenever each f_n is, the measurable functions turn out to work well to represent points in the metric spaces obtained by completing the integrable, compactly supported continuous or smooth functions (those satisfying $\int |f| < \infty$, $\int |f|^p < \infty$, etc.

Our next goal is to extend the theory of integration to measurable functions.

A *simple function* ϕ is a measurable function taking only finite many values. It has a canonical representation as $\phi = \sum_1^N a_i \chi_{E_i}$ where the a_i

enumerate the nonzero values of ϕ and $E_i = \{\phi = a_i\}$ are disjoint sets. The simple functions form a vector space.

Simple integration. For a simple function supported on a set of finite measure, we define

$$\int \phi = \int \sum a_i \chi_{E_i} = \sum a_i m(E_i).$$

We also define $\int_E \phi = \int \phi \chi_E$.

Example: $\int \chi_{\mathbb{Q}} = 0$.

Theorem 1.12 *Integration is linear on the vector space of simple functions.*

Proof. Clearly $\int a\phi = a \int \phi$. We must prove $\int \phi + \psi = \int \phi + \int \psi$.

First note that for any representation of ϕ as $\sum b_i \chi_{F_i}$ with the sets F_i disjoint, we have $\int \phi = \sum b_i m(F_i)$. Indeed,

$$\int \sum b_i \chi_{F_i} = \int \sum a_j \chi_{\bigcup_{b_i = a_j} F_i} = \sum a_j \sum_{b_i = a_j} m(F_i) = \sum b_i m(F_i).$$

Now take the finite collection of sets F_i on which ϕ and ψ are *both* constant, and write $\phi = \sum a_i \chi_{F_i}$ and $\psi = \sum b_i \chi_{F_i}$. Then

$$\int \phi + \psi = \sum (a_i + b_i) m(F_i) = \int \phi + \int \psi.$$

■

The Lebesgue integral. Now let E be a set of finite measure, let $f : E \to \mathbb{R}$ be a function and assume $|f| \le M$. We define the *Lebesgue integral* by

$$\int_E f = \inf_{\psi \ge f} \int_E \psi = \sup_{f \ge \phi} \int_E \phi,$$

assuming sup and inf agree. (Here ϕ and ψ are required to be simple functions.)

Theorem 1.13 *The two definitions of the integral of f above agree iff f is a measurable function.*

Proof. Suppose f is measurable. Since $\int \psi \geq \int \phi$, we just need to show the simple functions ϕ and ψ can be chosen such that their integrals are arbitrarily close. To this end, cut the interval $[-M, M]$ into N pieces $[a_i, a_{i+1})$ of length less than ϵ. Let E_i be the set on which $f(x)$ lies in $[a_i, a_{i+1})$. Then $\phi = \sum a_i \chi_{E_i}$ and $\psi = \sum a_{i+1} \chi_{E_i}$ satisfying $\phi \leq f \leq \psi$ and $\int (\psi - \phi) \leq \epsilon m(E)$, so we are done.

Conversely, if the sup and inf agree, then we can choose simple functions $\phi_n \leq f \leq \psi_n$ such that $\int (\psi_n - \phi_n) \to 0$. Let $\phi = \sup \phi_n$ and $\psi = \inf \psi_n$. Then ϕ and ψ are measurable, and $\phi \leq f \leq \psi$.

We claim $\phi = \psi$ a.e. (and thus f is measurable). Otherwise, there is a set of positive measure A and an $\epsilon > 0$ such that $\psi - \psi > \epsilon$ on A. But then $\epsilon \chi_A \leq \psi_n - \phi_n$ for all n, and thus $\int \psi_n - \phi_n \geq \epsilon m(A) > 0$. ■

Theorem 1.14 *Let f be a bounded function on an interval $[a, b]$, and suppose f is Riemann integrable. Then f is also Lebesgue integrable, and the two integrals agree.*

Proof. If f is Riemann integrable then there are **step functions** $\phi_n \leq f \leq \psi_n$ with $\int (\psi_n - \phi_n) \to 0$. Since step functions are special cases of simple functions, we see f is Lebesgue integrable.

It is now easy to check that the integral of bounded functions over sets of finite measure satisfies expected properties:

The integral is linear.

If $f \leq g$ then $\int f \leq \int g$.

In particular $|\int f| \leq \int |f|$, and if $A \leq f \leq B$ then $Am(E) \leq \int_E f \leq Bm(E)$.

For disjoint sets, $\int_{A \cup B} f = \int_A f + \int_B f$. The most interesting assertion is $\int (f + g) = \int f + \int g$. If $\psi_1 \geq f$ and $\psi_2 \geq g$ then $\psi_1 + \psi_2 \geq f + g$, so by the infimum definition of the integral we get $\int (f + g) \leq \int f + \int g$. To get the reverse inequality, use the supremum definition.

Theorem 1.15 (Bounded convergence) *Let $f_n \to f$ (pointwise) Theorem (Bounded convergence) Let $f_n \to f$ (pointwise) on a set of finite measure E, where $|f_n|, |f| \leq M$. Then $\int_E f_n \to \int_E f$.*

Proof. We will use Littlewood's 3rd Principle. Ignoring a set A of small measure, the convergence is uniform. Then

$$\left| \int_{E-A} f_n - f \right| \leq \int_{E-A} |f_n - f| \leq m(E - A) \sup_{E-A} |f_n - f| \to 0.$$

On the other hand, $|\int_A f_n|$ and $|\int_A f|$ are both less than $Mm(A)$, so ignoring A makes only a small change in the integrals and therefore we have convergence.

Banach limits. If we mimic the definition of the Lebesgue integral using a finitely-additive, non-atomic measure on $\mathbb{N}$ (i.e. a non-principal ultrafilter), then we obtain a linear map

$$L : \ell^\infty(\mathbb{N}) \to \mathbb{R}$$

with $L(a_n) \geq 0$ if $(a_n) \geq 0$, and with $L(a_n) = \lim a_n$ if the limit exists.

The general Lebesgue integral. For $f \geq 0$ we define $\int f = \sup_{0 \leq g \leq f} \int g$, where g ranges over bounded functions supported on sets of finite measure. Clearly this is the same as saying $\int f = \lim = \int f_M$, where $f_M = \min(f, M)|[-M, M]$.

For general f, we require that $\int |f| < \infty$ before $\int f$ is defined. Then writing $f = f_+ - f_-$, we define $\int f = \int f_+ - \int f_-$.

Linearity. Let us check that $\int (f+g) = \int f + \int g$. First suppose $f, g \geq 0$. Then from $0 \leq f_1 \leq f$ and $0 \leq g_1 \leq g$ we get $0 \leq f_1 + g_1 \leq f + g$, so $\int (f+g) \geq \int f + \int g$. On the other hand, given $0 \leq h \leq f + g$ we can write $h = f_1 + g_1$ with $f_1 = \min(f, h)$; then $f_1 \leq f$ and $g_1 \leq g$ so $\int f + g \leq \int f + \int g$. This completes the proof for positive functions.

For the general case, note that if $f = g - h$ with $g, h \geq 0$ integrable functions, then $\int f = \int g - \int h$. Indeed, we have $g \geq f_+$ and $h \geq f_-$, so their differences are positive, and indeed $(g - f_+) = (h - f_-)$. Thus by linearity for positive functions, we get

$$\begin{aligned} \int g - h &= \int (f_+ + g - f_+) - \int (f_- + h - f_-) \\ &= \left(\int f_+ - \int f_- \right) + \left(\int (g - f_+) - \int (h - f_-) \right) = \int f. \end{aligned}$$

Now to prove linearity, just note that if $f = g + h$, then $f = (g_+ + h_+) - (g_- + h_-)$ expresses f as a sum of two positive integrable functions. Integrating each one and using linearity for positive functions we get $\int f = \int g + \int h$.

Integrals and limits. In general from $f_n \to f$ we can deduce no relationship between $\int f$ and $\int f_n$. The basic example of the tent functions can be made positive or negative; we can even get $\int f_n$ to oscillate in an arbitrary way, while $f_n \to 0$ a.e. (This 'a.e.' often signals 'pointwise convergence'.)

Positive functions. The situation is better if $f, f_n \geq 0$, and $f_n \to f$. There are two main results:

Fatou's Lemma: $\int f \leq \liminf \int f_n$.
Monotone Convergence: if $f_1 \leq f_2 \leq \ldots$, then $\int f = \lim \int f_n$.

Proofs: For Fatou's lemma, let g be a bounded function with bounded support such that $g \leq f$ and $(\int f) - \epsilon \leq \int g$. Then $g_n = \min(g, f_n) \to g$ and $g_n \leq f_n$, so

$$\left(\int f\right) - \epsilon \leq \int g = \lim \int g_n \leq \liminf \int f_n.$$

Here we have used the Bounded Convergence Theorem to interchange integrals and limits.

Letting $\epsilon \to 0$ gives the result.

For monotone convergence: Since $f \geq f_n$ for all n, we have $\int f \geq \limsup \int f_n$, while $\int f \leq \liminf \int f_n$ by Fatou's Lemma. ∎

Theorem 1.16 (Modulus of integrability) *Let $f \geq 0$ be integrable. Then for any $\epsilon > 0$ there is a $\delta > 0$ such that $m(E) < \delta \implies \int_E f < \epsilon$.*

Corollary 1.17 *The function $F(t) = \int_{-\infty}^t f(x)\,dx$ is uniformly continuous on $\mathbb{R}$.*

Proof of the Theorem. Let $f_M = \min(M, f)$. Then $f_M \to f$ monotonely as $M \to \infty$, and thus $\int (f - f_M) \to 0$. Choose M large enough that $\int (f - f_M) < \epsilon/2$. Then for $m(E) < \delta = \epsilon/(2M)$, we have $\int_E f \leq \int_E (f - f_M) + M m(E) \leq \epsilon$. ∎

Dominated convergence. Let $f_n \to f$, with $|f_n|, |f| \leq g$ and $\int g < \infty$. Then $\int f_n \to \int f$.

Proof. Given $\epsilon > 0$ there is a $\delta > 0$ such that $\int_A g < \epsilon$ whenever $m(A) < \delta$. We can also choose M such that $\int_E g < \epsilon$ outside $[-M, M]$. Then by Littlewood's 3rd principle, there is a set $A \subset [-M, M]$ with $m(A) < \delta$ outside of which $f_n \to f$ uniformly. Thus

$$\limsup \left| \int f_n - f \right| \leq 2 \left(\int_{\mathbb{R} - [-M,M]} g + \int_A g \right) \leq 4\epsilon.$$

Since ϵ was arbitrary, $\int f_n \to \int f$.

Derivatives. Even if $f'(x)$ exists everywhere, the behavior of $f'(x)$ can be very wild – e.g. not integrable. For example, if $f(x)$ is any function smooth away from $x = 0$, and $|f(x)| \le |x|^2$, then f is differentiable at 0; but we can make $f'(x)$ wild, e.g. look at $f(x) = x^2 \sin(e^{1/x^2} x)$. In particular, $f'(x)$ need not be integrable.

Here is an easy theorem illustrating the preceding results.

Theorem 1.18 *Suppose $f(x)$ is differentiable on $\mathbb{R}$, vanishes outside $[0, 1]$ and $|f'(x)| \le M$. Then $\int_0^t f'(x)\,dx = f(t)$.*

Proof. Since f is differentiable it is continuous, and $f_n(x) = n(f(x+1/n) - f(x)) \to f'(x)$ pointwise. By the mean-value theorem, $|f_n(x)| = |f'(y)| \le M$ for some $y \in [x, x + 1/n]$. Thus $\int f_n \to \int f'$. But

$$\int_0^t f_n(x)\,dx = n \int_t^{t+1/n} f(t)\,dt \to f(t)$$

by continuity of f. ∎

Convergence in measure. All the theorems about pointwise convergence also hold for convergence in measure. This can be proved using the following useful fact.

Theorem 1.19 *If $f_n \to f$ in measure, then there is a subsequence such that $f_n \to f$ pointwise a.e.*

As a warm-up to this fact, we prove the easy part of the Borel-Cantelli lemma.

Lemma 1.20 *If $\sum m(E_n) < \infty$, then $\limsup E_n$, the set of points x that belong to E_n for infinitely many n has measure zero.*

Remark: $\chi_{\limsup E_n} = \limsup \chi_{E_n}$.

Proof. For any $N > 0$, we have

$$m(\limsup E_n) \le m(\bigcup_N^\infty E_n) \le \sum_N^\infty m(E_n) \to 0$$

as $N \to \infty$. ∎

Proof of the Theorem. For each $k > 0$ there is an $n(k)$ such that $E_k = \{|f - f_{n(k)}| > 2^{-k}\}$ satisfies $m(E_k) < 2^{-k}$. We claim $f_{n(k)}(x) \to f(x)$ a.e. as $k \to \infty$. Indeed, $\sum m(E_k) < \infty$, so almost every x belongs to but finitely many E_k. And fixing x, for all k large enough that $x \notin E_k$ we $|f(x) - f_{n(k)}| \le 2^{-k} \to 0$.

Sample application. Suppose g is integrable, $|f_n| \le g$ for all n and $f_n \to f$ in measure. Then $\int f_n \to \int f$.

Proof. Let A be any limit point of $\int f_n$, possibly $\pm\infty$. We will show $A = \int f$.

Pass to a subsequence such that $\int f_n \to A$. Pass to a further subsequence so $f_n \to f$ pointwise. By the pointwise theorem we get $\int f_n \to \int f$, so $A = \int f$. Since A was any limit point of the original sequence $\int f_n$, we have $\int f_n \to \int f$.

2 Differentiation and Integration

Functions that are differentiable everywhere. Even if $f'(x)$ exists everywhere, it does not have to be continuous. For example, if $|f(x)| \le x^2$, then no matter how badly $f'(x)$ oscillates near $x = 0$, we have $f'(0) = 0$.

As an application of interchange of integrals, we can ask: if $f'(x)$ exists everywhere, then can we assert

$$\int_a^b f'(x) = f(b) - f(a)?$$

The answer is *no* in general, since $\int_a^b |f'(x)|$ might be infinite. However, we can approach the problem by defining $g_n(x) = n(f(x + 1/n) - f(x))$. Then clearly $g_n(x) \to f'(x)$ pointwise, and by continuity of f it is easy to see

$$\int_a^b g_n(x) \to f(b) - f(a).$$

So it is simply a question of interchanging integration and limits. For example, if f is *Lipschitz*, then g_n is *bounded*, so by the bounded convergence theorem, f is the integral of f'. More generally, the same conclusion holds if we can find a locally integrable function h such that

$$|f(x + t) - f(x)| \le th(x).$$

for $|t| \le 1$.

A nowhere differentiable function.

Let $f(x) = \sum_1^\infty a_n \sin(b_n x)$, where $\sum a_n$ converges quickly but $b_n a_n \to \infty$ rapidly. For concreteness, we take $a_n = 10^{-n}$, $b_n = 10^{6n}$.

Then for any n, we can choose $t \approx 1/b_n$ such that $\Delta a_n \sin(b_n x) \asymp a_n$. For $k < n$, we have

$$\sum \Delta a_k \sin(b_k x) \leq \sum a_k b_k / b_n \asymp a_{n-1} b_{n-1}/b_n \ll a_n,$$

and for $k > n$ we have

$$\Delta a_k \sin(b_k x) \leq a_k \ll a_n$$

Thus $\Delta f / \Delta x \asymp a_n / t \asymp a_n b_n \to \infty$, so $f'(x)$ does not exists.

Riemann's 'example'. Riemann thought that the function

$$f(x) = \sum \exp(2\pi i n^2 x)/n^2$$

was nowhere differentiable. This is almost true, however it turns out that $f'(x)$ actually does exists at certain rational points.

Monotone functions. We say $f : [a,b] \to \mathbb{R}$ is *increasing* if $x \leq y \implies f(x) \leq f(y)$. If f or $-f$ is increasing then f is *monotone*.

Example: write $\mathbb{Q} = \{q_1, q_2, \ldots\}$ and set $f(x) = \sum_{q_i < x} 2^{-i}$. Then $f : \mathbb{R} \to \mathbb{R}$ is monotone increasing, and f has a dense set of points of discontinuity.

Theorem 2.1 *A monotone function $f : [a,b] \to \mathbb{R}$ is differentiable almost everywhere.*

Thus the oscillations of the preceding example are *necessary* to produce nowhere differentiability.

Gleason has remarked that this property of monotone functions helped lead him to his proof of Hilbert's 5th problem (which topological groups are Lie groups?).

The proof of the Theorem will use the Vitali covering lemma.

Vitali coverings. Here is use an important covering argument based on the 'greedy algorithm'.

Let K be a compact subset of a metric space (X,d). A collection of balls $\mathcal{B}$ forms a *Vitali covering* of K if for every $x \in K$ and $r > 0$ there is a $B \in \mathcal{B}$ with $x \in B \subset B(x,r)$.

We can be rather loose about the boundary of B: it is only necessary that $B(y,s) \subset B \subset \overline{B(y,s)}$ for some open ball $B(y,s)$. In the case of the real numbers, this means B can be any interval except a degenerate one $[a,a]$.

Theorem 2.2 *For any Vitali covering $\mathcal{B}$ of K, there is a sequence of* disjoint *balls $\langle B(y_i, r_i)\rangle$ in $\mathcal{B}$ such that $K \subset \bigcup B(y_i, 3r_i)$. In fact for any $N > 0$ we have*

$$K \subset \bigcup_1^N \overline{B(y_i, r_i)} \cup \bigcup_{N+1}^{\infty} B(y_i, 3r_i).$$

Proof. Since K is compact, we can assume $\mathcal{B}$ is a countable set of balls whose diameters tend to zero. (For each n, extract from K a finite subcover $\mathcal{B}_n$ by balls of diameter $< 1/n$, and replace $\mathcal{B}$ with $\bigcup_1^\infty \mathcal{B}_n$ — it is still a cover in the sense of Vitali.)

To construct the disjoint balls, we use the greedy algorithm. Let $B(y_1, r_1)$ be the largest ball in $\mathcal{B}$, and define $B(y_{n+1}, r_{n+1})$ inductively as one of the largest balls among those in $\mathcal{B}$ disjoint from the ones already chosen, $B(y_1, r_1), \ldots, B(y_n, r_n)$.

We claim $K \subset \bigcup B(y_i, 3r_i)$. Indeed, if $x \in K$ then x belongs to some ball $B(y, r) \in \mathcal{B}$. If $B(y, r)$ belongs to the sequence of chosen balls $B(y_i, r_i)$, then we are done — x is covered.

Otherwise, consider the first i for which $r_i < r$. Since $B(y, r)$ was not chosen at the ith stage in the inductive definition, it must meet one of the earlier balls — say $B(y_j, r_j)$, with $j < i$. But then we have $r_j \geq r$, and since they meet, $B(y_j, 3r_j)$ contains $B(y, r)$. In particular, it contains x.

Now suppose we have $N > 0$ and $x \in K - \bigcup_1^N B(y_i, r_i)$. Then since the union of the first N balls is closed, there is a ball $B(y, r) \in \mathcal{B}$ *disjoint* from the first N balls and containing x. Once again, by the nature of the greedy algorithm $B(y, r)$ must meet $B(y_i, r_i)$ for some i with $r_i \geq r$; but this time by our choice of $B(y, r)$ we can insure that $i > N$. Since $r_i \geq r$ we have $x \in \bigcup_{N+1}^\infty B(y_i, 3r_i)$. ∎

Theorem 2.3 (Vitali covering lemma) *For any Vitali covering $\mathcal{B}$ of a set $E \subset \mathbb{R}$ of finite measure, and $\epsilon > 0$, there is a finite collection of disjoint balls $B_1, \ldots, B_n$ in $\mathcal{B}$ with $m(E \triangle \bigcup_1^n B_i) < \epsilon$.*

Proof. Since $m(E)$ is finite, we can find a compact K and an open U such that $K \subset E \subset U$ and $m(K)$, $m(E)$ and $m(U)$ all differ by at most ϵ. Remove from $\mathcal{B}$ any balls that are not contained in U; then $\mathcal{B}$ is still a Vitali covering of E, and hence of K.

Now extract a sequence of disjoint balls $\langle B_i = B(y_i, r_i)\rangle$ from $\mathcal{B}$ by the greedy algorithm. Then by Vitali's Lemma, we have $m(\bigcup B_i) = \sum 2r_i \leq$

$m(U) < \infty$, so we can choose $n > 0$ such that $\sum_{n+1}^{\infty} r_i < \epsilon$. Then $K - \bigcup_1^n B_i$ is covered by $\bigcup_{n+1}^{\infty} B(y_i, 3r_i)$, and therefore

$$m(K - \bigcup_1^n B_i) \le \sum_{n+1}^{\infty} 6r_i < 6\epsilon.$$

We also have $m(E-K) < \epsilon$, so $m(E - \bigcup_1^n B_i) < 7\epsilon$. Finally $m(\bigcup_1^n B_i - E) \le m(U - E) < \epsilon$, so we conclude $m(E \triangle \bigcup_1^n B_i) < 8\epsilon$.

Proof of monotone differentiability. We will assume $f : [a,b] \to \mathbb{R}$ is monotone increasing. For any set $A \subset \mathbb{R}$ let $[A]$ be the smallest interval containing it, so $m[A] = \sup A - \inf A$.

Fix rational numbers $0 \le u < v$ and consider the set $E \subset [a,b]$ of those x for which there are arbitrarily small t such that $m[f(x+t), f(x)] \le ut$ and also arbitrarily small t such that $m[f(x-t), f(x)] \ge vt$. This means the derivative of f measured from above x wants to lie below u, but from the right is wants to lie above v, and we have $u < v$ so f is not differentiable at x.

The set of all points where f is not differentiable is a countable union of sets of the same basic form as E, so we will be content to show E has measure zero. Also the points of discontinuity of f are countable, so we can assume f is continuous on E.

The idea of the proof is to show that $m(f(E)) = um(E) = vm(E)$ and thus $m(E) = 0$.

More precisely, there is a Vitali covering $\mathcal{B}$ of E by intervals of the form $B = [x, x+t]$ with $m[f(B)]/m(B) < u$. From these extract a finite disjoint cover $B_1, \ldots B_n$ with $m(E \triangle \bigcup_1^n B_i) < \epsilon$. Then we have

$$\sum m[f(B_i)] \le u \sum m(B_i) \le u(m(E) + \epsilon).$$

Now let $A = E \cap \bigcup_1^n \operatorname{int}(B_i)$. There is a Vitali covering of A by intervals $C = [x-t, x]$ expanded under f by a factor of at least v, and with $C \subset B_i$ for some i. We can extract a finite union of disjoint balls $C_1, \ldots C_m$ such that $\sum m(C_i) > m(A) - \epsilon > m(E) - 2\epsilon$. Then we find

$$\sum m[f(C_i)] \ge v \sum m(C_i) \ge v(m(E) - 2\epsilon).$$

But each C_i is a subset of some B_j, so we have

$$\begin{aligned} v(m(E) - 2\epsilon) &\le \sum_1^m m[f(C_i)] \\ &\le \sum_1^n m[f(B_j)] \le u(m(E) + \epsilon). \end{aligned}$$

Letting $\epsilon \to 0$ we find $vm(E) \le um(E)$ and thus $m(E) = 0$.

Theorem 2.4 (Integral of the derivative) *If $f : [a,b] \to \mathbb{R}$ is monotone, then $\int_a^b f'(x)\,dx \le f(b) - f(a)$.*

Proof. Define $f_n(x) = n(f(x + 1/n) - f(x)) \ge 0$. Then $f_n(x) \to f'(x)$, so by Fatou's lemma we have $\int f' \le \liminf \inf f_n$. But $\int f_n$ is, for n large, the difference between the averages of f over two disjoint intervals, so it is less than or equal to the maximum variation $f(b) - f(a)$.

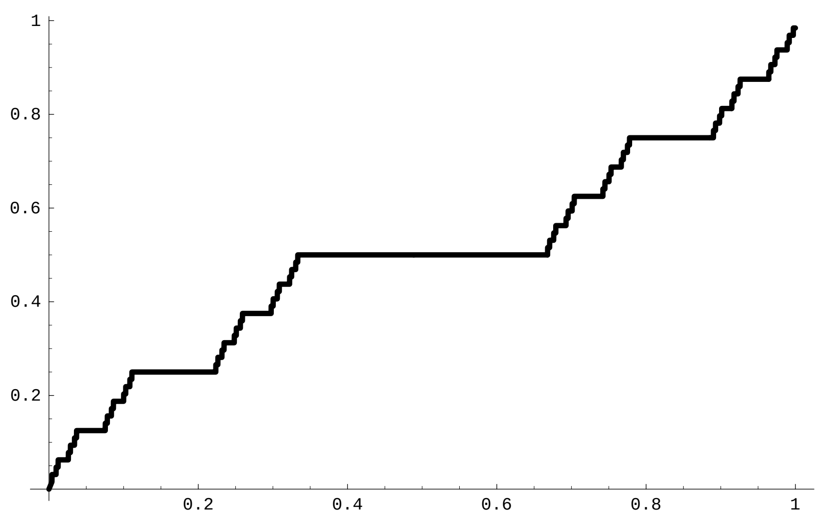

Figure 1. Cantor's function: the devil's staircase.

Singular functions. A monotone function is **singular** if $f'(x) = 0$ a.e. An example is the Cantor function or 'devil's staircase',

$$f(0.a_1a_2a_3\ldots) = \sum\{2^{-i} \;:\; a_i \le 1 \,\text{and}\, a_j \ne 1, 1 \le j < i.\}$$

where $x = 0.a_1a_2a_3\ldots$ in base 3.

This monotone function has the amazing property that it is continuous, and it climbs from 0 to 1, but $f'(x) = 0$ a.e. On the other hand, $f'(x)$ does not exist (or equals infinity) for x in the Cantor set (in fact f stretches intervals of length 3^{-n} to length 2^{-n}, and so even for a monotone function $f'(x)$ can fail to exist on an uncountable set (necessarily of measure zero).

There is a more sophisticated example, due to Whitney, of a function $f(x, y)$ on the plane whose derivatives exist everywhere, but which is not constant on its critical set. This function describes the topography of a hill with a (fractal) road running from top to bottom passing only along the level or flat parts of the hillside.

Bounded variation. We note that if $f = g - h$ where g and h are both monotone, then $f'(x)$ also exists a.e. So it is desirable to characterize the full vector space of functions spanned by the monotone functions.

A function $f : [a, b] \to \mathbb{R}$ has **bounded variation** if

$$\sup \sum_1^n |f(a_i) - f(a_{i-1})| = \|f\|_{BV} < \infty.$$

Here the sup is over all finite dissections of $[a, b]$ into subintervals, $a = a_0 < a_1 < \ldots a_n = b$. This supremum is called the *total variation* of f over $[a, b]$.

Theorem 2.5 *A function f is of bounded variation iff $f(x) = g(x) - h(x)$ where g and h are monotone increasing.*

Proof. Clearly $\|f\|_{BV} = f(b) - f(a)$ if f is monotone increasing, and thus f has bounded variation if it is a difference of monotone functions.

For the converse, define

$$f_+(x) = \sup \sum_1^n \max(0, f(a_i) - f(a_{i-1})),$$

over all partitions $a = a_0 < \ldots < a_n = x$, and similarly

$$f_-(x) = \sup \sum_1^n \max(0, -f(a_i) + f(a_{i-1})).$$

Clearly f_+ and f_- are monotone increasing, and they are bounded since the total variation of f is bounded.

We claim $f(x) = f(a) + f_+(x) - f_+(x)$. To see this, note that if we refine our dissection of $[a, b]$, then both f_+ and f_- increase. Thus for any $\epsilon > 0$, we can find a dissection for which *both* sums are within ϵ of their supremums.

But for a common partition, it is clear that

$$\begin{aligned}
&\sup \sum_1^n \max(0, f(a_i) - f(a_{i-1})) - \max(0, -f(a_i) + f(a_{i-1})) \\
&= \sup \sum_1^n f(a_i) - f(a_{i-1}) = f(x) - f(a).
\end{aligned}$$

Thus $f(x) = f(a) + f_+(x) - f_-(x)$. ∎

Corollary 2.6 *Any function of bounded variation is differentiable a.e.*

Theorem 2.7 (Lebesgue Density) *Let $E \subset \mathbb{R}$ be a measurable set. Then for almost every $x \in \mathbb{R}$,*

$$\lim_{r\to 0} \frac{m(E \cap B(x,r))}{m(B(x,r))} = \begin{cases} 1 & \text{if } x \in E, \\ 0 & \text{otherwise.} \end{cases}$$

Derivative of the integral. This theorem says that if we let $F(x) = \int_{-\infty}^{x} f(t)\,dt$, where $f(t) = \chi_E(t)$, then $F'(x)$ exists a.e. and $F'(x) = f(x)$ a.e. We will eventually prove that this is the case for **all** integrable f.

Proof of Lebesgue density. We consider the case $E \subset [0,1]$. Let E_n be the set of $x \in E$ such that $\liminf m(E \cap B(x,r))/2r < 1 - 1/n$. Using the Vitali lemma, we can find a finite set of disjoint intervals $I_1, \ldots I_N$ such that $m(E_n \triangle \bigcup I_i) < \epsilon$ and the density of E in I_i is less than $1 - 1/n$. Then

$$\begin{aligned}
m(E_n) - \epsilon &\le m(E \cap \bigcup I_i) = \sum \frac{m(E \cap I_i)}{m(I_i)} m(I_i) \\
&\le (1 - 1/n) \sum m(I_i) \le (1 - 1/n)(m(E_n) + \epsilon).
\end{aligned}$$

Since ϵ was arbitrary, we get $m(E_n) \le (1 - 1/n) m(E_n)$ and therefore $m(E_n)$=0.

Thus the limit in the Theorem exists and equals 1 for all $x \in E - \bigcup E_n$, so we have demonstrated half of the Theorem. For the other half, replace E with $[0,1] - E$. ∎

The Lebesgue density theorem has many basic applications in ergodic theory. Here is an example.

Theorem 2.8 *Let $\theta \in \mathbb{R} - \mathbb{Q}$ be an irrational number, and define $f : S^1 \to S^1$ by $f(x) = x + \theta \bmod 1$. Then f is ergodic: if $E \subset S^1$ has positive measure, and $f(E) \subset E$, then $m(E) = 1$.*

Proof. Let $\delta_r(x) = m(E \cap B(x,r))/m(B(x,r))$. Since E has positive measure, there is an $x_0 \in E$ such that for any $\epsilon > 0$ there is an $r > 0$ with $\delta_r(x_0) > 1 - \epsilon$. By invariance of E, δ_r is constant along the orbits of f. But each orbit is dense, so we have $\delta_r(x) > 1 - \epsilon$ along the dense set $\langle f^n(x) \rangle$. Since $\delta_r(x)$ is continuous, it is bounded below by $1 - \epsilon$ everywhere. Thus $\lim_{r \to 0} \delta_r(x) = 1 = \chi_E(x)$ a.e., and thus $m(E) = 1$.

Corollary. Any measurable function $h : S^1 \to \mathbb{R}$, invariant under the irrational rotational f, is constant a.e.

Proof. For any partition of $\mathbb{R}$ into disjoint intervals I_i of length ϵ, we have $m(f^{-1}(I_i)) = 1$ for exactly one i. As $\epsilon \to 0$, this distinguished interval shrinks down to the constant value assumed by f.

Absolute continuity. A function $F : [a,b] \to \mathbb{R}$ is **absolutely continuous** if for any $\epsilon > 0$, there is a $\delta > 0$ such that for any finite set of non-overlapping intervals (a_i, b_i), if $\sum_1^n |a_i - b_i| < \epsilon$ then $\sum_1^n |f(a_i) - f(b_i)| < \delta$.

Theorem 2.9 *An absolutely continuous function is continuous and of bounded variation.*

Proof. Continuity is clear. As for bounded variation, choose ϵ and δ as above; then over any interval of length δ, the total variation of f is at most ϵ, so over $[a,b]$ we have variation about $\epsilon(b-a)/\delta$. ∎

Theorem 2.10 *Let $F : [a,b] \to [c,d]$ be an absolutely continuous homeomorphism. Then $m(A) = 0$ implies $m(F(A)) = 0$.*

Proof. Given $\epsilon > 0$, choose δ as guaranteed by absolute continuity, and cover A by disjoint open intervals $I_i = (a_i, b_i)$ with $\sum_1^\infty |b_i - a_i| < \delta$. Since f is a homeomorphism, we have

$$m(f(A)) \leq m\left(\bigcup f(I_i)\right) \leq \sum m(f(I_i)) = \sum |f(b_i) - f(a_i)| < \epsilon.$$

Thus $m(f(A)) = 0$. ∎

Theorem 2.11 *The derivative $D(F) = f(x) = F'(x)$ establishes a bijective correspondence:*

$$\begin{aligned} D \ : \ & \{\textit{absolutely continuous } F : [a,b] \to \mathbb{R},\ F(a) = 0\} \leftrightarrow \\ & \{\textit{integrable } f : [a,b] \to \mathbb{R}\}. \end{aligned}$$

The inverse is given by $I(f) = F(x) = \int_a^x f(t)\, dt$.

Lemma 2.12 *If f is integrable then $F(x) = \int_a^x f$ is absolutely continuous.*

Proof. This follows from the fact that for any $\epsilon > 0$ there is a $\delta > 0$ such that $\int_A |f| < \epsilon$ whenever $m(A) < \delta$.

Lemma. If f is absolutely continuous, then it is of bounded variation, so $f'(x)$ is integrable.

Proof. The bounded variation is clear; then $f = f_+ - f_-$, both monotone increasing, and we have

$$\int |f'| \leq \int f'_+ + f'_- \leq f_+(b) - f_+(a) + f_-(b) - f_-(a) = \|f\|_{BV} < \infty.$$

Now we show the derivative of an integral gives the expected result. We have already proved this for the indicator function of a measurable set; the following argument gives a different proof.

Lemma. If $f : [a,b] \to \mathbb{R}$ is integrable, and $F(x) = \int_a^x f(t)\, dt = 0$ for all x, then $f = 0$. (This shows injectivity of the map I.)

Proof. Consider the collection of all sets over which the integral of f is zero. By assumption this contains all intervals in $[a,b]$, and it is closed under countable unions and complements. Thus it contains all closed sets in $[a,b]$. But if $f \neq 0$, then either $\{f > 0\}$ or $\{f < 0\}$ contains a closed set F of positive measure. Then $\int_F f \neq 0$, contradiction. Thus $f = 0$.

Theorem 2.13 *If f is bounded, then $F = I(f)$ is Lipschitz and satisfies $F'(x) = f(x)$ a.e.*

Proof. Suppose $|f| \leq M$; then clearly $|F(x+t) - F(x)| \leq Mt$. We will show $\int_a^c F'(x) - f(x)\, dx = 0$ for all c. To this end, just note that $F'(x) = \lim F_n(x) = n(F(x+1/n) - F(x))$ satisfies $|F_n| \leq M$, so it is a pointwise limit of bounded functions. Thus

$$\begin{aligned} \int_a^c F'(x)\, dx &= \lim \int_a^c F_n(x)\, dx = \lim n \left(\int_c^{b+1/n} F - \int_a^{a+1/n} F \right) \\ &= F(c) - F(a) = \int_a^c f(x)\, dx, \end{aligned}$$

by continuity of F. ■

Theorem 2.14 *Even if f is unbounded, but integrable, we have $D(I(f)) = f$.*

Proof. By linearity, it is enough to prove this assertion for positive f. Let $f_n = \min(n, f) \to f$, an increasing sequence, and let $F_n = I(f_n)$. Then $F = I(f - f_n) + F_n$ and since $I(f - f_n)$ is a monotone increasing function, we have

$$F'(x) \geq F_n'(x) = f_n(x)$$

a.e. (using the result $D(I(f_n)) = f_n$ for bounded f_n). Therefore

$$\int_a^c F'(x)\,dx \geq \int_a^c f_n(x)\,dx = F_n(c) \to F(c).$$

On the other hand,

$$\int_a^c F'(x)\,dx \leq F(c) - F(a) = F(c)$$

by our general results on integration of increasing functions. Thus equality holds, and we have shown

$$\int_a^c F'(x) - f(x)\,dx = 0$$

for all c. Thus $F'(x) = f(x)$ a.e.

Now we turn to the converse inequality: to show $I(D(F)) = F$ for all absolutely continuous F. This direction is a little easier.

Lemma. If F is absolutely continuous and $F'(x) = 0$ a.e. then F is constant.

Proof. (This verifies injectivity of D.) Pick any $c \in [a, b]$. Using the Vitali lemma, cover $[a, c]$ with a finite number of intervals $I_1, \ldots, I_n$ such that $|\Delta F/\Delta t| < \epsilon$ over these intervals, and what's left over has total measure at most ϵ. Then by absolutely continuity, the total variation of F over the complementary intervals is at most δ. Thus

$$|F(c) - F(a)| \leq \delta + \epsilon \sum m(I_i) \leq \delta + \epsilon(b - a),$$

and this can be made arbitrarily small so $F(c) = F(a)$. ■

Theorem 2.15 $I(D(F)) = F$.

Proof. Let $f = F'(x)$, and let $G = I(f)$; then we've seen that G is absolutely continuous, and $G'(x) = f(x) = F'(x)$ a.e., so $(G - F)'(x) = 0$ a.e. implies $G - F$ is a constant. By our normalization, $G(a) = F(a) = 0$, so $F = G = I(D(F))$. ■

Summary: letting M=monotone functions, we have

$$BV = M - M \supset AC \iff f = \int f'.$$

We will eventually see the differentiation form of this setup:

$$\{\text{signed } \mu\} = \{\mu - \nu \,:\, \mu, \nu \geq 0\} \supset \{\mu = f(x)\,dx\} \iff L^1(\mathbb{R}).$$

Convexity. A function $f : \mathbb{R} \to \mathbb{R}$ is *convex* if for all $x, y \in \mathbb{R}$ and $t \geq 0$ we have

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y).$$

In other words, the graph of the function f lies below every one of its chords.

Theorem 2.16 *The right and left derivatives of a convex function exist for all x, and agree outside a countable set.*

Proof. The secant lines move monotonely. ■

We have yet to use the Monotone Convergence Theorem. When can we assert that the approximations to the derivative, $f_t(x) = (f(x+t) - f(x))/t$, converge to $f'(x)$ *monotonely* as t decreases to zero?

Answer: when f is a convex function!

Theorem 2.17 *A function f is convex iff f is absolutely continuous and $f'(x)$ is increasing.*

Proof. Suppose f is convex. Then the slope of the secant line $f_t(x) = (f(x+t) - f(x))/t$ is an increasing function of t and of x. It follows that $f_t(x)$ is uniformly bounded on any compact interval $[c, d]$ in the domain of f. Thus f is Lipschitz, which implies it is absolutely continuously. Finally $f'(x)$ is increasing since it is a limit of increasing functions.

For the converse, just note that when f is absolutely continuous, the secant slope

$$f_t(x) = \frac{1}{t}\int_x^{x+t} f'(s)\,dx$$

is just the average of f'. But the average of an increasing function is itself increasing, so the secants of the graph of f have increasing slope, which implies f is convex. ∎

Corollary 2.18 *If $f(x)$ is convex, then $f''(x)$ exists a.e. and $f''(x) \geq 0$.*

Theorem 2.19 (Jensen's inequality) *If $f : \mathbb{R} \to \mathbb{R}$ is a convex function, and $X : [0,1] \to \mathbb{R}$ is integrable, then*

$$f\left(\int X\right) \leq \int f(X) = \int_0^1 f(X(t))\,dt.$$

Proof. First note that equality holds if f is a linear function. Also, both sides of the equation are linear functions of f (under pointwise addition). So it is enough to prove the result after modifying f by a linear function. To this end, let $m = \int X$ be the mean of X, take a linear supporting function $g(x) = ax + b$ with $g(m) = f(m)$ and $g(x) \leq f(x)$ otherwise; and replace f with $f - g$. Then $f(\int X) = 0$ but $\int f(X) \geq 0$. ∎

Probabilistic interpretation. If f is convex, then $E(f(X)) \geq f(E(X))$ for any random variable X. Jensen's theorem is this statement where the distribution of the random variable is dictated by the function $X : [0,1] \to \mathbb{R}$. **It includes** δ-masses as a special case, since these are obtained when X is a simple function.

The definition of convexity says the result holds for random variables assuming just two values x or y, with probabilities t and $(1-t)$ respectively.

A bettor's dilemma. You are about to gamble with \$100 at a fair game. A generous patron has offered to square your holdings. Do you ask for this boost before you start playing, to increase your stakes, or after you have gambled, to increase your payoff?

Answer: let X denote your payoff. Fairness means $E(X) = 100$. Since x^2 is convex, $E(X^2) \geq (E(X))^2$, so squaring your payoff is better on average.

Arithmetic/Geometric Mean. As is well-known, for $a, b > 0$, we have $\sqrt{ab} \leq (a+b)/2$, because:

$$0 \leq (\sqrt{a} - \sqrt{b})^2/2 = (a+b)/2 - \sqrt{ab}.$$

More generally, considering a random variable that assumes values $a_1, \dots a_n$ with equal likelihood, the concavity of the logarithm implies

$$\log\left(\frac{1}{n}\sum a_i\right) \geq \frac{1}{n}\sum \log a_i$$

and thus

$$\left(\prod a_i\right)^{1/n} \leq \frac{1}{n}\sum a_i.$$

Mnemonic: To remember the direction of this inequality, note that if $a_i \geq 0$ but $a_1 = 0$, then the geometric mean is zero but the arithmetic mean is not.

Appendix: Measure on $[a, b]$. There is a correspondence between monotone functions $f : [a, b] \to \mathbb{R}$ and positive, finite measures μ on $[a, b]$, namely:

$$f(x) = \mu[a, x).$$

(This function is always continuous on one side: we have $f(x_n) \to f(x)$ if $x_n \uparrow x$.)

Now we will later see that $\mu = \mu^a + \mu^s$, where μ^a and μ^s are absolutely continuous and singular with respect to Lebesgue measure. (That is, μ^s vanishes outside a set of Lebesgue measure zero, while $\mu^a(E) = \int_E g$ for some positive integrable g.) Also $[a, b]$ may contain countably many 'atoms' for μ, i.e. points with $\mu(p) > 0$ (delta masses).

Then we have the following dictionary:

Functions	**Measures**
f is monotone increasing	μ is a positive measure
f' exists a.e.	$f' = g = d\mu^a/dx$ a.e.
f is singular	$\mu = \mu^s$
f is absolutely continuous	$\mu = \mu^a$
f is discontinuous at p	p is an atom for μ
countably many discontinuities	countably many atoms

Finally one can also consider *signed measures*; these correspond to functions of bounded variation, and the canonical representation of f as a difference of monotone functions corresponds to the Hahn decomposition, $\mu = \mu_+ - \mu_-$, μ_+ and μ_- mutually singular positive measures.

3 The Classical Banach Spaces

A *normed linear space* is a vector space V over $\mathbb{R}$ or $\mathbb{C}$, equipped with a norm $\|v\| \geq 0$ defined for every vector, such that:

$\|v\| = 0 \implies v = 0$;

$\|\alpha v\| = |\alpha| \cdot \|v\|$; and

$\|v + w\| \leq \|v\| + \|w\|$. A norm is the marriage of metric and linear structures. It determines a distance by $d(v, w) = \|v - w\|$.

A *Banach space* is a complete normed linear space.

The unit ball. It is frequently useful to think of a norm in terms of its *closed unit ball*, $B = \{v \,:\, \|v\| \leq 1\}$. Then we can recover the norm by $\|v\| = \inf\{\alpha > 0 \,:\, v \in \alpha B\}$. The conditions above insure:

B contains no line through the origin;

B is symmetric; and

B is convex. Conversely, when checking the sub-additivity of a norm, it suffices to show B is convex.

Theorem 3.1 (Verifying completeness) *A normed linear space is complete iff $\sum \|a_i\| < \infty \implies$ there is an $a \in V$ such that $\sum_1^N a_i \to a$.*

Proof. If a_i is a Cauchy sequence in V we can pass to a subsequence such that $d(a_i, a_{i+1}) < 2^i$. Then $a_1 + (a_2 - a_1) + (a_3 - a_2) + \ldots$ is absolutely summable, so it sums to some $s \in V$, and $a_i \to s$. The converse is obvious. ∎

Example: $C(X)$. Let X be any compact Hausdorff space, and let $C(X)$ be the vector space of continuous functions $f : X \to \mathbb{R}$. Define $\|f\| = \sup_X |f|$. Then $\sum \|f_i\| < \infty$ implies the sum converges uniformly, and therefore $\sum f_i(x) = f(x)$ exists and is continuous; thus $C(X)$ is a Banach space.

Example: $\ell^p(I)$. For any index set I, and $1 \leq p \leq \infty$, we let $\ell^p(I)$ denote the space of sequences $a = \langle a_i \,:\, i \in I \rangle$ with

$$\begin{aligned} \|a\|_p &= \left(\sum_I |a_i|^p \right)^{1/p}, \\ \|a\|_\infty &= \sup_I |a_i|. \end{aligned}$$

The outer exponent is put in to give homogeneity of degree one.

Thus $\ell^p(2)$ gives a norm on $\mathbb{R}^2$, with the unit ball defined by $x^p + y^p \le 1$. Note that as p increases from 1 to ∞, these balls are all convex, and they move steadily from a diamond through a circle to a square. In $\mathbb{R}^3$ they move from an octahedron through a sphere to a cube.

The 'usual' ℓ^p spaces refer to $\ell^p(\mathbb{N})$. We also have $c = c(\mathbb{N}) = C(\mathbb{N} \cup \infty)$, the space of convergent sequences with the sup norm, and $c_0 \subset c$, the space of sequences converging to zero.

The L^p spaces. For any measurable subset $E \subset \mathbb{R}$, and $1 \le p \le \infty$, we define $L^p(E)$ as the set of measurable functions $f : E \to \mathbb{R}$ such that $\int_E |f|^p < \infty$; and set

$$\begin{aligned} \|f\|_p &= \left(\int_E |f|^p \right)^{1/p}, \\ \|f\|_\infty &= \inf\{M \ge 0 \,:\, |f| \le M \text{ a.e}\}. \end{aligned}$$

Actually for the norm of f to vanish, it is only necessary for f to vanish a.e., so the elements of L^p are technically equivalence classes of functions defined up to agreement a.e.

Indicator functions. If $m(A)$ is finite, then $\|\chi_A\|_p = (m(A))^{1/p} \to 1$ as $p \to \infty$.

The scale of spaces. If $m(E) < \infty$ then $L^\infty(E) \subset L^p(E) \subset L^1(E)$, i.e. the L^p spaces shrink as p rises.

If $m(E) = \infty$ then there is no comparison.

Theorem 3.2 *For $1 \le p \le \infty$, the space $L^p(E)$ with the norm above is a Banach space.*

Theorem 3.3 (Minkowski's inequality) $\|f + g\|_p \le \|f\|_p + \|g\|_p$. *For $1 < p < \infty$, equality holds iff f and g lie on a line in L^p.*

Proof. As mentioned above, it suffices to verify convexity of the unit ball; that is, assuming $\|f\| = \|g\| = 1$, we need only verify

$$\|tf + (1-t)g\| \le 1$$

for $0 < t < 1$. In fact by convexity of the function x^p, $p > 1$, we have

$$\int |tf + (1-t)g|^p \le \int t|f|^p + (1-t)|g|^p \le t + (1-t) = 1.$$

This proves B is convex. For $1 < p$, the strict convexity of x^p gives strict convexity of B, furnishing strict inequality unless f and g lie on a line. ∎

Completeness. It remains to show L^p is complete. Suppose $\sum \|f_i\|_p < \infty$. Let $F(x) = \sum |f_i(x)|$. Then by monotone convergence, $\int F^p \leq (\sum_1^\infty \|f_i\|_p)^p$, so $F(x)$ is finite a.e. and it lies in L^p. Therefore the same is true for $f(x) = \sum f_i(x)$, since $|f(x)| \leq F(x)$; and we have $\|f\|_p \leq \sum \|f_i\|_p$. By virtue of the last inequality we also have

$$\|f - \sum_1^n f_i\|_p \leq \sum_{n+1}^\infty \|f_i\|_p \to 0,$$

and thus every absolutely summable sequence is summable, and L^p is complete. ∎

Theorem 3.4 (Cauchy-Schwarz-Bunyakovskii inequality) *If f and g are in L^2, then fg is in L^1 and*

$$\langle f, g\rangle = \int fg \leq \|f\|_2 \|g\|_2.$$

Proof. We can assume $f, g \geq 0$. For any $t > 0$ we have

$$\|f + tg\|^2 \leq (\|f\| + t\|g\|)^2 \leq \|f\|^2 + 2t\|f\|\|g\| + O(t^2),$$

while at the same time

$$\|f + tg\|^2 = \|f\|^2 + 2t\langle f, g\rangle + t^2\|g\|^2 \geq \|f\|^2 + 2t\langle f, g\rangle;$$

comparing terms of size $O(t)$, we find $\|f\|\|g\| \geq \langle f, g\rangle$. ∎

This *inner product* $\langle f, g\rangle$ is a symmetric, definite bilinear form making L^2 into a *Hilbert space*. It is an infinite-dimensional analogue of the inner product in $\mathbb{R}^n$. For example, if E and F are disjoint measurable sets, then $L^2(E)$ and $L^2(F)$ are orthogonal subspaces inside $L^2(\mathbb{R})$.

Hölder's inequality. Let's try to mimic this argument for $f \in L^p$ and $g \in L^q$, with $1/p + 1/q = 1$. Then $f^{p/q} \in L^q$, and using the binomial expansion for $(a+b)^q$ we have:

$$\begin{aligned} \|f^{p/q} + tg\|_q^q &\leq \|f^{p/q}\|_q^q + qt\|f^{p/q}\|_q^{q-1}\|g\|_q + O(t^2) \\ &= \|f\|_p^p + qt\|f\|_p\|g\|_q + O(t^2) \end{aligned}$$

since $(q-1)/q = 1/p$. On the other hand for $f, g \geq 0$ we have (by convexity of x^p),

$$\begin{aligned} \|f^{p/q} + tg\|_q^q &= \int |f^{p/q} + tg|^q \geq \int (f^{p/q})^q + qt(f^{p/q})^{q-1} g \\ &\geq \|f\|_p^p + qt \int fg. \end{aligned}$$

Putting these inequalities together gives the theorem. ■

Young's inequality. One can also prove Hölder's inequality using the fact that $ab \leq a^p/p + b^q/q$; this is on the homework.

(Proof of Young's inequality. Draw the curve $y = x^{p-1}$, which is the same as the curve $x = y^{q-1}$. Then the area inside the rectangle $[0,a] \times [0,b]$ is bounded above by the sum of a^p/p, the area between the graph and $[0,a]$, and b^q/q, the area between the graph and $[0,b]$.)

Theorem 3.5 *If $f : [a,b] \to \mathbb{R}$ is absolutely continuous, and $f' \in L^p[a,b]$, then f is Hölder continuous of exponent $1 - 1/p$.*

If $p = 1$ we get no information. If $p = \infty$ we get Lipschitz continuity.

Proof. We have

$$|f(x) - f(y)| = \left| \int_x^y 1 \cdot f'(t)\, dt \right| \leq \|\chi_{[x,y]}\|_q \|f'\|_p \leq \|f'\|_p |x-y|^{1/q}.$$

■

Theorem 3.6 (Density of simple functions) *For any p and E, simple functions are dense in $L^p(E)$. For $p \neq \infty$, step, continuous and smooth functions are dense in $L^p(\mathbb{R})$.*

Proof. First we treat $f \in L^\infty$. Then f is bounded, so it is a limit of simple functions in the usual way (cut the range into finitely many small intervals and round f down so it takes values in the endpoints of these intervals).

Now for $f \in L^p$, $p \neq \infty$, we can truncate f in the domain and range to obtain bounded functions with compact support, $f_M \to f$. Since $f - f_M \to 0$ pointwise and $|f - f_M|^p \leq |f|^p$, dominated convergence shows $\|f - f_M\|_p \to 0$. Finally we can find step, continuous or smooth functions $g_n \to F_M$ pointwise, and bounded in the same way. Then $\int |g_n - F_M|^p \to 0$ by bounded convergence, so such functions are dense. ■

Note! The step, continuous and smooth functions are **not** dense in L^∞!

***L^p* as a completion.** Given say $V = C_0^\infty(\mathbb{R})$ with the L^2-norm, it is exceedingly natural to form the metric completion $\overline{V}$ of V and obtain a Banach space. But what are the elements of this space? The virtue of measurable functions is that they do suffice to represent all elements of $\overline{V}$.

It is this completeness that makes measurable functions as important as real numbers.

Duality. Given a Banach space X, we let X^* denote the **dual space** of bounded linear functionals $\phi : X \to \mathbb{R}$, with the norm

$$\|\phi\| = \sup_{x \in X - \{0\}} \frac{|\phi(x)|}{\|x\|}.$$

There is a natural map $X \to X^{**}$. If $X = X^{**}$ then X is *reflexive.*

Theorem 3.7 (Riesz-Fischer) *Let $1/p + 1/q = 1$, with $1 < p, q < \infty$. Then $L^p[a,b]^* = L^q[a,b]$ and vice-versa; in particular, L^p is reflexive.*

Proof. Suppose $p \neq \infty$. Let $\phi : L^p[a,b] \to \mathbb{R}$ be a bounded linear functional, with $|\phi(f)| \leq M\|f\|$. Define

$$G(x) = \phi(\chi_{[a,x]}).$$

Then for any collection of disjoint intervals (a_i, b_i) with $\sum |a_i - b_i| < \delta$, we have

$$\begin{aligned} \sum |G(a_i) - G(b_i)| &= \sum |\phi(\chi_{(a_i,b_i)})| = \phi\left(\sum \pm \chi_{(a_i,b_i)}\right) \\ &\leq M\left(\sum |a_i - b_i|\right)^{1/p} \leq M\delta^{1/p}, \end{aligned}$$

since $\phi(f) \leq M\|f\|_p$. Thus $G(x)$ is absolutely continuous, and thus there is an integrable function $g(x) = G'(x) \in L^1[a,b]$ such that

$$\phi(\chi_I) = \int_I G'(x)\,dx = \int \chi_I g$$

for any interval $I \subset [a,b]$. (For $p = \infty$, the indicator functions $\chi_{[a,x]}$ do not depend continuously on x, so this step fails!)

Next we check that $\phi(f) = \langle f, g\rangle$ for all $f \in L^\infty[a,b]$. Indeed, if $|f| \leq A$ there are step functions $f_n \to f$ with $f_n \to f$ pointwise and $|f_n| \leq A$. Then $f_n \to f$ in L^p, so by continuity $\phi(f_n) \to \phi(f)$. On the other hand,

$|f_n g| \le A|g|$, and $\int |g| < \infty$, so by the dominated convergence theorem we have $\langle f_n, g\rangle \to \langle f, g\rangle$. Since $\phi(f_n) = \langle f_n, g\rangle$, we have $\phi(f) = \langle f, g\rangle$.

Now let g_n be the truncation of g to a function with $|g_n| \le n$, and choose the sign of g_n and g_n^α below so the products below are positive. Then:

$$\begin{aligned}\int g_n^q &= \langle g_n^{q-1}, g\rangle = \phi(g_n^{q-1}) \le M\|g_n^{q-1}\|_p \\ &= M\left(\int |g_n|^{(q-1)p}\right)^{1/p} = M\left(\int |g_n|^q\right)^{1/p},\end{aligned}$$

since $pq = p + q$. Thus for every n we have

$$\int |g_n|^q \le M^{1/(1-1/p)} = M^q.$$

Taking the limit as $n \to \infty$ and applying Fatou's lemma or monotone convergence, we have $\|g\|_q \le M$.

Thus by Hölder's inequality, $\langle f, g\rangle$ defines a continuous linear functional on $L^p[a, b]$. Since $\phi(f) = \langle f, g\rangle$ on the dense set of bounded functions, we conclude that equality holds everywhere.

Finally, Hölder's inequality shows

$$\langle f, g\rangle \le \|f\|_p \|g\|_q,$$

with equality for $f = g^{q/p}$, so the dual norm on $L^p[a, b]^*$ agrees with the L^q norm, $\|g\|_q$. ∎

On the other hand, $(L^1)^* = L^\infty$ but $(L^\infty)^* \ne L^1$.

The first assertion follows by a modification of the proof above. For an indication of the second, recall the analogous fact that we used an ultrafilter to construct a bounded linear function $L : \ell^\infty \to \mathbb{R}$, extending the usual function $L(a_n) = \lim a_n$ on $c \subset \ell^\infty$. On the other hand, for any $b \in \ell^1$ we can find $a \in c$ such that $L(a) = 1$ but $\langle a, b\rangle$ is as small as we like (slide the support of a off towards infinity.) Thus L is a linear functional that is not represented by any element of ℓ^1.

A similar construction can be carried out extending the point evaluations from $C[a, b]$ to $L^\infty[a, b]$.

4 Baire Category

Theorem 4.1 *Let X be a complete metric space, and let U_i be a sequence of* dense *open sets in X. Then $\bigcap U_i$ is dense. In particular the intersection is nonempty (so long as X is nonempty).*

Proof. We will define a nested sequence of closed balls $B_0 \supset B_1 \supset \dots$ by induction. Let B_0 be arbitrary. Since U_n is dense, it meets the interior of B_n; choose B_{n+1} to be any ball contained in $B_n \cap U_n$, with $\operatorname{diam} B_{n+1} \le 1/(n+1)$.

Then (if $X \neq \emptyset$), the centers of the balls B_n form a Cauchy sequence, so they converge to a limit $x \in X$. By construction, $x \in B_0 \cap \bigcap U_i$. Since B_0 was arbitrary, this shows $\bigcap U_i$ is dense. ■

Category. A subset E of a topological space is **nowhere dense** if the interior of $\overline{E}$ is empty. A space is of *first category* if it is a countable union of nowhere dense sets; otherwise it is of *second category.*

For example, $\mathbb{Q}$ is of first category, but $\mathbb{Z}$ is not (since every point of $\mathbb{Z}$ is open).

In a complete metric space, a countable union of nowhere dense sets is said to be *meager*; the complement of such a set is *residual.* A property is *generic* if it holds outside a meager set.

Reformulations of Baire's theorem. Let X be a nonempty complete metric space, or locally compact space topological space.

X is of second category.
A countable intersection of dense G_δ's in X is again a dense G_δ.
If $X = \bigcup F_i$ then $\operatorname{int} \overline{F_i} \neq \emptyset$ for some i.

Measure and category. The sets of measure zero and the meager sets in $\mathbb{R}$ both form σ*-ideals* (in the ring of all subsets of $\mathbb{R}$). That is, they are closed under taking subsets and countable unions.

Theorem 4.2 (Uniform boundedness) *Let $\mathcal{F}$ be a collection of continuous functions on a (nonempty) complete metric space X, such that for each x the functions are bounded — i.e. $\sup_{\mathcal{F}} f(x) \le M_x$. Then there is a open set $U \neq \emptyset$ on which the functions are* uniformly *bounded: $\sup_U f(x) \le M$ for all $f \in \mathcal{F}$.*

Proof. Let $F_n = \{x : f(x) \le n \, \forall f \in \mathcal{F}\}$. Then F_n is closed and $\bigcup F_n = X$, so some F_M has nonempty interior U. ■

Diophantine approximation. A real number x is *Diophantine of exponent* α if there is a $C > 0$ such that

$$\left| x - \frac{p}{q} \right| > \frac{C}{q^\alpha}$$

for all rational numbers p/q.

Theorem 4.3 *If x is algebraic of degree $d > 1$, then it is Diophantine of exponent d.*

Proof. Let $f(t) = a_0 t^d + \dots a_d$ be an irreducible polynomial with integral coefficients satisfied by x. Then $|f(p/q)| \geq 1/q^d$. Since $|f'|$ is bounded, say by M, near x, we find

$$q^{-d} \leq |f(x) - f(p/q)| \leq M|x - p/q|$$

and thus $|x - p/q| \geq 1/(Mq^d)$. ∎

Roth has proved the deep theorem that any algebraic number is Diophantine of exponent $2 + \epsilon$.

A number is Diophantine of exponent 2 iff the coefficients in its continued fraction expansion are bounded. For quadratic numbers, these coefficients are pre-periodic.

Liouville numbers. We say x is *Liouville* if x is irrational but for any $n > 0$ there exists a rational number with $|x - p/q| < q^{-n}$. Such a number is not Diophantine for any exponent, so it must be transcendental.

For example, $x = \sum 1/10^{n!}$ is an explicit and easy example of a transcendental number.

Theorem 4.4 (Measure vs. Category) *A random $x \in [0,1]$ is Diophantine of exponent $2+\epsilon$ for all $\epsilon > 0$. However a generic $x \in [0,1]$ is Liouville.*

Proof. For the first part, fix $\epsilon > 0$, and let

$$E_q = \{x \in [0,1] \,:\, \exists p,\ |x - p/q| < 1/q^{2+\epsilon}\}.$$

Since there are only q choices for p, we find $m(E_q) = O(1/q^{1+\epsilon})$, and thus $\sum m(E_q) < \infty$. Thus $m(\limsup E_q) = 0$ (by easy Borel-Cantelli). But this means that for almost every $x \in [0,1]$, only finitely rationals approximate x to within $1/q^{2+\epsilon}$. Thus x is Diophantine of exponent $2+\epsilon$. Taking a sequence $\epsilon_n \to 0$ we conclude that almost every x is Diophantine of exponent $2+\epsilon$ for all $\epsilon > 0$.

For the second part, just note that

$$E_n = \{x \in [0,1] \,:\, \exists p, q,\ |x - p/q| < 1/q^n\}$$

contains the rationals and is open. Thus $\bigcap E_n = L$ is the set of Liouville numbers, and by construction it is a dense G_δ. ∎

Sets with no category.

Lemma 4.5 *The set of closed subsets of* $\mathbb{R}$ *has the same cardinality as* $\mathbb{R}$ *itself.*

Lemma 4.6 *A closed subset of* $\mathbb{R}$ *with no isolated points contains a Cantor set.*

Lemma 4.7 *Every uncountable closed set* E *in* $\mathbb{R}$ *contains a Cantor set.*

Proof. Consider the subset F of $x \in E$ such that $F \cap B(x, r)$ is uncountable It is easy to see that F is a nonempty, closed set, without isolated points, using the fact that countable unions preserve countable sets. Thus F contains a Cantor set. ∎

Corollary 4.8 *Every uncountable closed set satisfies* $|F| = |\mathbb{R}|$.

Corollary 4.9 *Every set of positive measure contains a Cantor set.*

Proof. It contains a compact set of positive measure, which is necessarily uncountable. ∎

By similar arguments, it is not hard to show:

Theorem 4.10 *Every dense* G_δ $X \subset [a, b]$ *contains a Cantor set.*

Theorem 4.11 *There exists a set* $X \subset \mathbb{R}$ *such that* X *and* X' *both meet every uncountable closed set.*

Proof. Use transfinite induction, choosing an isomorphism between $\mathbb{R}$ and the smallest ordinal c with $|c| = |\mathbb{R}|$. ∎

Such an X is called a *Bernstein set.*

Corollary 4.12 *If* X *is a Bernstein set, then for any interval* $[a, b]$, *neither* $X \cap [a, b]$ *nor* $\widetilde{X} \cap [a, b]$ *has first category.*

Proof. If $X \cap [a, b]$ has first category, then the complement of X contains a Cantor set K, contradicting the fact that X meets K. The same reasoning applies to $\widetilde{X}$. ∎

Thus X is an analogue, in the theory of category, of a non-measurable set. (One can think of a set X that meets **some** open set in a set of second category, as a set of positive measure).

Games and category. (Oxtoby, §6.) Let $X \subset [0,1]$ be a set. Players A and B play the following game: they alternately choose intervals $A_1 \supset B_1 \supset A_2 \supset B_2 \cdots$ in $[0,1]$, then form the intersection $Y = \bigcap A_i = \bigcap B_i$. Player A wins if Y meets X, otherwise player B wins.

Theorem 4.13 *There is a winning strategy for B iff X has first category.*

Proof. If X has first category then it is contained in a countable union of nowhere dense closed sets, $\bigcup F_n$. Player B simply chooses B_n so it is disjoint from F_n, and then $\bigcap B_n$ is disjoint from X.

Conversely, suppose B has a winning strategy. Then using this strategy, we can find a set of disjoint 'first moves' B_1^i that are dense in $[0,1]$. To see this, let $B_1(A)$ be B's move if $A_1 = A$. Let $J_1, J_2, \ldots$ be a list of the intervals with rational endpoints in $[0,1]$. Inductively define $B_1^1 = B(J_1)$ and $B_1^{i+1} = B(J_k)$ for the first k such that J_k is disjoint from $B_1^1, \ldots, B_1^i$. Then every J_k meets some B_1^i so $\bigcup B_1^i$ is dense.

Similarly, we can find disjoint second moves that are dense in B_1^i for each i. Putting all these together, we obtain moves B_2^i, each contained in some B_1^i, that are also dense in $[0,1]$.

Continuing in this way, we obtain a sequence B_k^i such that $U_k = \bigcup_i B_k^i$ is dense in $[0,1]$. Let $Z = \bigcap U_k$. Any point $x \in Z$ is contained in a unique nested sequence $B_1^{i_1} \supset B_2^{i_2} \supset \cdots$ obtained using B's winning strategy. Thus $x \notin X$. This shows X is disjoint from the dense G_δ Z, and thus X has first category. ∎

By the same reasoning we have:

Theorem 4.14 *Player A has a winning strategy iff there is an interval A_1 such that $I \cap A_1$ has second category.*

Corollary 4.15 *There exists a set X such that neither A nor B has a winning strategy!*

One might try to take X equal to a non-measurable set $P \subset [0,1) \cong S^1 = \mathbb{R}/\mathbb{Z}$ constructed so that $\mathbb{Q} + P = S^1$. By the Baire category theorem, P does not have first category, but it also does not have second category, since $P \cap P + 1/2 = \emptyset$.

However it might be the case that $P \cap I$ is small (even empty!) for some interval I. To remedy this, one considers instead a *Bernstein set*, i.e. a set X such that both X and its complement X' meet every uncountable closed subset of S^1. Then, as we have seen above, $X \cap [a, b]$ has neither first nor second category.

Poincaré recurrence. Let X be a finite measure space, and let $T : X \to X$ be a measure-preserving automorphism. Then for any set A of positive measure, there exists an $n > 0$ such that $m(A \cap T^n(A)) > 0$.

Proof. Let $E = T(A) \cup T^2(A) \cup \ldots$ be the strict forward orbits of the elements of A. Then, if A is disjoint from its forward orbit, we find A and E are disjoint sets and $T(A \cup E) = E$. Thus $m(A \cup E) = m(E) = m(E) + m(A)$, so $m(A) = 0$. ■

Recurrence and category. Now suppose X is also a compact metric space, $T : X \to X$ is a measure-preserving homeomorphism, and every nonempty open set has positive measure. We say $x \in X$ is *recurrent* if x is an accumulation point of the sequence $T^n(x), n > 0$.

Theorem 4.16 *The set of recurrent points is residual and of full measure.*

Proof. If x is not a recurrent point, then there is a positive distance from x to the closure of its forward orbit. That is, for some $r > 0$ we have

$$x \in E_r = \{y \ : \ d(y, T^n(y)) \geq r, \forall n > 0\}.$$

Note that E_r is closed, and hence compact. We claim $m(E_r) = 0$. If not, there is a ball such that $A = B(x, r/2) \cap E_r$ has positive measure. But then A is disjoint from its forward orbit, contrary to Poincaré recurrence.

Thus E_r is a closed set of measure zero, and hence nowhere dense. Since the non-recurrent points are exactly the set $\bigcup_i E_{1/i}$, we see the recurrent points are residual and of full measure. ■

The space of homeomorphisms. Let X be a compact metric space. Let us make the space $C(X, X)$ of all continuous maps $f : X \to X$ into a complete metric space by $d(f, g) = \sup d(f(x), g(x))$. What can we say about the subset $H(X)$ of homeomorphisms?

It is easy to see $H([0, 1])$ is already neither open nor closed. However it does consist exactly of the bijective maps in $C(X, X)$. Now surjectivity *is* a closed condition, and hence a G_δ-condition. What about injectivity? If f

is not injective, then there are two points at definite distance, x and y, that are identified. Thus the non-injective maps are a union of closed sets,

$$\bigcup_n \{f \,:\, \exists x, y \in X, d(x,y) \geq 1/n, f(x) = f(y)\}.$$

(The closedness uses compactness of X). Putting these observations together we have:

Theorem 4.17 *For any compact metric space X, the homeomorphisms $H(X)$ are a G_δ subset of the complete metric space $C(X,X)$.*

The property of Baire. A topological space has the property of Baire if it satisfies the *conclusion* of the category theorem: namely if any intersection of dense G_δ's is still dense.

Theorem 4.18 *If X is complete and $Y \subset X$ is a G_δ, then Y has the property of Baire.*

Proof. Apply the Baire category theorem to $\overline{Y}$, in which Y is a dense G_δ. ■

(Actually one can re-metrize Y so Y itself is a complete metric space.)

Transitive maps of the square. Oxtoby and Ulam proved that a generic measure-preserving automorphism of any manifold is ergodic. We will prove a weaker result that illustrates the method.

Theorem 4.19 *There exists a homeomorphism of* $[0,1]\times[0,1]$ *with a dense orbit.*

Let $X = [0,1] \times [0,1]$. Since $H(X)$ has the property of Baire, it would suffice to show that a generic homeomorphism has a dense orbit. But this is not true! Once there is a disk such that $f(D) \subset D$, any orbit that enters D can never escape. And in fact any homeomorphism can be perturbed slightly so that $f^n(D) \subset D$ for some disk D. The category method has failed!

Measure-preserving maps. What Oxtoby and Ulam proved is that the problem can be solved by making it harder.

Theorem 4.20 *A generic* measure preserving *homeomorphism of the square has a dense orbit.*

Proof. Let $M(X) \subset H(X)$ be the measure-preserving homeomorphisms. It also has the property of Baire, because it is a closed subset of $H(X)$.

Consider two balls, B_1 and B_2, and let $U(B_1, B_2) \subset M(X)$ be the set of $f : X \to X$ such that $f^n(B_1)$ meets B_2, for some $n > 0$. Clearly $U(B_1, B_2)$ is open; we will show it is dense.

To this end, fix $r > 0$, and consider any $f \in M(X)$. Choose a chain of points $x_0, \ldots, x_n$ with $x_0 \in B_1$, $x_n \in B_2$, and $d(x_i, x_{i+1}) < r$. Since a generic point is recurrent, we can also assume each x_i is recurrent. Then we can also find high iterates $y_i = T^{n_i}(x_i)$ such that $d(x_i, y_i) < r$.

Now choose a short path P_i (of length less than $2r$) from y_i to x_{i+1}, avoiding all other of the points we have considered; including $T^n(x_i)$ for $0 \le n \le n_i$. Construct a measure-preserving map within distance r of the identity, such that $g(y_i) = x_{i+1}$. This map g is supported close to $\bigcup P_i$.

Then $g \circ f$, under iteration, moves x_0 to $y_{n_0 - 1}$, then to $g(f(y_{n_0-1})) = g(y_{n_0}) = x_1$, and then x_1 to x_2), etc.; so that ultimately x_n is in the forward orbit of x_0, and hence f moves B_1 into B_2.

Using a countable base for X, we can now conclude that a generic $f \in M(X)$ has the property that for any two nonempty open sets $U, V \subset X$, there exists an $n > 0$ such that $f^n(U) \cap V \neq \emptyset$.

We claim any such f has a dense orbit. Indeed, consider for any open ball B the set $U(B)$ of x such that $f^n(x) \in B$ for some $n > 0$. The set $U(B)$ is open, and it is dense by our assumption on f. Intersecting these $U(B)$ over a countable base for X, we find a *generic* $x \in X$ has a dense orbit. ■

Open problem. Does a *generic* C^1 diffeomorphism of a surface have a dense orbit? It is known that a sufficiently smooth diffeomorphism does *not* (KAM theory).

For more discussion, see [Ox, §18] and [Me, Thm. 4.3].

5 Topology

Topological spaces. The collection of open sets $\mathcal{T}$ satisfies: $X, \emptyset \in \mathcal{T}$; and finite intersections and arbitrary unions of open sets are open. Metric spaces give particular examples.

Compactness. A space is compact if every open cover has a finite subcover. Equivalent, any collection of closed sets with the finite intersection property has a nonempty intersection.

Theorem 5.1 . *A subset $K \subset \mathbb{R}^n$ is compact iff K is closed and bounded.*

Theorem 5.2 *A metric space (X, d) is compact iff every sequence has a convergent subsequence.*

The first result does not hold in general metric spaces: for example, the unit ball in $\ell^\infty(\mathbb{N})$ is closed and bounded but not compact. Similarly, the sequence of functions $f_n(x) = x^n$ is bounded in $C[0,1]$, but has no convergent subsequence.

The second, we will also see, does not hold in general topological spaces. Nevertheless both results can be modified so they hold in a general setting.

Total boundedness. A metric space is *totally bounded* if for any $r > 0$, there exists a covering of X by a finite number of r-balls. In $\mathbb{R}^n$, boundedness and total boundedness are equivalent; but the latter notion is much stronger in infinite-dimensional spaces, and gives the correct generalization of Theorem 5.1.

Theorem 5.3 *A metric space (X, d) is compact iff X is complete and totally bounded.*

Arzela-Ascoli. Here is application of compactness to function spaces.

Let $C(X)$ be the Banach space of continuous functions on a compact metric space (X, d). When does a set of functions $\mathcal{F} \subset C(X)$ have compact closure? That is, when can we assure that every sequence $f_n \in \mathcal{F}$ has a convergent subsequence (whose limit may or may not lie in $\mathcal{F}$)?

Recall that $C(X)$ is complete, and that a metric space is compact iff it is complete and totally bounded. The latter property means that for any $r > 0$ there is a finite covering of X by r-balls.

The set $\mathcal{F}$ is *equicontinuous* if all the functions satisfy the same modulus of continuity: that is, if there is a function $m(s) \to 0$ as $s \to 0$ such that $d(x, y) < s$ implies $|f(x) - f(y)| < m(s)$ for all $f \in \mathcal{F}$. Of course $\mathcal{F}$ is *bounded* iff there is an M such that $|f(x)| \le M$ for all $f \in \mathcal{F}$.

Theorem 5.4 *$\mathcal{F} \subset C(X)$ has compact closure iff $\mathcal{F}$ is bounded and equicontinuous.*

Proof. First suppose $\overline{\mathcal{F}}$ is compact. Then clearly $\mathcal{F}$ is bounded. Now take any $\epsilon > 0$, and cover $\mathcal{F}$ by a finite collection of balls $B(f_i, \epsilon/3)$. Since X is compact, each f_i is uniformly continuous, so there is a δ such that

$$d(x, y) < \delta \implies \forall i,\ |f_i(x) - f_i(y)| < \epsilon/3.$$

Then for any $f \in \mathcal{F}$, we can find f_i with $d(f, f_i) < \epsilon/3$, and conclude that $|f(x) - f(y)| < \epsilon$ when $d(x, y) < \delta$. Thus $\mathcal{F}$ is equicontinuous.

Now suppose $\mathcal{F}$ is bounded by M, and equicontinuous. To show $\overline{\mathcal{F}}$ is compact, we need only show $\mathcal{F}$ is totally bounded. To this end, fix $r > 0$, and by equicontinuity choose $\delta > 0$ such that $d(x, y) < \delta \implies |f(x) - f(y)| < r$. By compactness of X, we can find a finite set $E \subset X$ such that $B(E, \delta) = X$. Similarly we can pick a finite set $F \subset [-M, M]$ that comes within r of every point.

For each map $g : E \to F$, let

$$B(g) = \{f \in \mathcal{F} \,:\, \sup_E |g - f| < r\}.$$

Since there are only finitely many maps g, and every f is close to some g, these sets give a finite cover $\mathcal{F}$. Finally if $f_1, f_2 \in B(g)$, then for any $x \in X$, there is an $e \in E$ within δ of x. We then have

$$|f_1(x) - f_2(x)| \le 2r + |f_1(e) - f_2(e)| \le 4r,$$

so $\operatorname{diam} B(g) \le 4r$. It follows that $\mathcal{F}$ is totally bounded, and thus $\overline{\mathcal{F}}$ is compact. ■

Example: Normal families. Let $\mathcal{F}$ be the set of all analytic functions on an open set $\Omega \subset \mathbb{C}$ with $|f(z)| \le M$. Then $\mathcal{F}$ is compact in the topology of uniform convergence on compact sets.

Note: The functions $f_n(z) = z^n$ do *not* converge uniformly on the whole disk, so the restriction to compacta is necessary.

Proof. By Cauchy's theorem, if $d(z, \partial\Omega) > r$, then

$$|f'(z)| = \left| \frac{1}{2\pi i} \int_{S^1(p,r)} \frac{f(\zeta)\, d\zeta}{(\zeta - z)^2} \right| \le \frac{2\pi r M}{2\pi r^2} = \frac{M}{r}\cdot$$

Thus we can pass to a subsequence converging uniformly on the compact set $K_r = \{z \in \Omega \,:\, d(z, \partial\Omega) \ge r\}$. Diagonalizing, we get a subsequence converging uniformly on compact sets. Analyticity is preserved in the limit, so $\mathcal{F}$ is a normal family. ■

Metrizability, Topology and Separation. Our next goal is to formulate purely topological versions of the best properties of metric spaces. This properties will help us recognize when a topological space $(X, \mathcal{T})$ is *metrizable*, i.e. when there is a metric d that determines the topology $\mathcal{T}$.

Given any collection $\mathcal{C}$ of subsets of X, there is always a weakest topology $\mathcal{T}$ containing that collection. We say $\mathcal{C}$ *generates* $\mathcal{T}$.

A *base* $\mathcal{B}$ for a topology is a collection of open sets such that for each $x \in U \in \mathcal{T}$, there is a $B \in \mathcal{B}$ with $x \in B \subset U$. Then U is the union of all the B it contains, so $\mathcal{B}$ generates $\mathcal{T}$. Indeed $\mathcal{T}$ is just the union of the empty set and all unions of subsets of $\mathcal{B}$.

If $\mathcal{B}$ is given, it is a base for some topology iff for any $x \in B_1, B_2$ there is a B_3 with $x \in B_3 \subset B_1 \cap B_2$.

A *sub-base* $\mathcal{B}$ for a topology $\mathcal{T}$ is a collection of open sets such that for any $x \in U \in \mathcal{T}$, we have $x \in B_1 \cap \cdots B_n \subset U$ for some $B_1, \ldots, B_n \in \mathcal{B}$. Any sub-base also generates $\mathcal{T}$. Conversely, any collection of set $\mathcal{B}$ covering X forms a sub-base for the topology $\mathcal{T}$ it generates.

Example: In $\mathbb{R}^n$, the open half-spaces $H = \phi^{-1}(a, \infty)$ for linear functions $\phi : \mathbb{R}^n \to \mathbb{R}$ form a sub-base for the topology. (By intersecting them we can make small cubes).

A *base at x* is a collection of open sets $\mathcal{B}_x$, all containing x, such that for any open U with $x \in U$, there is a $B \in \mathcal{B}_x$ with $x \in B \subset U$.

Example: in any metric space, the balls $B(x, 1/n)$ form a base at x.

Countability axioms. A topological space X is:

- *first countable* if every point has a countable base;
- *second countable* if there is a countable (sub-)base for the whole space; and
- *separable* if there is a countable dense set $S \subset X$.

Clearly a second countable space is separable: just choice one point from each open set.

Examples. Clearly any metric space is first countable.

A Euclidean spaces $\mathbb{R}^n$ are first and second countable, and separable.

The space $\ell^\infty(\mathbb{N})$ is *not* separable or second countable. The uncountable collection of balls $B(\chi_A, 1/2)$, as A ranges over all subsets of $\mathbb{N}$, are disjoint. On the other hand, $\ell^p(\mathbb{N})$ is separable for $1 \leq p < \infty$.

Theorem 5.5 *Any separable metric space is second countable.*

Proof. Let (x_i) be a countable dense set of let $\mathcal{B} = \{B(x_i, 1/n)\}$. Then if $x \in U$, we have $x \in B(x, r) \subset U$, and hence $x \in B(x_i, 1/n) \subset U$ as soon as $d(x_i, x) < 1/n$ and $2/n < r$. ■

Theorem 5.6 *The number of open (or closed) sets in a separable metric space (like $\mathbb{R}^n$) is at most $|\mathbb{R}|$.*

Proof. $|\mathcal{T}| \leq |\mathcal{P}(\mathcal{B})| \leq |\mathcal{P}(\mathbb{N})| = |\mathbb{R}|$.

Corollary 5.7 *There are more subsets of $\mathbb{R}$ than there are closed subsets.*

Question. Does first countable and separable imply second countable? No!

Example. The half-open interval topology. Let

$$\mathcal{B} = \{[a, b) \,:\, a < b\};$$

this is a base for a topology $\mathcal{T}$ on $\mathbb{R}$. In this topology, $x_n \to y$ iff x_n approaches y from above. Thus every strictly increasing sequence diverges.

This space is first countable and separable. (The rationals are dense.) But it is *not* second countable! If $a \in B \subset [a, b)$, then a must be the minimum of B. Thus for any base $\mathcal{B}$, the map $B \mapsto \inf B$ sends $\mathcal{B}$ onto $\mathbb{R}$, and therefore $|\mathcal{B}| \geq |\mathbb{R}|$.

Cor: $(\mathbb{R}, \mathcal{T})$ is not metrizable.

This space is sometimes denoted $\mathbb{R}_\ell$; for an extended discussion, see Munkres, *Topology.*

The Lindelöf condition. A topological space is said to be *Lindelöf* if every open cover has a *countable* subcover. A second countable space is Lindelöf. The space $\mathbb{R}_\ell$ above is also Lindelöf, but not second countable. It is interesting to note that the Sorgenfrey (carefree?) plane, $\mathbb{R}_\ell \times \mathbb{R}_\ell$ is **not** Lindelöf (cf. Munkres, *Topology*, p. 193).

Separation axioms $\mathbf{T}_{\boldsymbol{i}}$. (T for Tychonoff). Let us say disjoint subsets E, F of a topological space X can be *separated* if they lie in disjoint open sets. The separation axioms (or properties) are:

T_1 (Tychonoff): Points are closed.
T_2 (Hausdorff): Pairs of points x, y are separated.
T_3 (regular): Points are separated from closed sets, and points are closed.
T_4 (normal): Pairs of closed sets are separated, and points are closed.

Example: any metric space is normal. Given two closed sets A and B, they are separated by the open sets $U = \{x : d(x,A) < d(x,B)\}$ and $V = \{x : d(x,B) < d(x,A)\}$.

Zariski topology. Let k be a field. A natural example of a topology that is not Hausdorff is the Zariski topology on k^n. In this topology, a set is F *closed* if it is defined by system of polynomial equations: F is the zero set of a collection of polynomials $f_\alpha \in k[x_1, \ldots x_n]$.

A base for the topology consists of complements of hypersurfaces, $U_f = k^n - Z(f)$. Note that $U_f \cap U_g = U_{fg}$, so we indeed have a base.

By the Noetherian property, the ideal (f_α) is finitely generated, so only a finite number of polynomials are actually necessary to define F. Geometrically, this means any decreasing sequence of closed sets, $F_1 \supset F_2 \supset F_3 \ldots$, eventually stabilizes. In particular, $\mathbb{R}^2$ is **compact** .

On $\mathbb{R}$, the Zariski topology is the cofinite topology. On $\mathbb{R}^n$, any two nonempty open sets meet; i.e. $\mathbb{R}^n$ cannot be covered by a finite number of hypersurfaces. Thus the Zariski topology is T_1 but not T_2.

The spectrum of a ring. Given a ring A, one also defines the Zariski topology on the set $\operatorname{Spec} A$ of all prime ideals $p \subset A$, by taking the closed sets to have the form $V(a) = \{p : p \supset a\}$, where a ranges over all ideals in A. A point $p \in \operatorname{Spec} A$ is closed iff p is a maximal ideal.

Thus $\operatorname{Spec} A$ is usually not even Hausdorff. In fact, for any ring A, the 'generic point' p coming from the ideal (0) is dense; its closure is the whole space.

Theorem 5.8 *A compact Hausdorff space X is normal.*

Proof. We first show X is regular. Let p be a point outside a closed set F. Then for each $x \in F$ there are disjoint open sets $x \in U_x$ and $p \in V_x$. Passing to a finite subcover of F, we have $F \subset \bigcup_1^n U_i$ and $p \in \bigcap_1^n V_i$.

Now to prove normality, suppose E and F are disjoint closed sets. Then for each $x \in E$, there is are disjoint open sets with $x \in U_x$ and $F \subset V_x$. Passing to a finite subcover, we have $E \subset \bigcup_1^n U_i$ and $F \subset \bigcap_1^n V_i$. ∎

Theorem 5.9 (Urysohn's Lemma) *Let A, B be disjoint closed subsets of a normal space X. Then there exists a continuous function $f : X \to [0,1]$ such that $f(A) = 0$ and $f(B) = 1$.*

Proof. Let $U_0 = A$ and let $U_1 = X$. The closed set A is a subset of the open set $\widetilde{B}$. By normality, there exists an open set $U_{1/2}$ with $A \subset U_{1/2} \subset$

$\overline{U_{1/2}} \subset \widetilde{B}$. Iterating this construction, we obtain a family of open sets U_r indexed by the dyadic rationals in $[0,1]$ such that $U_r \subset \overline{U_r} \subset U_s$ whenever $r < s$. Now let $f(x) = \inf\{r : x \in U_r\}$. Then $\{x : f(x) < s\} = \bigcup_{r<s} U_r$ is open, and $\{x : f(x) \le s\} = \bigcap_{r>s} \overline{U_r}$ is closed, so f is continuous. ■

Corollary 5.10 *In a normal space, there are sufficiently many functions $f : X \to \mathbb{R}$ to generate the topology on X.*

Proof. We must show that every closed set A is the intersection of level sets of functions. But for any $p \notin A$ we can find a function with $f(A) = 0$, $f(p) = 1$, and so we are done. ■

Theorem 5.11 (Tietze Extension) *If X is normal and $A \subset X$ is closed, then every continuous function $f : A \to \mathbb{R}$ extends to a continuous function on X.*

Actually Tietze generalizes Urysohn, since the obvious function $f : A \sqcup B \to \{0,1\}$ is continuous and $A \sqcup B$ is closed.

Approximating sets by submanifolds. For any compact set $X \subset \mathbb{R}^n$, and $r > 0$, there exists a smooth compact submanifold lying within $B(X, r)$ and separating X from ∞.

Proof. Smooth the function given by Tietze and apply Sard's theorem.

Cor. Any compact set in $\mathbb{R}^2$ can be surrounded by a finite number of smooth loops. Any Cantor set in $\mathbb{R}^3$ can be surrounded by smooth closed surfaces; but their genus may tend to infinity! (Antoine's necklace).

Weak topology. Given a collection of functions $\mathcal{F}$ on a set X, we can consider the weakest topology which makes all $f \in \mathcal{F}$ continuous. A base for this topology is given by the sets of the form

$$f_1^{-1}(\alpha_1, \beta_1) \cap f_2^{-1}(\alpha_2, \beta_2) \cap \ldots f_n^{-1}(\alpha_n, \beta_n),$$

where $f_i \in \mathcal{F}$. We have $x_n \to x$ iff $f(x_n) \to f(x)$ for all $f \in \mathcal{F}$.

The *weak topology* on a Banach space X is the weakest topology making all $\phi \in X^*$ continuous. For example, $f_n \to f$ weakly in L^1 iff

$$\int f_n g \to \int f g$$

for all $g \in L^\infty$. This topology is weaker than norm convergence; e.g. the functions $f_n(x) = \sin(nx)$ converge weakly to zero in $L^1[0,1]$, but they do not convergence at all in the norm topology.

Products. Given any collection of topological spaces $\{X_\alpha\}$, the product $X = \prod X_\alpha$ can be endowed with the Tychonoff topology, defined by the sub-basic sets $B(U,\alpha) = \{x \in X \,:\, x_\alpha \in U\}$ where $U \subset X_\alpha$ is open.

This is the weakest topology such that all the projections $f_\alpha : X \to X_\alpha$ are continuous.

Example: For any set A, $\mathbb{R}^A$ is the set of all functions $f : A \to \mathbb{R}$, and $f_n \to f$ in the Tychonoff topology iff $f_n(a) \to f(a)$ for all $a \in A$. So the Tychonoff topology is sometimes called the topology of pointwise convergence.

Example: In $X = (\mathbb{Z}/2)^A \cong \mathcal{P}(A)$, we have $A_n \to A$ iff ($x \in A$ iff $x \in A_n$ for all $n \gg 0$).

Theorem 5.12 *If X_i is metrizable for $i = 1, 2, 3, \ldots$, then so is $\prod_1^\infty X_i$.*

Proof. Replacing the metric d_i by $\min(d_i, 1)$, we can assume each X_i has diameter at most 1. Then $d(x,y) = \sum 2^{-i} d(x_i, y_i)$ metrizes $\prod X_i$. ∎

For example, $\mathbb{R}^{\mathbb{N}}$ is metrizable.

Theorem 5.13 (Urysohn metrization theorem) *A second countable topological space X is metrizable iff X is normal.*

Proof. Clearly a metric space is normal. For the converse, let (B_i) be a countable base for X. For each pair with $\overline{B_i} \subset B_j$, construct a continuous function $f_{ij} : X \to [0,1]$ with $f_{ij} = 0$ on B_i and $f_{ij} = 1$ outside B_j. Let $\mathcal{F}$ be the collection of all such functions, and consider the natural continuous map $f : X \to [0,1]^{\mathcal{F}}$, sending x to $(f_{ij}(x))$. Since $\mathcal{F}$ is countable, $f(X)$ is metrizable; we need only show that the inverse map $f(X) \to X$ is defined and continuous.

To see the map $f(X) \to X$ is defined, we must show f is injective. But given any points $x \neq y$, we can find open sets with $x \in \overline{B_i} \subset B_j$ and y outside B_j; then f_{ij} separates x from y.

To see $f(X) \to X$ is continuous, we just need to show that the weakest topology $\mathcal{T}'$ making all the functions f_{ij} continuous is the original topology $\mathcal{T}$ on X. But if $x \in U \in \mathcal{T}$, then there are basis elements with $x \in \overline{B_i} \subset B_j \subset U$. Then $V = f_{ij}^{-1}[0,1/2)$ is in $\mathcal{T}'$, and we have $x \in B_i \subset V \subset B_j \subset U$. Since this holds for every $x \in U$, we conclude that $U \in \mathcal{T}'$ and thus $\mathcal{T} = \mathcal{T}'$. ∎

Regularity v. Normality. Tychonoff observed that Urysohn's metrization theorem also applies to regular spaces, since we have:

Theorem 5.14 *A regular space with a countable base is normal.*

Proof. Let A, B be disjoint closed sets in such a space. Then A is covered by a countable collection of open sets U_i whose closures are disjoint from B. There is a similar cover V_i of B by open sets whose closures are disjoint from A. Now set $U_i' = U_i - (\overline{V}_1 \cup \cdots \overline{V}_i)$, set $V_i' = V_i - (\overline{U}_1 \cup \cdots \overline{U}_i)$, and observe that $U = \bigcup U_i'$ and $V = \bigcup V_i'$ are disjoint open sets containing A and B. ∎

A non-metrizable product. Example: $(\mathbb{Z}/2)^{\mathbb{R}} \cong \mathcal{P}(\mathbb{R})$ is not metrizable because it is not first countable.

A base at the set $\mathbb{R}$ consists of the open sets $U(F)$, defined for each finite set $F \subset \mathbb{R}$ as

$$U(F) = \{A \subset \mathbb{R} \,:\, F \subset A\}.$$

Let $\mathcal{F}$ be the set of finite subset $A \subset \mathbb{R}$. Then $\mathcal{F}$ meets every $U(F)$ so $\mathbb{R} \in \overline{\mathcal{F}}$. But there is no sequence $A_n \in \mathcal{F}$ such that $A_n \to \mathbb{R}$! Indeed, if $A_n \in \mathcal{F}$ is given then we can pick $x \notin \bigcup A_n$, and A_n never enters the neighborhood $U(\{x\})$ of $\mathbb{R}$.

We will later see that that $\mathcal{P}(\mathbb{R})$ is compact. But it has sequences with no convergent subsequences! To see this, let A_n be the set of real numbers $x = 0.x_1x_2x_3\ldots$ such that $x_n = 1$. Given any subsequence n_k, we can find an x such that x_{n_k} alternates between 1 and 2 as $n \to \infty$. Suppose $A_{n_k} \to B$. If $x \in B$ then $x \in A_{n_k}$ for all $k \gg 0$, and if $x \notin B$ then $x \notin A_{n_k}$ for all $k \gg 0$. Either way we have a contradiction.

Nets. A *directed system* A is a partially ordered set so any two $\alpha, \beta \in A$ are dominated by some $\gamma \in A$: $\gamma \geq \alpha$ and $\gamma \geq \beta$.

A *net* x_α is a map $x : A \to X$ from a directed system into a topological space X.

Example: $\mathbb{N}$ is a directed system, and a sequence x_n is a net.

Convergence. We say $x_\alpha \to x \in X$ iff for any neighborhood U of x there is an α such that $x_\beta \in U$ for all $\beta > \alpha$.

Theorem 5.15 *In any topological space, $x \in \overline{E}$ iff there is a net $x_\alpha \in E$ converging to x.*

Proof. Let $\alpha = \alpha(U)$ range over the directed set of neighborhoods of x in X, and for each U let x_α be an element of $U \cap E$. Then $x_\alpha \to x$.

Conversely, if $x_\alpha \in E$ converges to x, then every neighborhood of x meets E, so $x \in \overline{E}$. ∎

Subnets. If B is also a directed system, a map $f : B \to A$ is *cofinal* if for any $\alpha_0 \in A$ there is a $\beta_0 \in B$ such that $f(\beta) \geq \alpha_0$ whenever $\beta \geq \beta_0$. Then $y_\beta = x_{f(\beta)}$ is a *subnet* of x_α.

Example: A function $f : \mathbb{N} \to \mathbb{N}$ is cofinal iff $f(n) \to \infty$. So subsequences are special cases of subnets.

Theorem 5.16 *X is compact iff every net has a convergent subnet.*

Proof. Let $\mathcal{F}$ be a collection of closed sets with the finite intersection property, and let α be the directed system of finite subsets of $\mathcal{F}$, and let x_α be a point lying in their common intersection. Then the limit point y of a convergent subset of x_α will lie in every element of $\mathcal{F}$, so $\bigcap \mathcal{F} \neq \emptyset$.

Conversely, let x_α be a net in a compact space X. For every α let

$$F_\alpha = \{x_\beta \ : \ \beta \geq \alpha\}.$$

Since the index set is directed, any finite set of indices has an upper bound, and thus the $\overline{F_\alpha}$ have the finite intersection property. Therefore there is a y in $\bigcap \overline{F_\alpha}$.

Now let B be a base at Y ordered by inclusion, and let $C = A \times B$ with the product ordering. (This means $(a, b) < (a', b')$ iff $a < a'$ and $b < b'$.) Then the projection $A \times B \to A$ is cofinal.

For every pair $\gamma = (\alpha, \beta(U))$ there is an element $y_\gamma = x_{f(\gamma)} \in U \cap F_\alpha$. Then y_γ is a subnet converging to y. ∎

Theorem 5.17 (Tychonoff) *A product $X = \prod_N X_n$ of compact sets is compact. (Here N is an arbitrary index set).*

Proof. By the Axiom of Choice we may assume the index set is an ordinal $N = \{0, 1, 2, \ldots\}$. Given a net $x_\alpha \in X$, we will produce a convergent subnet y_α, by transfinite induction over N. In the process we will define nets x^n for each $n \in N$, with x^n a subnet of x^i for $i < n$, and with each coordinate $x^n(i)$ converging for $i < n$. We will have $f_{ij} : A_i \to A_j$ denote the re-indexing function for $i \geq j$.

Let $y^0 = x$. Passing to a subnet, we obtain a net x^0_α indexed by $\alpha \in A_0$ and with $x^0_\alpha(0)$ converging in X_0.

Given $n \in N+1$, let $B_n = \sqcup_{i<n} A_i$, and let $y^n(\alpha) = x^i_\alpha$ for $\alpha \in A_i$. Order A_n by $\alpha \le \beta$ if α and β belong to A_i and A_j with $i \le j$, and if $f_{ji}(\beta) \ge \alpha$. Finally to make y^n a subnet of x^i, let $g_{nj}(\alpha) = f_{ij}(\alpha)$ if $\alpha \in A_i$, $i \ge j$, and specify $g_{nj}(\alpha) \in A_j$ arbitrarily if $\alpha \in A_i$ for $i < j$.

(Check that this is a subnet: given $\alpha_0 \in A_i$, if $\beta \ge \alpha_0$, then $\beta \in A_j$ for some $j \ge i$, and by definition of the ordering on B_n we have $f_{ji}(\beta) \ge \alpha_0$, so $g_{ni}(\beta) \ge \alpha_0$.)

Since y^n is a subnet of x^i, the net $y^n_\alpha(i)$ converges for all indices $i < n$. Let (x^n, A_n) be a subnet of (y^n, B_n) that converges in position n.

By induction we obtain, for the ordinal $N+1$, a subnet $y_\alpha = y^{N+1}_\alpha$ that converges in all coordinates. This means y_α converges in X. ■

Axiom of Choice. The use of the Axiom of Choice in the preceding proof cannot be dispensed with, in the strong sense that Tychonoff's theorem *implies* the Axiom of Choice. Note that this is stronger than the commonly-heard statement 'you need the Axiom of Choice to construct a non-measurable set'.

Partitions of unity.

Theorem 5.18 *Let X be a compact Hausdorff space, and let $\mathcal{U}$ be an open cover of X. Then there is a finite subcover U_i and functions $0 \le f_i(x) \le 1$ supported on U_i such that $\sum_1^n f_i(x) = 1$.*

Proof. For each $x \in X$ there is an open set $U \in \mathcal{U}$ and a continuous function $f \ge 0$ supported in U, such that $f(x) = 1$. By compactness there is a finite set of such functions such that the open sets $\{x : f_i(x) > 0\}$ cover X. Then $g(x) = \sum f_i(x) > 0$ at every point; replacing $f_i(x)$ by $f_i(x)/g(x)$ gives the desired result. ■

Lebesgue number. Corollary. Given an open covering $\mathcal{U}$ of a compact metric space X, there is an $r > 0$ such that for every $x \in X$, there is a $U \in \mathcal{U}$ with $B(x,r) \subset U$. The number r is called the *Lebesgue number* of $\mathcal{U}$.

Proof. Construct a partition of unity subordinate to $U_1, \ldots, U_n \in \mathcal{U}$; then for every x there is an i such that $f_i(x) \ge 1/n$, and by uniform continuity of the functions f_i there is an $r > 0$ such that $f_i(x) > 0$ on $B(x,r)$; then $B(x,r) \subset \{f_i > 0\} \subset U_i \in \mathcal{U}$. ■

Local constructions.

Theorem 5.19 *Any compact manifold X admits a metric.*

Proof. Take a finite collection of charts $\phi : U_i \to \mathbb{R}^n$, a partition of unity f_i subordinate to U_i, and let $g(v) = \sum f_i |D\phi_i(v)|^2$. ∎

Maximal ideals in $C(X)$.

Theorem 5.20 *Let X be a compact Hausdorff space; then the maximal ideals in the algebra $C(X)$ correspond to the point evaluations.*

Proof. Let $I \subset C(X)$ be a (proper) ideal. Suppose for all $x \in C(X)$, I is not contained in the maximal ideal M_x of functions vanishing at x. Then we can find for each x a function $f \in I$ not vanishing on a neighborhood of x. By compactness, we obtain $g = f_1^2 + \cdots + f_n^2$ vanishing nowhere. Then $(1/g)g \in I$ so $I = C(X)$, contradiction. So I is contained an some M_x. ∎

Spectrum. Given an algebra A over $\mathbb{R}$, let

$$\sigma(f) = \{\lambda \in \mathbb{R} : \lambda + f \text{ has no inverse in } A\}.$$

Then for $A = C(X)$, we have $\sigma(f) = f(X)$, and thus we can reconstruct $\|f\|_\infty$ from the algebraic structure on A.

Also for $A = C(X)$ we can let Y be the set of multiplicative linear functionals, and embed Y into $\mathbb{R}^A$ by sending ϕ to the sequence $(\phi(f) : f \in A)$. Then in fact $\phi(f) \in [-\|f\|, \|f\|]$, so Y is compact, and Y is homeomorphic to X.

Local compactness. A topological space X is **locally compact** if the open sets U such that $\overline{U}$ is compact form a base for the topology.

For example, $\mathbb{R}^n$ is locally compact.

Alexandroff compactification. Let X be a locally compact Hausdorff space, and let $X^* = X \cup \{\infty\}$, and define a neighborhood base at infinity by taking the complements $X^* - K$ of all compact sets $K \subset X$.

Theorem 5.21 *X^* is a compact Hausdorff space, and the inclusion of X into X^* is a homeomorphism.*

Proof. Compact: if you cover X^*, once you've covered the point at infinity, only a compact set is left. Hausdorff: because of local compactness, every $x \in X$ is contained in a U such that $\overline{U}$ is compact, and hence $V = X^* - \overline{U}$ is a disjoint neighborhood of infinity. ∎

This space is called the *one-point compactification* of X.

Examples: $\mathbb{N}^*$; $S^n = (\mathbb{R}^n)^*$.

Proper maps. A useful counterpart to local compactness is the notion of a *proper map.* A map $f : X \to Y$ is **proper** if $f^{-1}(K)$ is compact whenever K is compact. Intuitively, if x_α leaves compact sets in X, then $f(x_\alpha)$ leaves compact sets in $f(X)$. Thus $x_\alpha \to \infty$ implies $f(x_\alpha) \to \infty$, and so f extends to a continuous map from X^* to Y^*. This shows:

Theorem 5.22 *A continuous bijection between locally compact Hausdorff spaces is a homeomorphism iff it is proper.*

Example: There is a bijective continuous map $f : \mathbb{R} \to S^1 \cup [1, \infty) \subset \mathbb{C}$.

The Stone-Čech compactification.

Theorem 5.23 *Let X be a normal space. Then there is a unique compact Hausdorff space $\beta(X)$ such that:*

1. *X is dense in $\beta(X)$;*
2. *Every bounded continuous $f : X \to \mathbb{R}$ extends to a continuous function on $\beta(X)$;*
3. *If X is compactified by another Hausdorff space Y, in the sense that the inclusion $X \subset Y$ is dense, then $\beta(X)$ is bigger than Y: there is a continuous map $\phi : \beta(X) \to Y$.*

Proof. Let $\mathcal{F}$ be the family of *all* continuous $f : X \to [0, 1]$, let Z be the compact product $[0, 1]^{\mathcal{F}}$, and let $\beta(X) \subset Z$ be the closure of X under the embedding $x \mapsto (x_f)$ where $x_f = f(x)$. The first two properties are now evident.

Finally let Y be another compactification of X, and let $\mathcal{G}$ be the family of all continuous maps $g : Y \to [0, 1]$. Then there is an embedding $Y \subset [0, 1]^{\mathcal{G}}$, and the inclusion $\mathcal{G} \subset \mathcal{F}$ gives a natural projection map $[0, 1]^{\mathcal{F}} \to [0, 1]^{\mathcal{G}}$. This projection sends $\beta(X)$ into Y. ■

Example: $X = \beta(\mathbb{N})$. In this space, a sequence $x_n \in \mathbb{N}$ converges iff it is eventually constant. Thus X is compact but the sequence $x_n = n$ has no convergent subsequence! (However it does have convergent subnets; for such a net, $f(n_\alpha)$ converges for every $f \in \ell^\infty(\mathbb{N})$!)

Stone-Čech and dual spaces. Another way to look at $\beta(\mathbb{N})$ is that each $n \in \mathbb{N}$ provides a map $n : \ell^\infty(\mathbb{N}) \to \mathbb{R}$ by $a \mapsto a_n$, and that $\beta(\mathbb{N})$ is the closure of these maps. Note that the maps in the closure are **bounded, linear functionals** . A typical example is provided by the ultrafilter limit we constructed before. In general the closure consists of those finitely-additive measures on $\mathbb{N}$ such that $\mu(\mathbb{N}) = 1$ and $\mu(E) = 0$ or 1 for all $E \subset \mathbb{N}$.

Theorem 5.24 (Stone-Weierstrass) *Let X be a compact Hausdorff space, and let $A \subset C(X)$ be a subalgebra that contains the constants and separates points. Then A is dense in $C(X)$.*

Examples: in $C[0,1]$, the functions of bounded variation, or Lipschitz, or C^k, or Hölder continuous, or polynomials, or those that are real-analytic, all form subalgebras that separate points and contain the constant.

Lemma. The closure of A is a lattice.

Proof. We must show that $f, g \in A \implies f \vee g \in A$, where $(f \vee g)(x) = \max(f(x), g(x))$. Note that $f \vee g$ is the average of $f + g$ and $|f - g|$. So it suffices to show $f \in A \implies |f| \in A$. Now if $\epsilon < f < 1$, then $\sqrt{f} = \sqrt{1 - (1 - f)} \in A$, because $\sqrt{1 - x} = \sum a_n x^n$ can be expanded in a power series convergent in $B(0,1)$, and hence uniformly convergent in $B(0, 1 - \epsilon)$. Then

$$|f| = \lim_{\epsilon \to 0} \sqrt{f^2 + \epsilon},$$

so $|f|$ is in A, and hence $f \vee g$ is in A. ∎

Proof of Stone-Weierstrass. As above we replace A by its closure; then A is an algebra as well as a lattice.

Given $g \in C(X)$, let $\mathcal{F} = \{f \in A \, : \, f \geq g\}$. To show $g \in A$, it suffices to show for each x that $g(x) = \inf_{\mathcal{F}} f(x)$. Indeed, if that is the case, then for any $\epsilon > 0$ and $x \in X$, there is a neighborhood U of x and an $f \in \mathcal{F}$ such that $g \leq f \leq g + \epsilon$ on U. Taking a finite sub-cover, we obtain a finite number of functions such that $g \leq f_1 \wedge \cdots \wedge f_n \leq g + \epsilon$ on all of X. Since A is a lattice, we are done.

It remains to construct, for $\epsilon > 0$ and $x \in X$, and function $f \in A$ such that $f \geq g$ and $g(x) \leq f(x) \leq g(x) + \epsilon$. By replacing g with $ag + b$, we may assume $g(x) = 0$ and $\sup |g| \leq 1$.

Pick a neighborhood U of x on which $|g| < \epsilon$. Since A separates points, for each $y \notin U$ there is a function $h \in A$ with $h(x) = 0$, $h(y) = 2$. Taking a finite subcover of $X - U$ by balls on which $h > 1$, we obtain a function $f = h_1^2 + \cdots + h_n^2 + \epsilon$ with $f(x) = \epsilon = g(x) + \epsilon$, with $f \geq \epsilon > g$ on U, and with $f \geq 1 > g$ on $X - U$. Then $f \in \mathcal{F}$, and so $g(x) = \inf_{\mathcal{F}} f(x)$ as desired. ∎

Paracompactness. For local constructions like making a metric, what's needed is not so much compactness (finiteness of coverings) as paracompactness (local finiteness). This says that any open covering has a locally finite *refinement*. Using this property one can show, for example, that any *paracompact manifold* admits a metric.

All metric spaces are paracompact (a hard theorem). However there exists a manifold which is *not* paracompact, namely the *long line*. It is obtained from the first uncountable ordinal Ω by inserting an interval between any two adjacent points, and introducing the order topology.

This space X has the amazing property that **every sequence has a convergent subsequence.** Indeed, since a sequence is countable, it is bounded above by some countable ordinal α, and (by induction) the segment $[0, \alpha]$ is homeomorphic to $[0, 1]$, hence compact.

On the other hand, X is not compact, since the open covering by all intervals of the form $[0, \alpha)$ has no finite subcover. Thus X is not metrizable. Therefore X is not paracompact.

6 Banach Spaces

The theory of Banach spaces is a combination of infinite-dimensional linear algebra and general topology. The main themes are *duality*, *convexity* and *completeness*.

The first two themes lead into the Hahn-Banach theorem, separation theorems for convex sets, weak topologies, Alaoglu's theorem, and the Krein-Milman theorem on extreme points. The third theme leads to the '3 principles of functional analysis', namely the open mapping theorem, the closed graph theorem and the uniform boundedness principle. These three results all rest on the Baire category theorem and hence make crucial use of completeness.

Continuous linear maps. Let $\phi : X \to Y$ be a linear map between Banach spaces. The norm of ϕ, denoted $\|\phi\|$, is defined as the least M such that

$$\|\phi(x)\| \leq M\|x\|$$

for all $x \in X$. Note: if $Y = \mathbb{R}$ we use the usual absolute value on $\mathbb{R}$ as a norm.

A linear map is *bounded* if its norm is finite.

Theorem 6.1 *A linear map is bounded iff it is continuous.*

Proof. Clearly boundedness implies (Lipschitz) continuity. Conversely, if ϕ is continuous, then $\phi^{-1}B(0,1)$ contains $B(0,r)$ for some $r>0$ and then $\|\phi\| \leq 1/r$. ∎

Theorem 6.2 (Hahn-Banach) *Let $\phi : S \to \mathbb{R}$ be a linear map defined on a subspace $S \subset X$ in a Banach space such that $|\phi(x)| \leq M\|x\|$ for all $x \in S$. Then ϕ can be extended to a linear map on all of X with the same inequality holding.*

Proof. Using Zorn's lemma, we just need to show that any maximal such extension of ϕ is defined on all of X. So it suffices to consider the case $S \neq X$ and show that ϕ can be extended to the span of S and y where $y \in X - S$.

We may assume $M = 1$. The extension will be determined by its value $\phi(y) = z$, and the extension will continue to be bounded by $M = 1$ so long as we can insure that z is chosen so for all $s \in S$ we have:

$$-\|y+s\| \leq \phi(y+s) = z + \phi(s) \leq \|y+s\|.$$

To show such a z exists amounts to showing that for any $s, s' \in S$ we have

$$-\phi(s) - \|y+s\| \leq -\phi(s') + \|y+s'\|,$$

so that there is a number z between the sup and inf. Now this inequality is equivalent to:

$$\phi(s') - \phi(s) \leq \|y+s\| + \|y+s'\|,$$

and this one is in fact true, since

$$\phi(s'-s) \leq \|s'+y-y-s\| \leq \|s+y\| + \|s'+y\|.$$

∎

Linear functionals on $L^\infty[0,1]$. We can now show more rigorously that $L^1[0,1]$ is not reflexive: namely take point evaluation on $C[0,1]$, and extend it by Hahn-Banach to a linear functional ϕ on $L^\infty[0,1]$. It is clear then $\phi|C[0,1]$ is not given by an element in $L^1[0,1]$.

Embedding into X^{**}.

Theorem 6.3 *For any $x \in X$ there is a $\phi \in X^*$ such that $\|\phi\| = 1$ and $\phi(x) = \|x\|$.*

Proof. Define ϕ first on the line through x, then extend it by Hahn-Banach.

Corollary 6.4 *The embedding of X into X^{**} is isometric.*

L^p examples. Let $f \in L^p(\mathbb{R})$ with $\|f\|_p = 1$ and $1 < p < \infty$, there is a *unique* ϕ of norm 1 in the dual space such that $\phi(f) = 1$: namely $\phi = \operatorname{sign}(f)|f|^{p/q}$, which satisfies

$$\phi(f) = \int f\phi = \int |f|^p = 1.$$

This reflects the 'smoothness' of the unit ball in L^p: there is a unique supporting hyperplane at each point.

For L^1 things are different: for example, if $\operatorname{supp} f = [0,1]$ there is a huge space of $\phi \in L^\infty$ such that $\|\phi\|_\infty = 1$ and $\phi(f) = 1$.

Non-example: L^∞. Now let $f(x) = x$ in $L^\infty[0,1]$, and suppose $\phi \in L^1[0,1]$ has norm 1. Choose $a < 1$ such that $t = \int_0^a |\phi| > 0$. Then we have:

$$\phi(f) \le \int_0^a x|\phi| + \int_a^1 |\phi| \le at + (1-t) = 1 - (1-a)t < 1.$$

Thus $\phi(f)$ can never be 1! This reflects of course the fact that $X = L^1[0,1]$ is a proper subset of its double dual $X^{**} = L^\infty[0,1]^*$.

More non-reflexive spaces. For the little ℓ^p spaces we have the following, rather rich non-reflexive example:

$$c_0^* = \ell^1, (\ell_1)^* = \ell^\infty, (\ell^\infty)^* = m(\mathbb{Z}).$$

It turns out the last space can be identified with the space of finitely-additive measures on $\mathbb{Z}$.

Weak closure. The Hahn-Banach theorem implies:

Theorem 6.5 *Let $S \subset X$ be a linear subspace of a Banach space. Then S is weakly closed iff S is norm-closed.*

Proof. Any weakly closed space is norm closed. Conversely, if S is norm closed, for any $y \notin S$ we can find a linear functional $\phi : X \to \mathbb{R}$ that vanishes on S and sends y to 1, so y is not in the weak closure of S. ∎

(More generally, as we will see later, any norm-closed convex set is weakly closed.)

The weak* topology. We say $\phi_\alpha \to \phi$ in the weak* topology on X^* if $\phi_\alpha(x) \to \phi(x)$ for every $x \in X$.

Example: weak closures of continuous functions. The space $C[0,1]$ is dense in $L^\infty[0,1]$ in the weak* topology. Indeed, if $g \in L^\infty$ then there are continuous $f_n \to g$ pointwise a.e. with $\|f_n\|_\infty \le \|g\|_\infty$. Now for any $h \in L^1[0,1]$ the dominated convergence theorem implies

$$\langle h, f_n\rangle = \int h f_n \to \int hg = \langle h, g\rangle.$$

On the other hand, $C[0,1]$ is already *closed* in the weak topology, since it is norm closed.

Theorem 6.6 (Alaoglu) *The unit ball $B^* \subset X^*$ is compact in the weak* topology.*

Proof. Let B be the unit ball in X. Then there is a tautological embedding of B^* into $[-1,1]^B$. Since linearity and boundedness are preserved under pointwise limits, the image is closed. By Tychonoff, it is compact! ∎

Metrizability. Theorem. If X is separable, then the unit ball B in X^* is a compact metrizable space in the weak* topology.

Proof. Let x_n be a dense sequence in X; then the balls

$$B = \{\phi \,:\, |\phi(x_n) - p/q| < 1/r\}$$

form a countable base. By Urysohn's metrization theorem, B is metrizable.

Example: the space of measures. Naturally $C[0,1]$ is separable. Thus $P[0,1]$, the space of probability measures with the weak* topology, is a compact metric space. It can be thought of as a sort of infinite-dimensional simplex; indeed the measures supported on $\le n$ points form an $(n+1)$-simplex.

Banach limits.

Theorem 6.7 *There is a linear map* $\mathrm{Lim} : \ell^\infty(\mathbb{N}) \to \mathbb{R}$ *such that*

$\mathrm{Lim}(a_n) \ge 0$ *if* $a_n \ge 0$
$\mathrm{Lim}(1) = 1$*, and*
$\mathrm{Lim}(a_{n+1}) = \mathrm{Lim}(a_n)$.

Note that $|\operatorname{Lim}(a_n)| \leq \|a_n\|$ and that Lim extends the usual limit on c and agrees with the Césaro limit when that exists.

Proof. Let $\phi_N(a) = N^{-1}\sum_1^N a_n$ and let Lim be the limit point of a convergent subnet. Note that $\phi_N(a_{n+1}) - \phi_N(a_n) = O(1/N)$. ■

Stone-Čech compactification of $\mathbb{N}$. The unit ball B in $\ell^\infty(\mathbb{N})^*$, while compact, is not metrizable! Indeed, the integers embed via $\phi_n(a) = z_n$, but $\langle \phi_n \rangle$ has no convergent subsequence! (If ϕ_{n_k} is a subsequence, then we can choose $a \in \ell^\infty$ such that $a_{n_k} = (-1)^k$; then $\phi_{n_k}(a)$ does not converge, so ϕ_{n_k} does not converge in the weak* topology.

The Banach-Tarski paradox . Using the same construction on $\mathbb{Z}$ or $\mathbb{Z}^n$, we get finitely-additive measures by applying Lim to indicator functions. Because of these measures, you cannot cut $\mathbb{Z}$ into a finite number of sets, move them by translation and re-assemble them to form 2 copies of $\mathbb{Z}$.

However, this type of re-construction *is* possible for a free group G on 2 generators!

Suppose μ is a finitely-additive invariant probability measure on G. Let W_a, $W_{a'}$, W_b and $W_{b'}$ denote the partition of $G - \{e\}$ into reduced words beginning with a, a', b and b'. Then $a'W_a$ contains W_a, W_b, $W_{b'}$ and $\{e\}$. Since translation by a' preserves measures, we conclude that the extra sets W_b, $W_{b'}$ and $\{e\}$ have measure zero. By the same token, all the W's have measure zero, which contradicts the assumption that $\mu(G) = 1$.

Cutting up the sun. Note that $G = a'W_a \cup W_{a'}$, and similar for W_b and $W_{b'}$. Thus we can cut G into 5 pieces, discard one of them (e), and re-assemble the other two into two copies of G.

Now embed G into $SO(3)$ by taking two random rotations. Then G acts on S^2. Let $E \subset S^2$ be a transversal, consisting of one point from each G-orbit, so $S^2 = G \cdot E$. Now cut S^2 into pieces of the form $E_i = W_i \cdot E$, $i = 1, \ldots, 4$. (There will be some S^2 left over.) Applying the left action of G to these pieces — that is, applying rotations — we can re-arrange W_1 and W_2 to form G, and so re-arrange E_1 and E_2 to form S^2. Do the same thing with E_3 and E_4, and we can make a second sphere!

Three basic principles of functional analysis. Let $A : X \to Y$ be a linear map between Banach spaces X and Y. Then we have:

1. *The open mapping theorem.* If A is continuous and onto, then it is open; that is, $Ax = y$ has a solution with $\|x\| \leq C\|y\|$.

Corollaries: If A is continuous and bijective, then it is an isomorphism. If X is complete with respect to two norms, and $\|x\|_1 \leq C\|x\|_2$, then a reverse inequality holds.

2. *The closed graph theorem.* If the graph of A is closed — meaning $x_n \to x$, $Ax_n \to y$ implies $Ax = y$ — then A is continuous.

3. *The uniform boundedness principle.* Let $\mathcal{F} \subset X^*$ satisfy that for each $x \in X$, $|f(x)| \leq M_x\|x\|$ for all $f \in \mathcal{F}$. Then there is an M such that $\|f\| \leq M$ for all $f \in \mathcal{F}$.

   The same result holds if we replace X^* with $B(X, Y)$.

These principles should be compared to the following results that hold when X and Y are *compact*.

1. If $f : X \to Y$ is bijective and continuous, then f is a homeomorphism.

2. If $f : X \to Y$ has a closed graph, then f is continuous.

   (Note that $f(x) = 1/x$ for $x \neq 0$, $f(0) = 0$, gives a map $f : \mathbb{R} \to \mathbb{R}$ with a closed graph that is not continuous.)

3. Let $\mathcal{F} \subset C(X)$ satisfy $|f(x)| \leq M_x$ for all $f \in \mathcal{F}$ and $x \in X$. Then there is a nonempty open set $U \subset X$ and a constant $M > 0$ such that $|f|U| \leq M$ for all $x \in U$.

Open mapping theorem: proof. Let $D = B(0, 1)$ be the open unit ball about the origin in X. We must show $B = A(D)$ contains a neighborhood of the origin in Y. By surjectivity of A, we have $Y = \bigcup nB$, and thus by the Baire category theorem, $n\overline{B}$ has nonempty interior for some n; and thus $\overline{B}$ has non-empty interior U.

Since $\overline{B}$ is convex and symmetric, we have $(U - U)/2 \subset \overline{B}$ and so $\overline{B}$ contains a neighborhood of the origin, say $B(0, r)$.

We now proceed to solve the equation $Ax = y$. Set $M = 1/r$. Then there is an x_1 with $\|x_1\| \leq M\|y\|$ and $\|Ax_1 - y\|$ as small as we like; say less than $\|y\|/2$. Solving for the difference, we obtain an x_2 with $\|x_2\| \leq M/2\|y\|$ and $\|A(x_1+x_2)-y\| \leq \|y\|/4$. Proceeding by induction we obtain a geometrically convergent sequence, and by continuity of A we have $Ax = y$ where $x = \sum x_i$ satisfies $\|x\| \leq 2M$. Thus $A(D)$ contains a ball of radius at least $1/(2r)$ about the origin. ∎

Open-mapping theorem: application. The open mapping theorem implies:

Corollary 6.8 *If X is complete in two norms, and $\|x\|_1 \leq C\|x\|_2$, then there is a C' such that $\|x\|_2 \leq C'\|x\|_1$.*

Here is a nice application due to Grothendieck.

Theorem 6.9 *Let $S \subset L^2[0,1]$ be a closed subspace such that every $f \in S$ is continuous. Then S is finite-dimensional.*

Proof. We have $\|f\|_\infty \geq \|f\|_2$, so S is complete in both the L^2 and the L^∞ norms. Thus there is an $M > 0$ such that $M\|f\|_2 \geq \|f\|_\infty$.

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

$$\|\sum f_i(p) f_i\|_2 = (\sum |f_i(p)|^2)^{1/2},$$

and thus

$$M(\sum |f_i(p)|^2)^{1/2} \geq \|\sum f_i(p) f_i\|_\infty \geq \sum f_i(p)^2,$$

which implies $\sum f_i(p)^2 \leq M^2$. Integrating from 0 to 1 gives $n \leq M^2$. ■

Closed graph theorem: proof. Let $|x| = \|x\| + \|Ax\|$. Now if $|x_n|$ is Cauchy, then $x_n \to x$ in X and $Ax_n \to y$ in Y; since the graph of A is closed, we have $Ax = y$ and thus $|x_n - x| \to 0$. Thus X is complete in the $|\cdot|$ norm, so by the open mapping theorem we have $|x| \leq M\|x\|$ for some M; thus $\|A\| \leq M$ and A is continuous. ■

Uniform boundedness theorem: proof. Let $F_M = \{x : |f(x)| \leq M \,\forall f \in \mathcal{F}\}$. By Baire category, some F_M contains a ball $B(p,r)$. Then for $\|x\| \leq r$ we have $|f(x)| = |f(p+x) - f(p)| \leq M + M_p$ and thus $\|f\|$ is uniformly bounded by $(M + M_p)/r$. ■

Example. Let $\phi_n \in X^*$ have the property that $\psi(x) = \lim \phi_n(x)$ exists for every $x \in X$. Then $\|\phi_n\| \leq M$ and hence $\psi \in X^*$.

Corollary. You cannot construct an unbounded linear functional by taking a pointwise limit of bounded ones.

Remark: if a net satisfies $x_\alpha \to y$, is $\|x_\alpha\|$ necessarily bounded? No! Let α range over all finite subsets of $\mathbb{N}$, directed by inclusion, and let x_α be the minimum of α. Then $x_\alpha \to 0$ but $\sup x_\alpha = \infty$.

Theorem 6.10 (Toeplitz) *Let $T : H \to H$ be a symmetric linear operator on Hilbert space, meaning $(Tx, y) = (x, Ty)$. Then T is continuous.*

Proof. Suppose $x_n \to x$ and $Tx_n \to z$. Then for all $y \in H$, we have

$$(y, z) = \lim(y, Tx_n) = \lim(Ty, x_n) = (Ty, x) = (y, Tx).$$

Thus $(y, Tx - z) = 0$ for all $y \in H$. Taking $y = Tx - z$, we find $Tx = z$. Thus T has a closed graph, and hence T is continuous. ∎

Note: a typical symmetric operator is given by $(Tf)(x) = \int K(x, y)f(y)\,dy$, where the kernel $K(x, y)$ is symmetric.

Convexity. A subset $K \subset X$ is *convex* if $x, y \in K \implies tx + (1 - t)y \in K$ for all $t \in (0, 1)$.

Support.

Theorem 6.11 *Let $K \subset X$ be an open convex set not containing the origin. Then there is a $\phi \in X^*$ such that $\phi(K) > 0$.*

Proof. Geometrically, we need to find a closed, codimension-one hyperplane $H = \operatorname{Ker}\phi \subset X$ disjoint from K. Consider all subspaces disjoint from K and let H be a maximal one (which exists by Zorn's lemma). If H does not have codimension one, then we can consider a subspace $S \supset H$ of two dimensions higher and all extensions $H' = H + \mathbb{R}v_\theta$ of H to S, $\theta \in S^1$.

Now consider the set $A \subset S^1$ of θ such that $H + \mathbb{R}_+ v_\theta$ meets K. Then A is open, connected, and $A \cap A + \pi = \emptyset$; else K would meet H. It follows that A is an open interval of length at most π. Taking an endpoint of A, we obtain an extension of H to H', a contradiction.

Thus H has codimension one. Since K is open, $\overline{H}$ is also disjoint from K, and hence $H = \overline{H}$. Thus H is the kernel of the desired linear functional. ∎

Separation.

Theorem 6.12 *Let $K, L \subset X$ be disjoint convex sets, with K open. Then there is a $\phi \in X^*$ separating K from L; i.e. $\phi(K)$ and $\phi(L)$ are disjoint.*

Proof. Let $M = K - L$; this set is open, convex, and it does not contain the origin because K and L are disjoint. Thus by the support theorem, there is a linear functional with $\phi(M) \geq 0$. Then for all $k \in K$ and $\ell \in L$, we have $\phi(k - \ell) = \phi(k) - \phi(\ell) \geq 0$. It follows that $\inf \phi(K) \geq \sup \phi(L)$. Since $\phi(K)$ is open, these sets are disjoint. ∎

Weak closure. Recall that $K \subset X$ is **weakly closed** if whenever a net $x_\alpha \in K$ satisfies $\phi(x_\alpha) \to \phi(x)$ for all $\phi \in X^*$, we have $x \in K$. The weak closure of a set is generally larger than the strong (or norm) closure. For example, the sequence $f_n(x) = \sin(nx)$ in $L^1[0,1]$ is closed in the norm topology (it is discrete), but its weak closure adds $f_0 = 0$.

Another good image to keep in mind is that K is weakly closed if for any $x \notin K$, there is a continuous linear map $\Phi : X \to \mathbb{R}^n$ such that $\overline{\Phi(K)}$ is disjoint from $\Phi(x)$. This is just because a base for the weak topology consists of finite intersections of sets of the form $\phi^{-1}(\alpha, \beta)$.

Theorem 6.13 *A convex set $K \subset X$ is weakly closed iff K is strongly (norm) closed.*

Proof. A weakly closed set is automatically strongly closed. Now suppose K is strongly closed, and $x \notin K$. Then there is an open ball B containing x and disjoint from K. By the separation theorem, there is a $\phi \in X^*$ such that $\phi(x) > \phi(K)$, and thus x is not in the weak closure of K. Thus K is weakly closed.

Linear combinations. By the preceding result, we see that the weak closure of a set $E \subset X$ is contained in hull(E), the smallest norm-closed convex set containing E. Now hull(E) can be described as the closure of finite convex combinations of points in E. So as an example we have:

Proposition. For any $\epsilon > 0$ there exist constants $a_n \geq 0$, $\sum a_n = 1$, such that

$$\left\| \sum_1^N a_n \sin(nx) \right\|_1 < \epsilon.$$

Problem. Prove this directly!

(Solution. Just take $a_n = 1/N$ for $n = 1, \ldots, N$, and note that for orthgonal functions e_n the function $f = \sum a_n e_n$ satisfies

$$\|f\|_1^2 \leq \|f\_2^2\|1\|_2^2 = O(\sum |a_n|^2) = O(1/N).$$

Intuitively, $f(x)$ behaves like a random walk with N steps.

LCTVS. A *topological vector space* X is a vector space with a topology such that addition and scalar multiplication are continuous. By translation invariance, to specify the topology on X it suffices to give a basis at the origin.

A very useful construction comes from continuity of addition: for any open neighborhood U of the origin, there is a neighborhood V such that $V + V \subset U$.

Usually we assume X is Hausdorff (T_2). This is equivalent to assuming points are closed (T_1). Indeed, if points are closed and $x \neq y$, then we can find a balanced open neighborhood U of the origin such that $y+U$ is disjoint from X. We can then find a balanced open V such that $V+V \subset U$, and then $x+V$ is disjoint from $y+V$.

Warning: Royden at times implicitly assumes X is Hausdorff. For example, if X is not Hausdorff, then an extreme point is not a supporting set, contrary to the implicit assumption in the proof of the Krein-Milman theorem.

Let X be a Banach space. Then the weak topology on X and the weak* topology on X^* are Hausdorff and locally convex. All the results like the Hahn-Banach theory, the separation theorem, etc. hold for locally convex topologies as well as the norm topology and weak topology.

Extreme points. Let K be convex. A point $x \in K$ is an *extreme point* if there is no open interval in K containing X. More generally, a *supporting set* $S \subset K$ is a closed, convex set with the property that, whenever an open interval $I \subset K$ meets S, then $I \subset S$. One should imagine a face of ∂K or a subset thereof.

Example: Let K be a convex compact set. Then the set of points where $\phi \in X^*$ assumes its maximum on $K \subset X$ is a supporting set. In particular, any compact convex set has nontrivial supporting sets.

Theorem 6.14 (Krein-Milman) *Let K be a compact convex set in a locally convex (Hausdorff) topological vector space X. Then K is the closed convex hull of its extreme points.*

Remark. The existence of *any* extreme points is already a nontrivial assertion.

Proof. We will show any supporting set contains an extreme point. Indeed, consider any minimal nonempty supporting set $S \subset K$; these exist by Zorn's lemma, using compactness to guarantee that the intersection of a nested family of nonempty supporting sets is nonempty. Now if S contains two distinct points x and y, we can find a $\phi \in X^*$ (continuous in the given topology) such that $\phi(x) \neq \phi(y)$. Then the set of points in S where ϕ assumes its maximum is nonempty (by compactness) and again a supporting set, contrary to minimality.

Now let $L \subset K$ be the closed convex hull of the extreme points. If there is a point $x \in K - L$, then we can separate x from L by a linear functional, say $\phi(x) > \phi(L)$. But then the set of points where ϕ assumes its maximum is a supporting set, and therefore it contains an extreme point, contrary to the assumption that ϕ does not assume its maximum on L.

Therefore $L = K$. ∎

Prime example: The unit ball in X^*, in the weak* topology.

What are the extreme points of the unit ball B in $L^2[0,1]$? Every point in ∂B is extreme! Because if $\|f\|_2 = 1$ then for ϵ and $g \neq 0$, we have

$$\|f \pm \epsilon g\|_2^2 = \|f\|^2 \pm 2\epsilon\langle f, g\rangle + \epsilon^2\|g|^2$$

and this is $> \|f\|^2 = 1$ if the sign is chosen properly.

What about the unit ball in $L^\infty[0,1]$? Here the extreme points are functions with $|f| = 1$ a.e. Picture the finite-dimensional case — a cube.

What about in $L^1[0,1]$? Here there are *no* extreme points! For example, if $f = 1$, then $f(x) + a\sin(2\pi x)$ has norm one for all small a, so f is not extreme. Similarly, for any $f \neq 0$ we can find a set A of positive measure on which $f > a > 0$ (or $0 > a > f$), and then a function g of mean zero supported on A such that $\|f \pm g\| = \|f\|$.

This fact is compatible with Krein-Milman only because $L^1[0,1]$ is *not* a dual space. In fact the preceding remark *proves* that for any Banach space X, the dual X^* is not isomorphic to $L^1[0,1]$.

For $X = C[0,1]$, the dual X^* consists of signed measures of total mass one, and the extreme points are $\pm\delta_x$.

Stone-Weierstrass revisited. Let X be a compact Hausdorff space, and let $A \subset C(X)$ be an algebra of real-valued functions containing the constants and separating points. Then A is dense in $C(X)$.

Proof (de Brange). Let $A^\perp \subset M(X)$ be the set of measures that annihilate A. By the Hahn-Banach theorem, to show A is dense it suffices to show that $A^\perp$ is trivial.

Let K be the intersection of $A^\perp$ with the unit ball. Then K is a closed, compact, convex set in the weak* topology. Thus K is the closed convex hull of its extreme points.

Suppose $\mu \in K$ is a nonzero extreme point. We will deduce a contradiction.

First, let $E \subset X$ be the support of μ (the smallest closed set whose complement has measure zero). Suppose E is not a single point. Choose a function $f \in A$ such that $f|E$ is not constant, and $|f| < 1$. Consider the two measures

$$\sigma = (1 + f)\mu/2, \quad \tau = (1 - f)\mu/2.$$

Since A is an algebra, both σ and τ are in $A^\perp$, and of course we have $\sigma + \tau = \mu$. Moreover, since $1 \pm f > 0$, we have

$$\|\mu| = \|\sigma\| + \|\tau| = 1.$$

Thus μ is a convex combination:

$$\mu = \|\sigma\| \frac{\sigma}{\|\sigma\|} + \|\tau\| \frac{\tau}{\|\tau\|}.$$

Since μ is an extreme point, it follows that $\mu = \sigma = \tau$. Therefore f is constant a.e. on E, a contradiction.

It follows that μ is a delta-mass supported on a single point. But the μ is not in $A^\perp$, since it pairs nontrivially with the constant function in A. ■

Haar measure. As a further application of convexity, we now develop the Kakutani fixed-point theorem and use it to prove the existence of Haar measure on a compact group. Our treatment follows Rudin, *Functional Analysis*, Chapter 5.

Theorem 6.15 (Milman) *Let $K \subset X$ be a compact subset of a Banach space and suppose $H = \overline{\mathrm{hull}(K)}$ is compact. Then the extreme points of H are contained in K.*

Proof. Suppose x is an extreme point of H that does not lie in K, and let $r = d(x, K)$. Then by compactness we can cover K by a finite collection of balls $B(x_i, r)$, $i = 1, \ldots, n$. Let H_i be the closed convex hull of $K \cap B(x_i, r)$. Since the ball is compact, we have $H_i \subset B(x_i, r)$.

Now $H = \mathrm{hull}(\bigcup_1^n H_i)$, and thus $x = \sum t_i h_i$ is a convex combination of points $h_i \in H_i \subset H$. But x is an extreme point, so $x = h_i$ for some i. This implies $x \in B(x_i, r)$, contradicting the fact that $d(x, K) = 2r$. ■

Theorem 6.16 (Kakutani) *Let $K \subset X$ be a nonempty compact convex subset of a Banach space, and let G be a group of isometries of X leaving K invariant. Then there exists an $x \in K$ fixed by all $g \in G$.*

Proof. Let $L \subset K$ be a minimal, nonempty, compact convex G-invariant set; such a set exists by the Axiom of Choice. If L consists of a single point, we are done. Otherwise there are points $x \neq y$ in L. Let $z = (x+y)/2$. Then by minimality of L, we have $L = \overline{\mathrm{hull}(G \cdot z)}$. Let z' be an extreme point of L. By Milman's theorem, z' is a limit of points in $G \cdot z$. By compactness of K, we can choose $g_n \in G$ such that $g_n z \to z'$, $g_n x \to x'$ and $g_n y \to y'$. But then $z' = (x' + y')/2$, so z' is not an extreme point. ■

Theorem 6.17 *Let G be a compact Hausdorff group. Then there is a unique left-invariant Borel probability measure μ on G, and μ is also right invariant.*

Proof. Let G be a compact topological group, and let $X = C(G)$. For each $g, h \in G$, the shift operators $L_g(f) = f(g^{-1}x)$ and $R_h(f) = f(xh)$ are isometries, and they commute. The only fixed-points for G are the constant functions.

Now fix $f \in C(G)$. Then f, and all its translates, are equicontinuous, and thus

$$L(f) = \overline{\text{hull}(G \cdot f)} \subset C(G)$$

is compact. Similarly, the closed convex hull of the right translates, $R(f) = \overline{\text{hull}(f \cdot G)}$, is also compact. By Kakutani's fixed-point theorem, each of these convex sets contains at least one constant function, $l(f)$ and $r(f)$.

The constant $l(f)$ can be approximated by averages of the form

$$T(f) = \sum \alpha_i L_{a_i}(f),$$

and similarly for $r(f)$. But the right and left averages commute, and leave the constants invariant, so $l(f) = r(f)$. Thus there is a unique constant function, $M(f)$, contained in both $L(f)$ and $R(f)$.

To show $M(f)$ corresponds to Haar measure, we must show $M(1) = 1$, $M(f) \geq 0$ when $f \geq 0$, and M is linear. The first two assertions are immediate. To show $M(f + h) = M(f) + M(h)$, choose a left-averaging operator T such that $M(f) \approx T(f)$. Then $T(h) \in L(h)$, so $M(T(h)) = M(h)$. Thus there is a second left-averaging operator S such that $S(T(h)) \approx M(h)$. But then $S(T(f + h)) \approx M(f) + M(h) \in L(f + h)$, so $M(f + h) = M(f) + M(h)$. ∎

Examples of compact groups: Finite groups, products such as $(\mathbb{Z}/2)^{\mathbb{N}}$, inverse limits such as $\mathbb{Z}_p = \varprojlim (\mathbb{Z}/p^n)$; Lie groups such as $\mathrm{SO}(n, \mathbb{R})$ and $\mathrm{SU}(n, \mathbb{C})$; p-adic Lie groups such as $\mathrm{SL}_n(\mathbb{Z}_p)$.

Here is a description of Haar measure on $G = \mathrm{SO}(n, \mathbb{R})$. Consider the Lie algebra $g = \mathrm{so}(n, \mathbb{R})$; it is the space of trace-zero matrices satisfying $A^t = -A$. There is a natural inner product on this space, given by $\langle A, B \rangle = \mathrm{tr}(AB)$. This inner product is invariant under the adjoint action of G, so it gives rise to an invariant quadratic form on every tangent space $T_g G$. In the case of $\mathrm{SO}(n)$, this metric is negative definite. Thus its negative determines a bi-invariant **metric** on $\mathrm{SO}(n, \mathbb{R})$, and hence an invariant measure.

This measure can be described as follows: to choose a random frame in $\mathbb{R}^n$, one first pick a point at random on S^{n-1}, then a point at random on the orthogonal S^{n-2}, etc., using the rotation-invariant probability measures on each sphere. There is a unique choice for the final point on S^0 that makes the frame positively oriented.

Unimodularity. More generally, any locally compact group G carries right and left invariant measures, unique up to scale, but they need not agree. When they do, the group is *unimodular.* For example, the group $\mathrm{SL}_2(\mathbb{R})$ is unimodular, but its upper-triangular subgroup AN is not.

7 Hilbert space

Of great importance in analysis are the Hilbert spaces, such as $L^2(\mathbb{R}^n)$.

Abstractly, a Hilbert space is a Banach space H equipped with a symmetric bilinear form (x,y) such that $(x,x)=\|x\|^2$. Examples:

$\mathbb{R}^n$, $(x,y)=\sum x_i y_i$.
$\ell^2(\mathbb{N})$, $(x,y)=\sum x_i y_i$.
$L^2(\mathbb{R}^n)$, $(f,g)=\int f(x)g(x)\,dx$.

Theorem 7.1 (Bunyiakowsky-Cauchy-Schwarz) $|(x,y)|\le\|x\|\cdot\|y\|$.

Proof. For all t we have $0\le(x+ty,x+ty)=(x,x)+t^2(y,y)+2t(x,y)$, so the discriminant of this quadratic polynomial must be non-positive. Thus $0\ge b^2-4ac\ge 4(x,y)^2-4(x,x)(y,y)$. ∎

An *orthonormal set* is a collection of unit vectors x_i in H, $i\in I$, with $(x_i,x_j)=\delta_{ij}$. The index set I can be finite, countable or even larger.

Given an orthonormal set, we define the 'Fourier coefficients' of $x\in H$ by $a_i=(x,x_i)$.

Lemma (Bessel). $\sum|a_i|^2\le\|x\|^2$.

Proof. For any finite sum we have

$$0\le(x-\sum a_i x_i, x-\sum a_i x_i)=(x,x)-2\sum|a_i|^2+\sum|a_i|^2.$$

∎

A *basis* for H is a maximal orthonormal set (x_i). By Zorn's Lemma, every Hilbert space has a basis. The elements of a basis are at distance $\sqrt{2}$ from one another. Thus if H is separable, it has a countable basis.

Given a basis, we can use Bessel's inequality to show $\sum a_i x_i$ converges for any $(a_i)\in\ell^2(I)$. Moreover, the norm of the sum in H coincides with the norm in $\ell^2(I)$. Finally, if $x \in H$ has Fourier coefficients a_i, then $y = x - \sum a_i x_i$ has Fourier coefficients zero, i.e. it is orthogonal to all x_i. Since the (x_i) are a maximal orthonormal set, $y = 0$. This shows:

Theorem 7.2 *For any orthogonal basis, and any $x \in H$, we have $x = \sum a_i x_i$ in H.*

Theorem 7.3 *For any basis $(x_i, i \in I)$ of a Hilbert space H, there is a natural isomorphism between H and the Hilbert space $\ell^2(I)$.*

Examples: On $S^1 = \mathbb{R}/\mathbb{Z}$, we can take 1, $\cos(2\pi n x)$ and $\sin(2\pi n x)$ as an orthonormal basis. Completeness follows from Stone-Weierstrass.

On $[-1, 1]$ we can apply Gram-Schmidt to the polynomials to obtain an orthonormal basis of Legendre polynomials $p_n(x)$. of degree n. Again Stone-Weierstrass yields completeness.

Complex Hilbert spaces. Over the field $\mathbb{C}$, the natural form for a Hilbert space is a Banach space H with a *Hermitian form* $\langle x, y\rangle$ such that $\langle x, x\rangle = \|x\|^2$. In this case, $(x, y) = \mathrm{Re}\langle x, y\rangle$ makes H into a Hilbert space over $\mathbb{R}$. Examples:

$\langle x, y\rangle = \sum x_i \overline{y}_i$ on $\mathbb{C}^n$ or $\ell^2(\mathbb{N}) \otimes \mathbb{C}$.
$\langle f, g\rangle = \int f\overline{g}$ on $L^2(\mathbb{R}^n)$.

The Hardy space. In $L^2(S^1) \otimes \mathbb{C}$, a natural orthonormal basis is given by $f_n(z) = z^n/2\pi$. The span of $f_n, n \geq 0$ is a closed subspace $H^2(S^1)$ known as the *Hardy space* of the circle. Every $f \in H^2(S^1)$ is the boundary value of a holomorphic function on S^1.

Fourier series. For a function $f(z)$ in $L^2(S^1)$, we have $f = \sum a_n z^n$ (norm convergent in L^2) where

$$a_n = \frac{1}{2\pi i} \int_{S^1} f(z) z^{-n} \frac{dz}{z}.$$

Note the analogy with Laurent series.

Passing to the coordinate x where $z = \exp(ix)$, we can think of $L^2(S^1)$ as the subspace of $L^2(\mathbb{R})$ consisting of functions that are periodic under $x \mapsto x + 2\pi$. Since $z^n = \cos(nx) + i\sin(nx)$, the Fourier series now becomes a sum of sines and cosines.

If we restrict to *odd* functions — where $f(-x) = -f(x)$ — then only sine terms appear, and we can identify this subspace with $L^2[0, \pi]$. Thus a function on $[0, \pi]$ has a natural Fourier series:

$$f(x) = \sum_{n=1}^{\infty} a_n \sin(nx).$$

Since $\int_0^\pi \sin^2(x)\, dx = \pi/2$ (its average value is $1/2$), we have

$$a_n = \frac{2}{\pi} \int_0^\pi f(x) \sin(nx)\, dx.$$

Example: if the graph of f is a triangle with vertex (p, x), then

$$a_n = \frac{2h \sin(np)}{n^2 p(\pi - p)}.$$

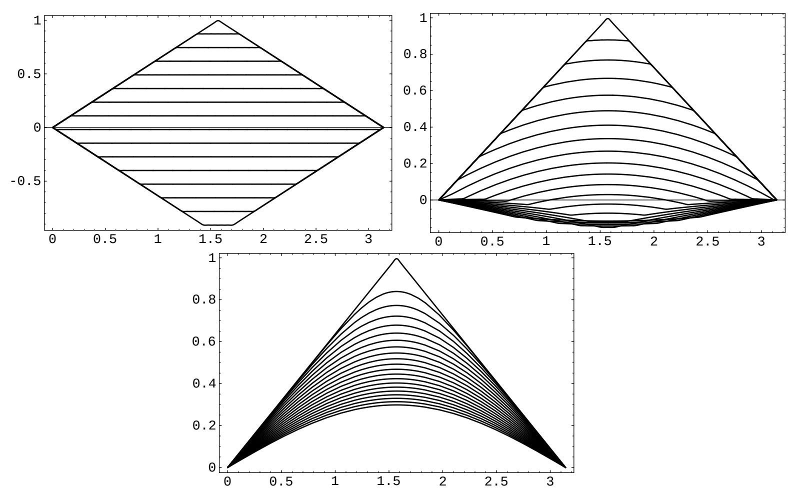

Figure 2. Solutions to the wave equation (undamped and damped) and the heat equation.

The wave equation and the heat equation. A typical problem in PDE is to solve the *wave equation* with given initial data $f(x) = u(x, 0)$ on $[0, \pi]$. This equation, which governs the motion $u(x, t)$ of a vibrating string, is given by

$$u_{tt} = u_{xx}$$

(where the subscripts denote differentiation). If we think of $u(x,t)$ as the motion of a string with fixed end points, it is natural to impose the boundary conditions $u(0,t) = u(\pi,t) = 0$. We will also assume $u_t(x,0) = 0$, i.e. the string is intially stationary.

Since the wave equation is linear, it suffices to solve it for the Fourier basis functions $f(x) = \sin(nx)$. And for these we have simply

$$u(x,t) = \cos(nt)\sin(nx).$$

This solution can be discovered by separation of variables; the key is that $f(x)g(t)$ solves the wave equation if f and g are eigenfunctions with the same eigenvalues.

These basic solutions are 'standing waves' corresponding to the bass note and then the higher harmonics of the string.

The solution to the wave equation for 'general' $f(x)$ is then given by:

$$u(x,t) = \sum_{n=1}^{\infty} a_n \cos(nt)\sin(nx).$$

Note that $u(x, t + 2\pi) = u(x,t)$, i.e. the string has a natural frequency.

The *heat equation*

$$u_t = u_{xx}$$

governs the evolution of temperature with respect to time. In the case at hand the boundary conditions mean that the ends of the interval are kept at a constant temperature of zero. Now the basic solutions are given by

$$u(x,t) = e^{-n^2 t}\sin(nx).$$

and thus

$$u(x,t) = \sum_{n=1}^{\infty} a_n e^{-n^2 t}\sin(nx).$$

Note that the Fourier coefficients are severely damped for any positive time; $u(x,t)$ is in fact a real-analytic function of x for $t > 0$.

An actual plucked guitar string does not have a periodic motion but a motion that smooths and decays with time. It obeys a combination of the heat and wave equations:

$$u_{tt} + 2\delta u_t = u_{xx}.$$

Here the basic solutions are given by

$$u(x,t) = \exp(\alpha_n t)\sin(nx)$$

where $\alpha_n^2 + 2\delta\alpha_n + n^2 = 0$. So long as $0 \le \delta \le 1$ we get

$$\alpha_n = -\delta \pm i\sqrt{n^2 - \delta^2}$$

and thus the solution with $u_t(x,0) = 0$ has the form

$$u(x,t) = \exp(-\delta t)\cos(\omega_n t + \sigma_n)\sin(nx),$$

where $\omega_n = \sqrt{n^2 - \delta^2}$ and $\tan(\sigma_n) = -\delta/\omega_n$. Note that the frequencies are now slowed and out of harmony — their ratios are no longer rational — and that $u(t,x)$ is damped but not smoothed out over time!

Discrete Fourier series. Similarly, given $a_n \in \ell^2(\mathbb{Z})$, there is a function $f \in L^2(S^1)$ such that

$$a_n = \int_{S^1} f(x) e^{-inx} dx.$$

In other words, a_n is a 'continuous superposition' of the sequences $b_n^x = e^{inx}$.

Note that while $z^n/2\pi$ is a basis for $L^2(S^1)$, the sequences b_n^x are not even *in* $\ell^2(\mathbb{Z})$.

Convergence of Fourier series. One of the main concerns of analysts for 150 years has been the following problem: given a function $f(x)$ on S^1, in what sense is f represented by its Fourier series $\sum a_n \exp(inx)$?

It is traditional to write $S_N(f) = \sum_{-N}^{N} a_n(f)\exp(inx)$. The simplest answer to the question is the one we have just seen: so long as $f \in L^2(S^1)$, we have

$$\int |f - S_N(f)|^2 \to 0$$

as $N \to \infty$.

The question of pointwise convergence is equally natural: how can we extract the value $f(x)$ from the numbers a_n? Of course, if f is discontinuous this might not make sense, but we might at least hope that when $f(x)$ is continuous we have $S_N(f) \to f$ pointwise, or maybe even uniformly. In this direction we have:

Theorem 7.4 *If $f(x)$ is C^2, then $a_n = O(1/n^2)$ and thus $S_N(f)$ converges to f uniformly.*

In fact we have:

Theorem 7.5 (Dirichlet) *If $f(x)$ is C^1, then $S_N(f)$ converges uniformly to f.*

Dirichlet's proof ... left open the question as to whether the Fourier series of every Riemann integrable, or at least every continuous, function converged. At the end of his paper Dirichlet made it clear he thought that the answer was yes (and that he would soon be able to prove it). During the next 40 years Riemann, Weierstrass and Dedekind all expressed their belief that the answer was positive. —Körner, *Fourier Analysis*, §18.

In fact this is false!

Theorem 7.6 (DuBois-Reymond) *There exists an $f \in C(S^1)$ such that $\sup_N |S_N(f)(0)| = \infty$.*

Use of functional analysis. To see this, suppose to the contrary that the sup above is finite for all continuous f. That is, suppose the values of the linear functionals $f \mapsto S_N(f)(0)$ are bounded by M_f. Then, by the uniform boundedness principle, they are uniformly bounded:

$$\sup_N |S_N(f)(0)| \leq M\|f\|_\infty.$$

There is nothing special about the point zero, so in fact we have:

$$\|S_N(f)\|_\infty \leq M\|f\|_\infty$$

where M is independent of N.

Now let us further note that every L^∞ function is the limit in measure of a uniformly bounded sequence of continuous functions. (Put differently, $C(S^1)$ is dense in $L^\infty(S^1$ in the weak* topology.) Since each Fourier coefficient varies continuously under such weak* limits, we have $\|S_N(f)\|_\infty \leq M\|f\|_\infty$ for all $f \in L^\infty(S^1)$.

Next we note that $S_N(f)(0) = \sum_{-N}^N a_n(f)$ is simply the sum of the Fourier coefficients of f,

$$a_n(f) = \frac{1}{2\pi}\int_{S^1} f(x)\exp(-inx)\,dx.$$

Thus we can write $S_N(f)(0) = (1/2\pi)\int f D_N$, where D_N is the *Dirichlet kernel*

$$D_N(x) = \sum_{-N}^{N} \exp(-inx).$$

Then we have shown that

$$\left| \int D_N f \right| \leq M \|f\|_\infty$$

for all $f \in L^\infty$. But this implies $\|D_N\|_1 \leq M$.

We now show that in fact, $\|D_N\|_1 \to \infty$ as $N \to \infty$. Setting $q = \exp(inx)$, we have

$$D_N(x) = \sum_{-N}^{N} q^{-n} = q^{-N} \frac{1 - q^{2N+1}}{1 - q} = \frac{q^{N+1/2} - q^{-N-1/2}}{q^{1/2} - q^{-1/2}} = \frac{\sin((N+1/2)x)}{\sin(x/2)}.$$

Clearly all the action occurs near $x = 0$; indeed, we have $|D_N(x)| = O(1/|x|)$ on $[-\pi, \pi]$. But near $x = 0$, there are periodic intervals on which $|\sin((N+1/2)x)| > 1/2$. On these intervals, $|1/\sin(x/2)| \approx 2/|x|$. Since $\int dx/|x|$ diverges, we have $\|D_N\|_1 \to \infty$. In fact, the L^1-norm behaves like $\int_{1/N}^1 dx/x \asymp \log(N)$. ■

After this phenomenon was discovered, a common sentiment was that it was only a matter of time before a continuous function would be discovered whose Fourier series diverged everywhere. Thus it was even more remarkable when L. Carleson proved:

Theorem 7.7 *For any $f \in L^2(S^1)$, the Fourier series of f converges to f pointwise almost everywhere.*

The proof is very difficult.

The Fejér kernel. However in the interim Fejér, at the age of 19, proved a very simple result that allows one to reconstruct the values of f from its Fourier series for any continuous function.

Theorem 7.8 *For any $f \in C(S^1)$, we have*

$$f(x) = \lim \frac{S_0(f) + \cdots + S_{N-1}(f)}{N}$$

uniformly on the circle.

This expression is a special case of *Césaro summation*, where one replaces the sequence of partial sums by their averages. This procedure can

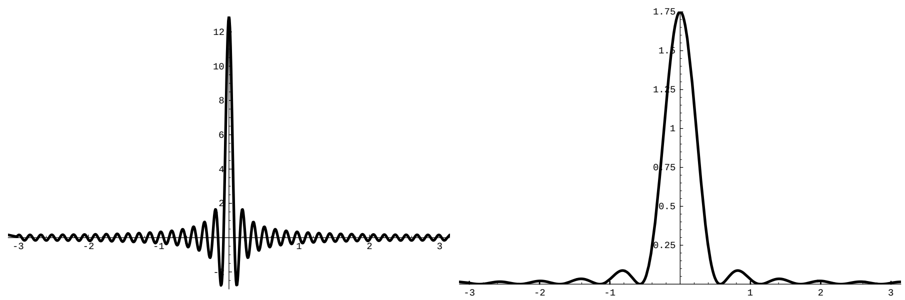

Figure 3. The Dirichlet and Fejér kernels.

be iterated. In the case at hand, it amounts to computing $\sum_{-\infty}^{\infty} a_n$ as the limit of the sums

$$\frac{1}{N}\sum_{i=-N}^{N}(N-|i|)a_n.$$

Approximate identities. To explain Fejér's result, it is useful to first understand the idea of convolution and approximate identities.

Writing S^1 as an additive group, for $f, g \in L^1(S^1)$ we let

$$(f * g)(x) = (1/2\pi)\int_{S^1} f(x)g(y-x)\,dx.$$

Note that $f * g = g * f$. It is easy to show that $(f * g)(x)$ is a *continuous* function of x; thus convolution is a *smoothing* operator.

We say f_n is an *approximation to the identity* if $f_n \geq 0$, $(1/2\pi)\int f_n = 1$ for all n and $f_n \to 0$ uniformly on compact sets outside $x = 0$.

Theorem 7.9 *If f_n is an approximation to the identity, and $g \in C(S^1)$, then $f * g \to g$ uniformly on S^1.*

Proof. Think of $f_n * g$ are a sum of the translates $g(x-y)$ weighted by $f_n(y)$. The translates with y small are uniformly close to g because g is continuous. The translates with y large make a small contribution because their total weight is small. Thus $f_n * g$ is uniformly close to g. ■

If we let

$$T_N(f) = \frac{S_0(f) + \cdots + S_{N-1}(f)}{N},$$

then $T_N(f) = f * F_N$ where the *Fejér kernel* is given by

$$F_N = (D_0 + \cdots D_{N-1})/N.$$

Of course $\int F_N = 1$ since $\int D_n = 1$. But in addition, F_N is *positive* and *concentrated near* 0, i.e. it is an approximation to the identity. Indeed, we have:

$$F_N(x) = \frac{\sin^2(Nx/2)}{N\sin^2(x/2)}\cdot$$

To see the positivity more directly, note for example that

$$\begin{aligned}(2N+1)F_{2N+1} &= z^{-2N} + 2z^{-2N+1} + \cdots + (2N+1) + \cdots 2z^{2N-1} + z^{2N} \\ &= (z^{-N} + \cdots z^N)^2 = D_N^2,\end{aligned}$$

where $z = \exp(ix)$.

Fourier transform. One of the great ideas in analysis is the *Fourier transform* on $L^2(\mathbb{R})$. We define it on $f \in L^2(\mathbb{R})$ by

$$\widehat{f}(\xi) = \int_{\mathbb{R}} f(x)e^{-ix\xi}\,dx.$$

This integral at least makes sense when f is smooth and compactly supported.

We claim $(\widehat{f}, \widehat{f})$ is a constant multiple of (f, f). Indeed, on any interval $[-\pi M, \pi M]$ large enough to contain the support of f, we have an orthonormal basis $g_i = e^{inx/M}/\sqrt{2\pi M}$; writing $f = \sum a_i g_i$ we find $a_i = \widehat{f}(n/M)/\sqrt{2\pi M}$, and thus

$$(f, f) = \sum |a_i|^2 = \frac{1}{2\pi M}\sum \widehat{f}(n/M) \to \frac{1}{2\pi}\int |\widehat{f}|^2$$

as $n \to \infty$. Thus f extends to all of L^2 as an isometry.

Fourier transform and differential equations. The Fourier transform reverse small-scale and large scale features of f. It turns differentiation d/dx_i into multiplication by x_i. Thus $\widehat{f}(0) = \int f$; if f is smooth then $\widehat{f}$ decays rapidly at infinity; etc.

Since differentiation is turned into multiplication, it becomes easy to solve PDEs. For example, to solve $\Delta u = f$, you just pass to the transform side and divide $\widehat{f}$ by $\sum x_i^2$. There is no difficultly near infinity for the result to be in L^2; this reflects that fact that Δ is a smoothing operator. There *is* difficulty near 0: both $\int f$ and the moments $\int f x_i$ should vanish for u to be in L^2.

Spherical harmonics. We can also look from $L^2(S^1)$ in another direction — towards $L^2(S^{n-1})$, where the domain remains compact but its symmetry group becomes larger $G = \mathrm{SO}(n)$. How do Fourier series generalize to the higher-dimensional spheres?

The case of a sphere is especially convenient because we can regard S^{n-1} as the unit ball $|x| = 1$ in $\mathbb{R}^n$. Let P_d denote the space of homogeneous polynomials of degree d on $\mathbb{R}^n$. We have

$$\dim P_d = \binom{d+n-1}{n-1}.$$

The *Laplacian* $\Delta = \sum d^2/dx_i^2$ maps P_d to P_{d-2}; its kernel H_d is the space of *harmonic polynomials* of degree d. The key property of the Laplace operator is that it is $\mathrm{SO}(n)$-invariant.

Theorem 7.10 *We have* $L^2(S^{n-1}) = \oplus_{d=0}^{\infty} H_d$.

Generally a function $f \in H_d$ or its restriction to S^{n-1} is called a *spherical harmonic*. It can be shown that f, considered as a function the sphere, is actually an *eigenfunction* of the spherical Laplacian.

One can also study issues of pointwise convergence in this setting, for example one has:

Theorem 7.11 *If* $f \in C^2(S^{n-1})$ *then its Fourier series converges uniformly.*

To begin the proof that the spherical harmonics forms a 'basis' for $L^2(S^{n-1})$, we first show there is no relation between them.

Proposition 7.12 *The restriction map from* $\mathbb{R}^n$ *to* S^{n-1} *is injective on* $\oplus H_d$.

Proof. A harmonic polynomial which vanishes on the sphere is everywhere zero, by the maximum principle. ∎

Raising operator. Of course this result fails for general polynomails, because $r^2 = \sum x_i^2$ is constant on S^{n-1}. To take this into account, we introduce the *raising operator*

$$L : P_d \to P_{d+2}$$

defined by $L(f) = r^2 f$. Here are some of its key properties and their consequences.

1. If f is harmonic, then $\Delta L(f) = 2(n+d)f$. This is because

$$\Delta(r^2 f) = (\Delta r^2)f + (\nabla r^2) \cdot (\nabla f) + r^2 \Delta f$$

2. More generally, we have $\Delta L(f) = 2(n+d)f + L\Delta f$, i.e. $[\Delta, L] = 2(n+d)$. From this we find inductively:

$$\Delta L^{k+1} = C_k L^k + L^{k+1}\Delta,$$

   where $C_k \neq 0$. This shows:

$$\Delta(r^{2k+2} H_d) = r^{2k} H_d$$

   (and of course the map is an isomorphism because both sides have the same dimension).

3. We can now prove by induction:

$$P_d = H_d \oplus r^2 H_{d-2} \oplus r^4 H_{d-4} \oplus \cdots$$

   Indeed, once this is known for P_d we simply consider $\Delta : P_{d+2} \to P_d$. This map has kernel H_{d+2} and maps $r^2 H_d$ bijectively to H_d, etc.

4. As a Corollary we immediately see that $\oplus H_d | S^{n-1}$ is the same space of functions as $\oplus P_d | S^{n-1}$, since $r = 1$ on S^{n-1}. In particular, $\oplus H_d$ is dense in $L^2(S^{n-1})$.

5. It remains to check that H_d and H_e are orthogonal for $d \neq e$. One way is to consider the spherical Laplacian and note that these are eigenspaces with different eigenvalues. Another way is to consider the character of SO(2) acting on H_d.

6. The combination of these observations proves the spherical harmonics form a basis for $L^2(S^{n-1})$.

Low-dimensional examples. For example, when $n = 2$ we have $\dim H_0 = 1$ and $\dim H_d = 2$ for $d > 0$. A basis is given by $\operatorname{Re} z^d$ and $\operatorname{Im} z^d$.

For $n = 3$ we have $\dim h_d = 2d+1 = 1, 3, 5, \ldots$. It is traditional to form a complex basis Y_{md} for H_d where $-d \le m \le d$, and

$$Y_{md}(x, y, z) = (x \pm iy)^{|m|} P_d^m(z).$$

Here $P_d^m(z)$ is a Legendre polynomial.

The hydrogen atom. The simplest model for the hydrogen atom in quantum mechanics has as states of pure energy the functions f on $\mathbb{R}^3$ which satisfy

$$\Delta f + r^{-1} f = E f.$$

It turns out a basis for such functions has the form of products of radial functions with spherical harmonics. The energy is proportional to $1/N^2$ where N is the principal quantum number. For a given N, the harmonics with $0 \le d < N-1$ all arise, each with multiplicity $2d+1$, so there are N^2 independent states altogether. The states with $d = 0, 1, 2, 3, \ldots$ are traditionally labelled s, p, d, f, g, h.

Irreducibility. Is there a finer Fourier series that is still natural with respect to rotations? The answer is no:

Theorem 7.13 *The action of* $\mathrm{SO}(n)$ *on* H_d *is irreducible.*

Proof. There are many proofs of irreducibility; here is a rather intuitive, analytic one.

Suppose the action of $\mathrm{SO}(n)$ on H_d splits nontrivially as $A \oplus B$. Then we can find in each subrepresentation a function such that $f(N) = 1$, where N is the 'north pole' stabilized by $\mathrm{SO}(n-1)$; and by averaging over $\mathrm{SO}(n-1)$, we can assume f is constant on each sphere $S^{n-2} \subset S^{n-1}$ centered at N. In particular, if we consider a ball $B \subset S^{n-1}$ centered at N and of radius $\epsilon > 0$, we can find a nonzero $f \in H_d$ with $f|\partial B = 0$ and $\max f|B = 1$.

Considere the cone $U = [0, 2]B \subset \mathbb{R}^n$. Then f is a harmonic function which vanishes on all of the boundary of B except the cap $2B$. By homogeneity, $\max f|2B = 2^d$. In addition, there is an $x \in B$ where $f(x) = 1$. By the mean value property of harmonic functions, $f(x)$ is the average of the values $f(y)$ over the points y where a random path initiated at x first exists U. But the probability that the path exits through the cap $2B$ is $p(\epsilon) \to 0$ as $\epsilon \to 0$. Thus

$$1 = f(x) \le 2^d p(\epsilon) \to 0,$$

a contradiction. ∎

(Note: this argument gives *a priori* control over the diameter of a closed 'nodal set' for an eigenfunction of the Laplacian on S^{n-1} in terms of its eigenvalue.)

Spherical Laplacian. Here is a useful computation for understanding spherical harmonics intrinsically.

To compute the Laplacian of $f|S^{n-1}$, we use the formula:

$$\Delta_s(f) = \nabla \cdot \pi_s(\nabla f),$$

where

$$\pi_s(\nabla f) = \nabla f - (\widehat{r} \cdot \nabla f)\widehat{r}$$

is the projection of ∇f to a vector field tangent to the sphere. Using the fact that $\nabla \cdot \widehat{r} = n-1$, this gives:

$$\Delta_s(f) = \Delta(f) - (n-1)(df/dr) - d^2 f/dr^2.$$

Now suppose f is a spherical harmonic of degree ℓ. Then $\Delta(f) = 0$, $df/dr = \ell f$, and $d^2 f/dr^2 = \ell(\ell-1)f$, which yields:

Theorem 7.14 *If $f \in H_\ell(\mathbb{R}^n)$ then $f|S^{n-1}$ is an eigenfunction of the spherical Laplacian, satisfying*

$$\Delta_s(f) = -\ell(\ell+n-2)f.$$

8 General Measure Theory

Measures. A measure $(X, \mathcal{B}, m)$ consists of a map $m : \mathcal{B} \to [0, \infty]$ defined on a σ-algebra of subsets of X, such that $m(\emptyset) = 0$ and such that $\sum m(B_i) = m(\bigcup B_i)$ for countable unions of disjoint $B_i \in \mathcal{B}$.

Countable/Co-countable measure. An example is the measure defined on any uncountable set X by taking $\mathcal{B}$ to be the σ-algebra generated by singletons and $m(B) = 0$ or ∞ depending on whether B is countable or $X - B$ is countable.

Hausdorff measure. This is defined on the Borel subsets of $\mathbb{R}^n$ by

$$m_\delta(E) = \lim_{r \to 0} \inf_{E = \bigcup E_i} \sum \operatorname{diam}(E_i)^\delta,$$

where $\operatorname{diam}(E_i) \le r$. Appropriately scaled, m_n is equal to the usual volume measure on $\mathbb{R}^n$.

Dimension; the Cantor set. The Hausdorff *dimension* of $E \subset \mathbb{R}^n$ is the infimum of those δ such that $m_\delta(E) = 0$.

For example, the usual Cantor set E can be covered by 2^n intervals of length $1/3^n$, so its dimension is at most $\log 2/\log 3$. On the other hand, there is an obvious measure on E such that $m(A) \leq C(\operatorname{diam} E)^{\log 2/\log 3}$ and from this it is easy to prove the dimension is equal to $\log 2/\log 3$.

Linear maps and dimension. Clearly Hausdorff measure satisfies $m_\delta(\alpha E) = \alpha^\delta m(E)$. So for the Cantor set E built on disjoint subintervals of lengths a, b $a + b < 1$ in $[0, 1]$, one has $a^\delta + b^\delta = 1$ *if* $0 < m_\delta(E) < \infty$.

This makes it easy to guess the dimension of self-similar fractals. The self-affine case is much harder; cf. the M curve, of dimension $1 + 2^{\log 2/\log 3}$.

Signed measures. To make the space of all measure into a linear space, we must allow measures to assume negative values.

A *finite signed measure* m on a σ-algebra $\mathcal{B}$ is a map $m : \mathcal{B} \to [-M, M]$, such that for any sequence of disjoint B_i we have

$$\sum m(B_i) = m(\bigcup B_i).$$

Note that the sum above converges absolutely, since the sum of its positive terms individually is bounded above by M, and similar for the negative terms.

A general signed measure is allowed to assume at most one of the values $\pm\infty$, and the sum above is required to converge absolutely when $m(\bigcup B_i)$ is finite.

A *measure* is a signed measure assuming no negative values.

For simplicity we will restrict attention to **finite** signed measures.

Positive sets. Given a signed measure m, a set P is *positive* if $m(A) \geq 0$ for all $A \subset P$.

Theorem 8.1 *If $m(A) > 0$ then there is a positive set $P \subset A$ with $m(P) \geq m(A)$.*

Proof. Let $\lambda(A) = \inf\{m(B) : B \subset A\} \geq -M$. Pick a set of nonpositive measure, $B_1 \subset A$, with $m(B_1) < \lambda(A) + 1$. By induction construct a set of nonpositive measure $B_{n+1} \subset A_n = A - (B_1 \cup \ldots \cup B_n)$ with $m(B_{n+1}) < \lambda(A_n) - 1/n$. Then $\sum |m(B_i)| < \infty$, so $m(B_i) \to 0$ and thus $\lambda(A_i) \to 0$.

Letting $P = \bigcap A_n$, we have $P \subset A_n$ so $\lambda(P) \geq \lim \lambda(A_n) = 0$. Thus P is a positive set, and $m(P) \geq m(A)$ since $m(B_i) \leq 0$ for each i. ∎

The Hahn Decomposition.

Theorem 8.2 *Given a finite signed measure m on X, there is a partition of X into a pair A, B of disjoint sets, one positive and one negative.*

Proof. Let $p = \sup m(P)$ over all positive sets $P \subset X$. We claim p is achieved for some positive set A. Indeed, we can choose positive sets A_i with $m(A_i) \to p$ and just let $A = \bigcup A_i$.

Now let $B = X - A$. Then B contains a set of positive measure, then it contains a positive set P of positive measure; then $m(A \cup P) > m(A) = p$, contrary to the definition of p. Thus B is negative. ∎

Jordan decomposition.

Theorem 8.3 *Let m be a signed measure on X. Then m can be uniquely expressed as $m = p - n$, where p and n are mutually singular (positive) measures.*

Here *mutually singular* means p and n are supported on disjoint sets.

Proof. Let $p = m|A$ and $n = -m|B$, where $A \cup B$ is the Hahn decomposition of X (unique up to null sets). This shows p and n of the required form exist.

Now assuming only that $m = p - n$, where p and n are mutually singular, we can assert that $p(A) = \sup\{m(B) \ : \ B \subset A\}$, and thus p is unique. Similarly n is unique. ∎

Absolute continuity. Given a pair of measures μ and λ, we say $\mu \ll \lambda$, or μ is *absolutely continuous* with respect to λ, if $\lambda(E) = 0 \implies \mu(E) = 0$.

For example, $X = [0, 1]$ and $\mu(E) = \int_E f(x)\, dx$ for $f \in L^1[0, 1]$, then $\mu \ll \lambda$ if λ is Lebesgue measure on $[0, 1]$. In fact the converse holds.

The Radon-Nikodym theorem.

Theorem 8.4 *If $\mu \ll \lambda$ then there is an $f \geq 0$ such that*

$$\mu(E) = \int_E f(x)\, d\lambda.$$

Proof. If f has the form above, then the Hahn decomposition of μ is $\{f < 0\} \cup \{f > 0\}$. Similarly the Hahn decomposition of $\mu - \alpha\lambda$ is $\{f < \alpha\} \cup \{f > \alpha\}$.

So for each rational number α, let P_α be the positive set for the Hahn decomposition of $\mu - \alpha\lambda$. Then the P_α are nested (up to null sets). Define $f(x) = \sup\{\alpha \ : \ x \in P_\alpha\}$, and set $\nu(A) = \int_A f\, d\lambda$.

Now notice that for $\alpha < \beta$, for any A contained in

$$\{\alpha \leq f \leq \beta\} = P_\alpha - P_\beta,$$

we have $\nu(A)$ and $\mu(A)$ both contained in $[\alpha, \beta]\lambda(A)$. Chopping $[0, \infty]$ into intervals of length $1/n$, and pulling these intervals back to a decomposition E_i of a set E, we find that $\mu(E)$ is sandwiched between the upper and lower approximations to $\int_E f\, d\lambda$. Therefore equality holds. ∎

Derivatives. The function f defined above is commonly written $f = d\mu/d\lambda$, so we have

$$\mu = \frac{d\mu}{d\lambda}\, d\lambda.$$

Absolutely continuous/singular decomposition. Given a pair of measures μ, ν on $(X, \mathcal{B})$, we can naturally decompose $\nu = \nu_a + \nu_s$ with $\nu_a \ll \mu$ and $\nu_s \perp \mu$.

To do this, just let $\pi = \mu + \nu$, and write $d\mu = f\, d\pi$, $\nu = g\, d\pi$ (using Radon-Nikodym derivatives). Then we have $\nu = g/f\, d\mu$ on the set where $f > 0$, and $\nu \perp \mu$ on the set where $f = 0$. These two restrictions give the desired decomposition of ν.

Baire measures. We now pass to the consideration of measures μ on a *compact* Hausdorff space X compatible with the topology. The natural domain of such a measure is not the Borel sets but the **Baire** sets $\mathcal{K}$, the smallest σ-algebra such that all $f \in C(X)$ are measurable.

A *Baire measure* is a measure m on $(X, \mathcal{K})$.

What's the distinction? In $\mathbb{R}$, all closed sets are $G_\delta' s$, so their preimages under functions are also G_δ. Thus $\mathcal{K}$ is generated by the closed G_δ's in X, rather than all closed sets.

In a compact metric space, the Borel and Baire sets coincide.

Regular contents. It is useful to have a characterization of those functions $\lambda : \mathcal{F} \to X$ defined on the closed (hence compact) sets $\mathcal{F}$ in X such that λ extends to a Baire measure. Here it is:

Theorem 8.5 *Let $\lambda(K) \geq 0$ be defined for all compact G_δ sets $K \subset X$ and satisfy:*

(i) $\lambda(K_1) \leq \lambda(K_2)$ if $K_1 \subset K_2$;
(ii) $\lambda(K_1 \cup K_2) = \lambda(K_1) + \lambda(K_2)$ if K_1 and K_2 are disjoint; and
$\lambda(K) = \inf \lambda(\overline{U})$ over all open sets $U \supset K$.

Then there is a unique Baire measure μ such that $\mu(K) = \lambda(K)$ for all compact K.

Such a λ is called a *regular content* on X.

Sketch of the proof. Given λ, we can define a set-function (inner measure) by

$$\mu_*(E) = \sup_{K \subset E} \lambda(K),$$

define a set A to be measurable if $\mu_*(A \cap E) + \mu_*((X - A) \cap E) = \mu_*(E))$ for all E, show that the measurable sets contain the Baire sets and that $\mu = \mu_*$ is a Baire measure extending λ.

Positive functionals.

Theorem 8.6 *Let $\phi : C(X) \to \mathbb{R}$ be a linear map such that $f \geq 0 \implies \phi(f) \geq 0$. Then there is a unique Baire measure μ on X such that*

$$\phi(f) = \int_X f \, d\mu.$$

Proof. Let us say $f \in C(X)$ is *admissible* for a compact G_δ set K if $f \geq 0$ and $f \geq 1$ on K. Define $\lambda(K)$ as $\inf \phi(f)$ over all admissible f.

We claim λ is a regular content. (i) is clear; as for (ii), $\lambda(K_1 \cup K_2) \leq \lambda(K_1) + \lambda(K_2)$ is obvious. For the reverse inequality, use normality of X to get $g_1 + g_2 = 1$, $g_1 g2 = 0$, $0 \leq g_i \leq 1$ with $g_i = 1$ on K_i. Then given f for $K_1 \cup K_2$ we get competitors $f_i = g_i f$ for K_i with $\phi(f_1) + \phi(f_2) \leq \phi(f)$, so $\lambda(K_1 \cup K_2) \leq \lambda(K_1 \cup K_2)$.

Finally for (iii): choose f admissible for K with $\phi(f) \leq \lambda(K) + \epsilon$. Let $U = \{f > 1 - \epsilon\}$. Then $f/(1-\epsilon)$ is admissible for $\overline{U}$, so $\lambda(\overline{U}) \leq (\lambda(K) + \epsilon)/(1-\epsilon)$, and therefore $\lambda(K) = \inf \lambda(\overline{U})$.

Thus λ extends to a Baire measure μ. To show that integration against μ reproduces ϕ, first note that for any K there exist admissible f_n decreasing to χ_K pointwise, with f_n eventually vanishing on any compact set L disjoint from K (since K is a G_δ), and for which $\int f_n$ and $\phi(f_n)$ both converge to $\mu(K) = \lambda(K)$.

Thus we can approximate χ_K by a continuous function f with $\phi(f) \approx \int f$. Now approximate g from above by sums of indicator functions of compact sets, and approximate these from above by admissible functions f; then we get $\phi(g) \leq \phi(f) \approx \int g \, d\mu$. Doing the same from below we find that $\phi(g) = \int g \, d\mu$. ∎

Theorem 8.7 (Riesz) *Let X be a compact Hausdorff space. Then the dual of $C(X)$ is the space of Baire measures on X, with $\mu = |\mu|(X)$.*

Proof. One shows a linear functional can be decomposed into a positive and negative part, each of which is represented by a measure. ■

Corollary. The space of measures on a compact Hausdorff space is compact in the weak* topology.

Functions of bounded variation and signed measures on $[a, b]$. We can now address afresh the theory of differentiation of $f : [a, b] \to \mathbb{R}$. To each signed measure μ we can associate the function $f(x) = \mu[a, x]$. This function is continuous from above and of bounded variation. Conversely, to each such f one can attach a measure df. The weak topology is the one where $f_n \to f$ iff $f_n(x) \to f(x)$ for each x such that f is continuous at x.

Now signed measure correspond to functions in BV; absolutely continuous measures, to absolutely continuous functions; $f'(x)$ is $d\mu_a/d\lambda$; discontinuities correspond to atoms; singular measure correspond to f with $f' = 0$ a.e.

Compactness. As an alternative proof of compactness: consider a sequence of monotone increasing functions $f : [a, b] \to [0, 1]$ with $f(b) = 1$. (I.e. a sequence of probability measures.) Passing to a subsequence, we can get $f_n(x)$ to converge for all rational $x \in [a, b]$. Then there is a monotone limit g, which can be arranged to be right-continuous, such that $f_n \to g$ away from its discontinuities.

Integration. Given a function f of bounded variation and $g \in C^\infty[a, b]$, we can define

$$I = \int_a^b g(x)\, df(x) \quad = \quad -\int_a^b f(x) g'(x)\, dx.$$

Now breaking $[a, b]$ up into intervals $[a_i, a_{i+1}]$ we get the approximation:

$$\begin{aligned} I &= -\sum f(a_i)(g(a_{i+1}) - g(a_i)) \\ &= +\sum (f(a_{i+1}) - f(a_i)) g(a_i) = O(\|f\|_{BV} \|g\|_\infty). \end{aligned}$$

Thus integration against df gives a bounded linear functional on a dense subset of $C[a, b]$, so it extends uniquely to a measure.

This idea is the beginnings of the theory of distributions.

Sample application: Let $f : X \to X$ be a homeomorphism. Then there exists a probability measure μ on X such that $\mu(A) = \mu(f(A))$.

Proof. Take any probability measure — such as a point mass δ; average it over the first n iterates of f; and take a weak* limit. ■

Haar measure. If G is a compact Hausdorff topological group, for each open neighborhood U of the origin we define $\lambda_U(K) = [K : U]/[G : U]$ where $[E : U]$ is the minimal number of left translates gU needed to cover E. Then as U shrinks towards the identity, we can extract **some** (Banach) limit of λ_U, which turns out to be a content λ. In this way we obtain a left-invariant measure on G.

References

[Con] J. H. Conway. *On Numbers and Games.* Academic Press, 1976.

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

[Ox] J. C. Oxtoby. *Measure and Category.* Springer-Verlag, 1980.