A Second Course in Real Analysis

Gaurav Goel

June 2022

Contents

1 Introduction to Measure Spaces and Measurable Functions

1.1 Measurable and Measure Spaces

We define an *interval* in $\mathbf{R}$ to be a connected subset; these are precisely the subsets of the form

$$\emptyset, \mathbf{R}, \{x\}, [a,b], (a,b), [a,b), (a,b], (a,\infty), [a,\infty), (-\infty,b), \text{ or } (-\infty,b]$$

for $x, a < b \in \mathbf{R}$. We define a notion of length for these as usual.

Attempt 1.1.1. We want a function $\mu : 2^{\mathbf{R}} \to [0,\infty)$ such that:

(a) If $I \subset \mathbf{R}$ is a bounded interval, then $\mu(I) = \ell(I)$. (This implies that $\mu(\emptyset) = 0$.)
(b) If $A \subseteq B$, then $\mu(A) \leq \mu(B)$.

This is a problem: since $[-n,n] \subseteq \mathbf{R}$ for all $n \geq 1$, we must have $2n = \mu[-n,n] \leq \mu(\mathbf{R})$ for all $\mathbf{R}$, which no number in $[0,\infty)$ can do. Therefore, extend to *extended real numbers* $[0,\infty]$. This has the great property that every series of numbers in $[0,\infty]$ converges to a number in it; further, rearrangement is fine, so that every countable collection of elements of $[0,\infty]$ can be summed to an element of $[0,\infty]$.

Attempt 1.1.2. We want a function $\mu : 2^{\mathbf{R}} \to [0,\infty]$ such that:

(a) If $I \subseteq \mathbf{R}$ is any interval, then $\mu(I) = \ell(I)$.
(b) If $A \subseteq B$, then $\mu(A) \leq \mu(B)$.
(c) If $A = \coprod_\alpha A_\alpha$ is any disjoint union, then $\mu(A) = \sum_\alpha \mu(A_\alpha)$.

Can we ask (c) for an arbitrary disjoint union? Well, no. This is because every $x \in \mathbf{R}$ gives an interval $[x,x] = \{x\}$ of length 0, so that $\mu(\{x\}) = 0$. On the other hand, if $A \subseteq \mathbf{R}$ is any subset, then $A = \coprod_{x\in A}\{x\}$, so this would imply that $\mu(A) = 0$ for any $A \subseteq \mathbf{R}$, which is absurd and contradicts (a). Besides, we only really know how to sum series; I don't know how to sum over uncountable sets! Therefore, reasonable that we ask (c) to hold only for countable disjoint unions. This automatically implies the result along with (a) for finite disjoint unions: take the remaining to be $\emptyset, \emptyset, \ldots$. On the other hand, the finite result implies (b): indeed, $B = A \coprod (B \setminus A)$, so that $\mu(A) \leq \mu(A) + \mu(B \setminus A) = \mu(B)$ and $\mu(B \setminus A) = \mu(B) - \mu(A)$ when $\mu(A) < \infty$; therefore, if I ask for this, I can remove (b) from my axioms.

Attempt 1.1.3. We want a function $\mu : 2^{\mathbf{R}} \to [0,\infty]$ such that:

(a) If $I \subseteq \mathbf{R}$ is any interval, then $\mu(I) = \ell(I)$.
(b) If $A = \coprod_i A_i$ is any countable disjoint union, then $\mu(A) = \sum_i \mu(A_i)$.
(c) If $x \in \mathbf{R}$ and $A \subseteq \mathbf{R}$, then $\mu(A + x) = \mu(A)$.

Theorem 1.1.4 (Vitali)**.** There does not exist a function $\mu : 2^{\mathbf{R}} \to [0,\infty]$ with the above properties.

Proof. Suppose there were; we will construct a pathological subset $V \subset \mathbf{R}$ which will lead to a contradiction. Partition $\mathbf{R}$ into equivalence classes where $r \sim r' \Leftrightarrow r - r' \in \mathbf{Q}$, and for each class pick a representative in $[0,1]$. Define $V \subseteq [0,1]$ to be the union of these representatives. If $(q_i)_i$ is an enumeration of $[-1,1] \cap \mathbf{Q}$, define $V_i := V + q_i$. Then by definition these are disjoint, and we have $[0,1] \subseteq \coprod_i V_i \subseteq [-1,2]$, so that $1 \leq \sum_i \mu(V) \leq 3$, a contradiction. ∎

What do we do? We restrict our attention to a large class $\mathcal{M} \subset 2^{\mathbf{R}}$ of well-behaved subsets whose measures we can ask for and which does not contain pathologies like V. What properties do we want it to satisfy? Well, it should contain all the sets we're interested in, like intervals, open subsets, and closed subsets, etc. We want the empty set to be measurable, we want the complement of a measurable set to be measurable, and we want the union of a collection of measurable sets to be measurable. For the same reason as before, we want to allow only countable unions.

Definition 1.1.5. Given a set X, a collection $\mathcal{M} \subseteq 2^X$ of subsets is called a *σ-algebra* if the following hold:

(a) $\emptyset \in \mathcal{M}$,
(b) if $A \in \mathcal{M}$, then $X \setminus A \in \mathcal{M}$, and
(c) if $A_i \in \mathcal{M}$ is a countable collection, then $\bigcup_i A_i \in \mathcal{M}$.

We call a subset $A \subseteq X$ to be $\mathcal{M}$*-measurable* (or simply *measurable* if $\mathcal{M}$ is clear from context) iff $A \in \mathcal{M}$. Such a pair $(X, \mathcal{M})$ is called a *measurable space*. Given measurable spaces $(X, \mathcal{M})$ and $(Y, \mathcal{N})$, a function $f : X \to Y$ is called *measurable* iff for each $B \in \mathcal{N}$, the preimage $f^{-1}(B) \in \mathcal{A}$. In this case, we write $f : (X, \mathcal{M}) \to (Y, \mathcal{N})$.

The conditions (b) and (c) on $\mathcal{M}$ imply that a countable intersection of measurable sets is measurable. A composition of measurable functions is measurable, so we get the category of measurable spaces.

Definition 1.1.6. A *measure* on a σ-algebra $\mathcal{M}$ on a set X is a function $\mu : \mathcal{M} \to [0, \infty]$ such that

(a) $\mu(\emptyset) = 0$, and
(b) if $A = \coprod_i A_i$ is a countable disjoint union of $A_i \in \mathcal{M}$, then $\mu(A) = \sum_i \mu(A_i)$.

Such a triple $(X, \mathcal{M}, \mu)$ is called a *measure space*.

This is a little like metrizable vs. metric spaces. Again, countable additivity implies finite additivity, and that if $A \subseteq B$ are both in $\mathcal{M}$, then so is $B \setminus A = B \cap (X \setminus A)$ and so $\mu(A) \leq \mu(B)$ with $\mu(B \setminus A) = \mu(B) - \mu(A)$ when $\mu(A) < \infty$ as before. We also get a few more properties:

Lemma 1.1.7. Let $(X, \mathcal{M}, \mu)$ be a measure space.

(a) If $A = \bigcup_i A_i$ is any countable union of $A_i \in \mathcal{M}$, then $\mu(A) \leq \sum_i \mu(A_i)$.
(b) (PIE) If $A_i \in \mathcal{M}$ for $i = 1, \ldots, n$ with $\mu(A_i) < \infty$ for all i, then

$$\mu\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{j=1}^{n} (-1)^{j-1} \sum_{|I|=j} \mu(A_I),$$

where $A_I := \bigcap_{i \in I} A_i$.
(c) If A_i is an increasing sequence of sets in $\mathcal{M}$, then $\mu(\bigcup_i A_i) = \lim_i \mu(A_i)$.
(d) If A_i is a decreasing sequence of sets in $\mathcal{M}$, then $\mu(\bigcap_i A_i) = \lim_i \mu(A_i)$ if[1] $\mu(A_i) < \infty$ for some i.

Proof.

(a) Define $B_i := A_i \setminus \bigcup_{j=1}^{i-1} A_i$ and apply countable additivity and $\mu(B_i) \leq \mu(A_i)$.
(b) Induct on n.
(c) Define $B_i = A_i \setminus A_{i-1}$ for $i \geq 1$. Then $\mu(\bigcup_i A_i) = \sum_j \mu(B_j) = \lim_i \sum_{j=1}^{i} \mu(B_j) = \lim_i \mu(\bigcup_{j=1}^{i} B_i) = \lim_i \mu(A_i)$.
(d) WLOG $\mu(A_1) < \infty$. For each i, define $C_i := A_1 \setminus A_1$. From (c), we get

$$\mu(A_1) - \mu(\cap_i A_i) = \mu(A_1 \setminus \cap_i A_i) = \mu(\cup_i C_i) = \lim_i \mu(C_i) = \lim_i \mu(A_1 \setminus A_i) = \mu(A_1) - \lim_i \mu(A_i).$$

■

Example 1.1.8.

(a) The counting measure on any set $(X, 2^X, |\cdot|)$.
(b) For $x \in X$, the *point mass at x* or *Dirac delta at x* by $(X, 2^X, \delta_x)$ where $\delta_x(A) = \mathbb{1}_{x \in A}$.
(c) A nonnegative combination of measures is again a measure. In particular, you can get a finite collection of point masses with different weights. In fact, infinite collections work too (c.f. Cohn, Ex. 6 and 7 of §1.2).
(d) If $Z \subseteq X$ is any subset, can define $\mathcal{M}_Z := \{Z \cap A : A \in \mathcal{M}\}$; this is a σ-algebra on Z and so $(Z, \mathcal{M}_Z)$ is a *measurable subspace.* If $\mathcal{M}$ is assigned a measure, then so can $\mathcal{M}_Z$ be in a compatible way iff Z is itself measurable.
(e) More generally, if $f : X \to Y$ is any function and $(Y, \mathcal{N})$ is a measurable space, then there is a unique smallest σ-algebra $f^*\mathcal{N}$ on X such that $f : (X, f^*\mathcal{N}) \to (Y, \mathcal{N})$ is measurable. If $(X, \mathcal{M})$ is already measurable, then an arbitrary function $f : X \to Y$ is measurable iff $f^*\mathcal{N} \subset \mathcal{M}$.

How do we actually specify such a collection $\mathcal{M}$ which is not 2^X?

Observation 1.1.9. If $\mathcal{M}_\alpha$ is a collection of σ-algebras, then the intersection $\bigcap_\alpha \mathcal{M}_\alpha$ is also one. In particular, given any subset $\mathcal{G} \subset 2^X$, there is a unique smallest σ-algebra containing $\mathcal{G}$; this is called the σ*-algebra* generated by $\mathcal{G}$ and will be denoted by $\langle \mathcal{G} \rangle$. A function $f : (X, \mathcal{M}) \to (Y, \langle \mathcal{G} \rangle)$ is measurable iff for each $B \in \mathcal{G}$, the preimage $f^{-1}(B) \in \mathcal{A}$.

[1]This is necessary, take $A_i = [i, \infty)$ on $\mathbf{R}$.

Example 1.1.10. Given a topological space X, the *Borel σ-algebra* is $\mathcal{B}(X) := \langle\mathcal{T}(X)\rangle$. Note that $\mathcal{B}(\mathbf{R}^d)$ can be generated by all closed subsets, or all half-spaces (i.e. $\{x_i \leq b\}$ or $\{x_i < b\}$ or the other way), or all half-open rectangles $\prod_i (a_i, b_i]$. Another space that will be particularly useful to us will be the space $[0, \infty]$ or $[-\infty, \infty]$, where a neighborhood basis of ∞ is given by sets of the form $(x, \infty]$ for $x \in \mathbf{R}_{>0}$ say (and similarly for $-\infty$). Again, this Borel σ-algebra is generated by any of the four types of subspaces: $[-\infty, b)]$ and $[(b, \infty]$ for $b \in \mathbf{R}$.

Optional: explanation of F_σ, G_δ, etc. and how this relates to $\langle\mathcal{N}\rangle$. Construction of the Borel σ-algebra using transfinite induction.

Example 1.1.11. Some measurable functions.

(a) A continuous map of topological spaces is measurable with respect to the Borel σ-algebras.
(b) Let $I \subseteq \mathbf{R}$ be an interval and $f : I \to \mathbf{R}$ say nondecreasing. Then for each $b \in \mathbf{R}$, the preimage $f^{-1}(-\infty, b)$ is an interval and so f is Borel measurable.
(c) Let $(X, \mathcal{M})$ be a measurable space. Given an $A \subseteq X$, the characteristic function $\chi_A : X \to \mathbf{R}$ is measurable with respect to $\mathcal{M}$ and the Borel algebra iff $A \in \mathcal{M}$.

Definition 1.1.12. Let $(X, \mathcal{M})$ be a measurable space. A function $f : X \to Y$ from X to a topological space Y is said to be a *simple* function if the image of X under f consists of finitely many points.

Let μ be a measure on $(X, \mathcal{M})$. If $\mathbf{F} = \mathbf{R}$ or $\mathbf{C}$, a function $f : X \to \mathbf{F}$ is called a *step* function if $f = \sum_{j=1}^n \alpha_j \mathbb{1}_{A_j}$ for some $\alpha_j \in \mathbf{F}$ and measurable $A_j \subseteq X$ with with $\mu(A_j) < \infty$.

In particular, step functions are simple functions. We typically care about simple measurable functions with codomain Y one of $\mathbf{R}, \mathbf{C}, [0, \infty)]$ and $[(-\infty, \infty)]$.

Definition 1.1.13. A measure space $(X, \mathcal{M}, \mu)$ is *finite* if $\mu(X) < \infty$ or equivalently $\mu : \mathcal{M} \to [0, \infty)$; it is *$\sigma$-finite* if there is an increasing countable collection $A_i \in \mathcal{M}$ with $\bigcup_i A_i = X$ and each $(A_i, \mathcal{M}_{A_i}, \mu|_{A_i})$ finite (i.e. $\mu(A_i) < \infty$).

The measure on any finite measure space $(X, \mathcal{M}, \mu)$ can be normalized so that $\mu(X) = 1$. In this case, we call $(X, \mathcal{M}, \mu)$ a *probability space*, X the space of events, $\mathcal{M}$ the collection of *possible outcomes*, and μ a *probability measure.*

The goal of the first few lectures will be to show:

Theorem 1.1.14. There is a σ-algebra $\mathcal{M}$ of *Lebesgue measurable subsets* of $\mathbf{R}$ such that (i) this contains $\mathcal{B}(\mathbf{R})$ and (ii) there is a normalized translation-invariant measure μ on $\mathcal{M}$ called the *Lebesgue measure*. Further, any two normalized translation-invariant measures on $\mathcal{B}(\mathbf{R})$ agree.

The same can be applied to $\mathbf{R}^d$, and indeed to any locally compact Hausdorff topological group.

We need a few technical definitions:

Definition 1.1.15. Let $(X, \mathcal{M}, \mu)$ be a measure space. A subset $N \subseteq X$ is said to be a *null set* if it is contained in some measurable subset of measure 0. A property $\mathcal{P}$ on X is said to hold *μ-almost everwhere* abbreviated μ-a.e. or simply a.e. when μ is understood) if it holds everywhere outside some null subset of X.

Definition 1.1.16. Let $(X, \mathcal{M}, \mu)$ be a measure space. The measure μ is said to be *complete* if any null set is measurable (and hence of measure 0).

When $(X, \mathcal{M}, \mu)$ is complete, a property $\mathcal{P}$ holds a.e. iff the subset $\{x \in X : \neg\mathcal{P}(x)\}$ has measure 0. We end this section with a simple lemma that follows immediately from the definitions.

Lemma 1.1.17 (Borel-Cantelli)**.** Let $(X, \mathcal{M}, \mu)$ be a measure space and let $A_i \subseteq X$ be a countable collection of measurable subsets of X such that $\sum_i \mu(A_i) < \infty$. Then almost all $x \in X$ belong to at most finitely many of the A_i's.

Proof. It suffices to show that the subset $\{x \in X : x$ lies in infinitely many A_i's$\} = \bigcap_i \left(\bigcup_{j\geq i} A_j\right)$ has measure 0. For each i, we have that $\mu(\cup_{j\geq i} A_j) \leq \sum_{j\geq i} \mu(A_j) < \infty$. Now, by Lemma 1.1.7(d), we have that

$$\mu\left(\bigcap_i \left(\bigcup_{j\geq i} A_j\right)\right) = \lim_i \mu\left(\bigcup_{j\geq i} A_j\right) \leq \lim_i \sum_{j\geq i} \mu(A_j) = 0.$$

∎

1.2 Properties of Measurable Functions

Here we first need some standard properties of $[-\infty,\infty]$-valued measurable functions. For any topological space Y, the collection of Y-valued functions on a measurable space $(X,\mathcal{M})$ that are measurable with respect to the Borel measure on Y will be denoted $\mathcal{M}(X,Y)$. Note first that a constant function is always measurable.

Theorem 1.2.1 (Restriction of Measurable Functions)**.** Let $(X,\mathcal{M})$ be a measurable space, $A\in\mathcal{M}$, and $f:A\to[-\infty,\infty]$.

(a) If f is measurable, then $f|_B$ is measurable for any measurable $B\subseteq A$.
(b) If $(A_i)_i$ is a collection of sets in $\mathcal{M}$ with $A=\bigcup_i A_i$ and each $f|_{A_i}$ is measurable, then f is measurable.

Proof. For (a), we have $f|_B^{-1}[-\infty,b)=f^{-1}[-\infty,b)\cap B$. For (b), note that $f^{-1}[-\infty,b)=\bigcup_i f|_{A_i}^{-1}[-\infty,b)$. ∎

Theorem 1.2.2 (New Measurable Functions from Old)**.** Let $(X,\mathcal{M})$ be a measurable space an $A\in\mathcal{M}$.

(a) Let $f,g:A\to[-\infty,\infty]$ be measurable[2] Then:
  (i) The sets $\{x\in A: f(x)<g(x)\}$, $\{x\in A: f(x)\le g(x)\}$ and $\{x\in A: f(x)=g(x)\}$ belong to $\mathcal{M}$.
  (ii) The functions $\min(f,g),\max(f,g):A\to[-\infty,\infty]$ are measurable.
(b) If $(f_n)_n$ is a sequence of $[-\infty,\infty]$-valued measurable functions on A, then:
  (i) The functions $\sup f_n$, $\inf f_n$, $\overline{\lim} f_n$ and $\underline{\lim} f_n$ are all measurable.
  (ii) The function $\lim f_n:\{x\in A:\underline{\lim}_n f_n(x)=\overline{\lim}_n f_n(x)\}\to[-\infty,\infty]$ is measurable.
(c) Let $f,g:A\to[0,\infty]$ be measurable functions. Then for any $\alpha\in[0,\infty]$, the functions αf and $f+g$ are also measurable, i.e. $\mathcal{M}(A,[0,\infty])$ is closed under $[0,\infty]$-linear combinations. Further, the function $\sqrt{f}:A\to[0,\infty]$ is also measurable.
(d) Let $f:A\to[-\infty,\infty]$ be any function. Then f is measurable iff both $f^+:=\max(f,0)$ and $f^-:=-\min(f,0)$ are. In this case, $|f|=f^++f^-$ is also measurable.
(e) The set $\mathcal{M}(A,\mathbf{R})$ is an $\mathbf{R}$-algebra. Further, if $f,g:A\to\mathbf{R}$ are measurable, then so is $f/g:A\setminus g^{-1}(0)\to\mathbf{R}$.
(f) A function $f:A\to\mathbf{R}^d$ is measurable iff each function $f_i:A\to\mathbf{R}$ is, so that $\mathcal{M}(A,\mathbf{R}^d)=\mathcal{M}(A,\mathbf{R})^d$. In particular, this is also an algebra and closed under formation of limits as in (b).
(g) The set $\mathcal{M}(A,\mathbf{C})$ is a $\mathbf{C}$-algebra, and if $f,g:A\to\mathbf{C}$ are measurable, then so is $f/g:A\setminus g^{-1}(0)\to\mathbf{C}$ and $|f|:A\to[0,\infty)$.

Proof.

(a) For (i), note that $f(x)<g(x)$ iff there is an $q\in\mathbf{Q}: f(x)<q<g(x)$. Therefore,

$$\{x\in A: f(x)<g(x)\}=\bigcup_{q\in\mathbf{Q}} f^{-1}[-\infty,q)\cap g^{-1}(q,\infty],$$

a countable union of subsets in $\mathcal{M}$ and hence in $\mathcal{M}$. The set $\{x\in A: f(x)\le g(x)\}=A\setminus\{x\in A: g(x)<f(x)\}$ is hence also in $\mathcal{M}$. Finally, the set $\{x\in A: f(x)=g(x)\}=\{x\in A: f(x)\le g(x)\}\setminus\{x\in A: f(x)<g(x)\}$. For (ii), note that $\max(f,g)^{-1}[-\infty,b)=f^{-1}[-\infty,b)\cap g^{-1}[-\infty,b)$, and $\min(f,g)^{-1}[-\infty,b)=f^{-1}[-\infty,b)\cup g^{-1}[-\infty,b)$.
(b) For (i), note that $(\sup f_n)^{-1}[-\infty,b]=\bigcap_n f_n^{-1}[-\infty,b]$ and $(\inf f_n)^{-1}[-\infty,b)=\bigcup_n f_n^{-1}[-\infty,b)$. Then $\overline{\lim} f_n=\inf_N\sup_{n\ge N} f_n$ and $\underline{\lim} f_n=\sup_N\inf_{n\ge N} f_n$ are also measurable. For (ii), note that by (a)(i) and (b)(i) we conclude that the domain of $\lim f_n$ is also measurable; since there is given simply by the restriction of either $\overline{\lim}$ or $\underline{\lim}$, we are done by Theorem 1.2.1(a).
(c) Note that $(\alpha f)^{-1}[0,b)=f^{-1}[0,b/\alpha)$. Further, $(f+g)(x)<b$ iff there is a $q\in\mathbf{Q}: f(x)<q$ and $g(x)<b-q$, so

$$(f+g)^{-1}[0,b)=\bigcup_{q\in\mathbf{Q}} f^{-1}[0,q)\cap g^{-1}[0,b-q).$$

Finally, $\sqrt{f}^{-1}[0,b)=f^{-1}[0,b^2)$.
(d) If f is measurable, then clearly f^+ and f^- and hence $|f|$ are by (a)(ii). Conversely, if both f^+ and f^- are measurable, then

$$f^{-1}[-\infty,b)=\begin{cases}(f^-)^{-1}(-b,\infty], & \text{if } b<0,\\ (f^-)^{-1}[0,\infty]\cup(f^+)^{-1}[0,b), & \text{if } b\ge 0.\end{cases}$$

[2] Here "measurable" means with respect to the subspace σ-algebra on A and the Borel σ-algebra on $[-\infty,\infty]$.

(e) The proof measurability of αf and $f+g$ for $\alpha \in \mathbf{R}$ is basically the same as above; for $\alpha < 0$, we just need to be careful to flip the direction of the inequality. Then $f - g = f + (-1)g$. Next, we show the measurability of the product fg; for this, we'll show first that $h : A \to \mathbf{R}$ is measurable then so is h^2; this will suffice by $fg = \frac{1}{2}((f+g)^2 - f^2 - g^2)$. This is clear from

$$(h^2)^{-1}[-\infty, b) = \begin{cases} \emptyset, & \text{if } b \leq 0, \\ h^{-1}(-\sqrt{b}, \sqrt{b}), & \text{if } b > 0. \end{cases}$$

By (a)(i), $g^{-1}(0) \in \mathcal{M}$, so that $A \setminus g^{-1}(0) \in \mathcal{M}$. Now, the result follows from

$$(f/g)^{-1}[-\infty, b) = \left(g^{-1}(0,\infty] \cap \{x \in A : f(x) < bg(x)\}\right) \cap \left(g^{-1}[-\infty, 0) \cap \{x \in: A : f(x) > bg(x)\}\right).$$

(f) The projection map $\pi_i : \mathbf{R}^d \to \mathbf{R}$ is measurable simply because $\pi_i^{-1}(-\infty, b) = \{x_i < b\}$; a composite of measurable functions is measurable. Conversely, if each f_i is measurable, then $f^{-1}\{x_i < b\} = f_i^{-1}(-\infty, b)$.

(g) By (f), a complex-valued function f is measurable iff $\mathrm{Re}(f)$ and $\mathrm{Im}(f)$ are. The rest of the proof follows immediately from (e) except possibly the last statement, which follows from $|f| = \sqrt{\mathrm{Re}(f)^2 + \mathrm{Im}(f)^2}$, which uses (c).

■

Theorem 1.2.3 (Simple/Step Approximation). Let $(X, \mathcal{M})$ be a measurable space and $A \in \mathcal{M}$. Let $f : A \to [0,\infty]$ be measurable. Then there is a sequence $(f_n)_n$ of simple functions $f_n : A \to [0, \infty)$ such that $f_1 \leq f_2 \leq \cdots \leq f$ pointwise and $f = \lim_n f$. When A is σ-finite, the sequence $(f_n)_n$ can be chosen to consist of step functions.

Proof. For each $n \geq 1$ and $k = 0, 1, 2, \ldots, n2^n - 1$, let $A_{n,k} := f^{-1}[k2^{-n}, (k+1)2^{-n})$. These are all in $\mathcal{M}$. Define the function f_n to have value $k2^{-n}$ on $A_{n,k}$ and n in $A \setminus \bigcup_k A_{n,k}$. Then the f_n are simple, measurable because the preimages of each of the values of f_n are the $A_{n,k}$'s and $A \setminus \bigcup_k A_{n,k}$ which are measurable. To see that $f_n \leq f_{n+1}$, note first that this is automatic on $f^{-1}[n, \infty)$ since $f_n = n$ there but $f_{n+1} \geq 2n(2^{-n-1}) = n$. On $[0, n)$ this follows from observing that for any $k = 0, \ldots, n2^n - 1$ we have

$$f^{-1}[k2^{-n}, (k+1)2^{-n}) = f^{-1}[(2k)2^{-n-1}, (2k+1)2^{-n-1}) \coprod f^{-1}[(2k+1)2^{-n-1}, (2k+2)2^{-n-1}),$$

and f_n and f_{n+1} agree on the first piece and $f_n = k2^{-n} < (2k+1)2^{-n-1} = f_{n+1}$ on the second piece. It is also clear that $f_n \leq f$ for any n.

Finally, to show that $f = \lim f_n$, note first that if $f(x) = \infty$, then for any $n \geq 1$ we have that $x \in A \setminus \bigcup_k A_{n,k}$ so that $f_n(x) = n$ and hence $\lim_n f_n(x) = \infty$. Next, suppose that $f(x) = b < \infty$ and let $\varepsilon > 0$. Pick $n \geq 1$ big enough so that both $n > b$ and $2^{-n} < \varepsilon$. Then there is a unique $k = 0, \ldots, n2^n - 1$ such that $b \in [k2^{-n}, (k+1)2^{-n})$ (indeed, $k = \lfloor b2^n \rfloor \leq b2^n < n2^n$), and then we have that $f_n(x) = k2^{-n}$ with $f(x) - f_n(x) = b - k2^{-n} < 2^{-n} < \varepsilon$.

When A is σ-finite, we can write $A = \bigcup_{n\geq 1} A_n$ for a countable sequence of increasing measurable A_n with each $\mu(A_n) < \infty$ and, and so in the above simply replace f_n by $f_n \mathbb{1}_{A_n}$. ■

We now turn a little towards a.e. equality of functions.

Theorem 1.2.4. Let $(X, \mathcal{M}, \mu)$ be a *complete* measure space.

(a) If $f, g : X \to [-\infty, \infty]$ are equal a.e., then f is measurable iff g is.

(b) If $(f_n) : X \to [-\infty, \infty]$ is a sequence of a measurable functions and $f : X \to [-\infty, \infty]$ such that $f = \lim f_n$ a.e. Then f is measurable.

Proof. For (a), suppose that f is measurable. Let $C \subseteq X$ be measurable with $\mu(C) = 0$ and such that $f(x) = g(x)$ for all $x \in X \setminus C$. Then

$$g^{-1}[-\infty, b) = \left(f^{-1}[-\infty, b) \cap (X \setminus C)\right) \cup (g^{-1}[-\infty, b) \cap C).$$

The measurability of f and of C implies that $f^{-1}[-\infty, b) \cap (X \setminus C)$ is measurable, whereas the completeness of C implies that $g^{-1}[-\infty, b) \cap C$ is measurable; this implies that $g^{-1}[-\infty, b)$ is measurable for any b and so so is g. For (b), note that $\underline{\lim} f_n$ is also measurable; so by part (a), since $f = \underline{\lim} f$ a.e., we conclude that f is measurable. ■

Remark 1. The assumption on completeness cannot be removed; indeed, suppose that $(X, \mathcal{M}, \mu)$ is not complete. Then there is a sequence $N \subseteq C \subseteq X$ where $C \in \mathcal{M}, \mu(C) = 0$ but $N \notin \mathcal{M}$. Then χ_N and 0 agree almost everywhere, but 0 is measurable while χ_N is not. Simiarly, the sequence which is identically 0 converges to χ_N almost everywhere, but χ_N is not measurable. Therefore, the previous theorem can be taken to be a characterization of complete measure spaces. However, complete measure spaces are not hard to come by, as observed in the next section.

1.3 Completions

2 Outer Measures and The Carathéodory Condition

Now we look at a standard technique of constructing measures.

Definition 2.0.1. Let X be any set. An *outer measure* on X is a function $\mu^* : 2^X \to [0, \infty]$ such that

(a) $\mu^*(\emptyset) = 0$,
(b) if $A \subseteq \bigcup_i A_i$, i.e. A is covered by a countable collection A_i, then $\mu^*(A) \le \sum_i \mu^*(A_i)$.

Note that (b) in particular implies that $A \subseteq B \Rightarrow \mu^*(A) \le \mu^*(B)$. The genius of Carathéodory was recognizing the correct condition defining measurability.

Definition 2.0.2 (Carethéodory)**.** Let X be a set and μ^* an outer measure on it. A subset $A \subseteq X$ is called μ^*-*measurable* if for all subsets $T \subseteq X$ we have that

$$\mu^*(T) = \mu^*(T \cap A) + \mu^*(T \cap A^c).$$

(Informally, A separates arbitrary subsets of X well.)

To check measurability of A, it suffices to check that $\mu^*(T) \ge \mu^*(T \cap A) + \mu^*(T \cap A^c)$ for all T of finite outer measure. Measurable sets possess *excision*, i.e. if A is a measurable subset of finite outer measure and B is any subset containing A then $\mu^*(B \setminus A) = \mu^*(B) - \mu^*(A)$. Indeed, because A is measurable we have

$$\mu^*(B) = \mu^*(B \cap A) + \mu^*(B \cap A^c) = \mu^*(A) + \mu^*(B \setminus A).$$

Theorem 2.0.3 (Carathéodory's Theorem)**.** Let X be a set and $\mu^* : 2^X \to [0, \infty]$ an outer measure on X. Then the collection $\mathcal{M}$ of μ^*-measurable subsets of X forms a σ-algebra and $(X, \mathcal{M}, \mu^*|_{\mathcal{M}})$ is a complete measure space.

Proof. First, we show that $\mathcal{M}$ is a σ-algebra. First note that for any subset B, we have that $\mu^*(B) = 0 + \mu^*(B) = \mu^*(\emptyset \cap B) + \mu^*(X \cap B)$, so that $\emptyset \in \mathcal{M}$. The definition of μ^*-measurability is symmetric in A and A^c, so the second condition is trivial. Next we show that $\mathcal{M}$ is closed under the formation of finite unions, so assume that $A_1, A_2 \in \mathcal{M}$; we have to show that $A_1 \cup A_2 \in \mathcal{M}$, for which let $T \subseteq X$. Then μ^*-measurability of A_2 implies that

$$\mu^*(T \cap (A_1 \cup A_2)) = \mu^*(T \cap (A_1 \cup A_2) \cap A_1) + \mu^*(T \cap (A_1 \cup A_2) \cap A_1^c) = \mu^*(T \cap A_1) + \mu^*(T \cap (A_2 \setminus A_1)).$$

Therefore,

$$\begin{aligned} \mu^*(T \cap (A_1 \cup A_2)) + \mu^*(T \cap (A_1 \cup A_2)^c)) &= \mu^*(T \cap A_1) + \mu^*(T \cap A_1^c \cap A_2) + \mu^*(T \cap A_1^c \cap A_2^c) \\ &= \mu^*(T \cap A_1) + \mu^*(T \cap A_1^c) \\ &= \mu^*(T). \end{aligned}$$

Next, suppose that A_i is a countable collection of μ^*-measurable sets; we need to show that $\bigcup_i A_i$ is μ^*-measurable. By replacing A_i by $A_i' = A_i \setminus (\bigcup_{j<i} A_j)$, we can assume WLOG that the A_i are disjoint. In this case, it follows by induction induction that for any arbitrary $T \subseteq X$ we have that

$$\mu^*(T) = \sum_{i=1}^{n} \mu^*(T \cap A_i) + \mu^*(T \cap (\cap_{i=1}^n A_i^c)),$$

where the key step is that $\mu^*(T \cap (\cap_{i=1}^n A_i^c) = \mu^*(T \cap (\cap_{i=1}^n A_i^c) \cap A_{n+1}) + \mu^*(T \cap (\cap_{i=1}^{n+1} A_i^c))$, and the first term is $\mu^*(T \cap A_{n+1})$ by the assumption on disjointness. The right hand side is at least as big as $\sum_{i=1}^n \mu^*(T \cap A_i) + \mu^*(T \cap (\bigcap_i A_i^c))$ for any n, and so we conclude that

$$\mu^*(T) \ge \sum_i \mu^*(T \cap A_i) + \mu^*(T \cap (\cup_i A_i)^c) \ge \mu^*(T \cap (\cup_i A_i)) + \mu^*(T \cap (\cup_i A_i)^c) \ge \mu^*(T),$$

completing the proof that $\mathcal{M}$ is a σ-algebra. Finally, to show the restriction of μ^* to $\mathcal{M}$ is a measure, we need to show countable additivity; this follows from taking $T = \bigcup_i A_i$ above to get

$$\mu^*(\cup_i A_i) \ge \sum_i \mu^*(A_i) + \mu^*(\emptyset) \ge \mu^*(\cup_i A_i) + 0.$$

Finally, to show completeness, we will show the stronger statement that any subset of Lebesgue outer measure 0 is Lebesgue measurable; this suffices. Indeed, let $A \subseteq X$ has $\mu^*(A) = 0$; then for any $T \subseteq X$ we have that $\mu^*(T \cap A) = 0$ as well, so that automatically $\mu^*(T) \ge \mu^*(T \cap A^c)$ as needed. ∎

The most important example of this is the Lebesgue outer measure on $\mathbf{R}^d$.

Definition 2.0.4. For each $A \subseteq \mathbf{R}^d$, define its *Lebesgue outer measure*

$$\mu^*(A) := \inf \left\{ \sum_i \operatorname{vol}(B_i) : A \subseteq \bigcup_i B_i \right\},$$

where the sum is over all countable collections of open boxes B_i which cover A.

Lemma 2.0.5 (Elementary Properties of Lebesgue Outer Measure)**.**

(a) This is an outer measure which assigns to every box in $\mathbf{R}^d$ its volume.
(b) If $C \subset \mathbf{R}^d$ is countable, then $\mu^*(C) = 0$.
(c) If $A \subseteq \mathbf{R}^d$ is any subset, then $\mu^*(A) = \inf_{U \supseteq A}\{\mu^*(U)\}$ where the infimum is over all open $U \supseteq A$.
(d) If $A, B \subseteq \mathbf{R}^d$ are subsets with the distance $d(A, B) > 0$, then $\mu^*(A \cup B) = \mu^*(A) + \mu^*(B)$.
(e) If a set $A \subseteq \mathbf{R}^d$ is a countable union of almost disjoint boxes B_i then $\mu^*(A) = \sum_i \operatorname{vol}(B_i)$.

Proof. For (a), first note that $\mu^*(\emptyset) = 0$ since the empty set is covered by the empty collection. For the second condition, note that it clearly holds if $\sum_i \mu^*(A_i) = \infty$, so assume that $\sum_i \mu^*(A_i) < \infty$; in particular, $\mu^*(A_i) < \infty$ for all i. Given any $\varepsilon > 0$, for each $i \geq 0$, pick a countable collection $\{B_{ij}\}_j$ of open boxes covering A_i such that $\sum_j \operatorname{vol}(B_{ij}) < \mu^*(A_i) + 2^{-i-1}\varepsilon$. Then the collection $\{B_{ij}\}_{i,j}$ is a countable collection of boxes covering A with $\sum_{i,j} \operatorname{vol}(B_{ij}) < \sum_i(\mu^*(A_i) + 2^{-i-1}\varepsilon) = \sum_i \mu^*(A_i) + \varepsilon$; this is true for each $\varepsilon > 0$, so it follows that $\mu^*(A) \leq \sum_i \mu^*(A_i)$.

To show that this outer measure assigns to each box $B \subseteq \mathbf{R}^d$ its volume, assume first that B is bounded; then it is trivial to find an open box $B_\varepsilon^+ \supseteq B$ with $\operatorname{vol}(B_\varepsilon^+) \leq \operatorname{vol}(B) + \varepsilon$ for any $\varepsilon > 0$; it follows that $\mu^*(B) \leq \operatorname{vol}(B)$. For the other direction, by monotonicity, it suffices to assume B is closed; indeed, in general, we can find a closed box $B_\varepsilon^- \subseteq B$ with $\operatorname{vol}(B) - \varepsilon \leq \operatorname{vol}(B_\varepsilon^-)$, in which case it would follow that

$$\operatorname{vol}(B) - \varepsilon \leq \operatorname{vol}(B_\varepsilon^-) = \mu^*(B_\varepsilon^-) \leq \mu^*(B).$$

Since this would be true for any $\varepsilon > 0$, this would imply that $\operatorname{vol}(B) \leq \mu^*(B)$ as needed. Therefore, assume that $B \subset \mathbf{R}^d$ is a closed bounded box, and suppose that $\{B_i\}$ is a countable cover of B by open boxes. By compactness of B, a finite number of these cover B, say the first n for some $n \geq 1$. Then B can be decomposed into a finite collection S_j of subboxes that overlap only along their boundaries and such that for each j the interior of S_j is contained in some B_j for $j \leq n$; then it follows that $\operatorname{vol}(B) = \sum_j \operatorname{vol}(S_j) \leq \sum_{i=1}^n \operatorname{vol}(B_i) \leq \sum_i \operatorname{vol}(B_i)$, where the first inequality follows from grouping the S_j according to what B_i they lie in. This implies that $\operatorname{vol}(B) \leq \mu^*(B)$ and hence that $\mu^*(B) = \operatorname{vol}(B)$ for all bounded boxes $B \subset \mathbf{R}^d$. Now if B is any unbounded box, then it contains a bounded subbox B_N with $\operatorname{vol}(B_N) \geq N$ for any $N \in \mathbf{R}$; this implies from $N \leq \operatorname{vol}(B_N) = \mu^*(B_N) \leq \mu^*(B)$ for any $N \in \mathbf{R}$ that $\mu^*(B) = \infty = \operatorname{vol}(B)$, so the result holds for unbounded boxes as well.

For (b), enumerate $C = \{c_i\}$. For any $\varepsilon > 0$, pick a box B_i around c_i with $\operatorname{vol}(B_i) \leq 2^{-i-1}\varepsilon$. Then $\mu^*(C) \leq \sum_i \operatorname{vol}(B_i) \leq \varepsilon$; since this holds for any $\varepsilon > 0$ we must have $\mu^*(C) = 0$.

For (c), note that $\mu^*(A) \leq \inf$ is clear by monotonicity. For the converse, let $\varepsilon > 0$ be given and let B_i be a countable collection of open boxes covering A such that $\sum_i \operatorname{vol}(B_i) < \mu^*(A) + \varepsilon$. Then if $U = \bigcup_i B_i$, then $\mu^*(U) \leq \sum_i \mu^*(B_i) = \sum_i \operatorname{vol}(B_i) < \mu^*(A) + \varepsilon$, so that $\inf \leq \mu^*(A)$ as well.

For (d), we know first that $\mu^*(A \cup B) \leq \mu^*(A) + \mu^*(B)$. For the opposite inequality, pick a ρ with $0 < \rho < d(A, B)$. Next, given any $\varepsilon > 0$, pick a covering of $A \cup B$ by open boxes B_i such that $\sum_i \operatorname{vol} B_i \leq \mu^*(A \cup B) + \varepsilon$; by subdividing, assume WLOG that each diameter $\operatorname{diam} B_i < \rho$. Then each box B_i can intersect at most one of A and B. If $J_A = \{i : B_i \cap A \neq \emptyset\}$ and similarly J_B, then $J_A \cap J_B = \emptyset$ and $A \subseteq \bigcup_{i \in J_A} \operatorname{vol} B_i$ and similarly for B. Therefore, $\mu^*(A) + \mu^*(B) \leq \sum_{i \in J_A} \mu^*(B_i) + \sum_{i \in J_B} \mu^* B_i \leq \sum_i \mu^*(B_i) = \sum_i \operatorname{vol} B_i \leq \mu^*(A \cup B) + \varepsilon$; since this holds for every $\varepsilon > 0$, we are done.

For (e), given any $\varepsilon > 0$ pick for each i an open box $B_i' \subset B_i$ such that $\operatorname{vol} B_i' \leq \operatorname{vol} B_i + \varepsilon 2^{-i}$. Then for every N the boxes B_i' for $i = 1, \ldots, N$ are at positive distances from each other and hence by (d) we get $\mu^*(A) \supseteq \mu^*\left(\bigcup_{i=1}^N B_i'\right) = \sum_{i=1}^N \operatorname{vol} B_i' \geq \sum_{i=1}^N (\operatorname{vol} B_i - \varepsilon 2^{-i}) \geq \sum_{i=1}^N \operatorname{vol} B_i - \varepsilon$. Therefore, $\sum_{i=1}^N \operatorname{vol} B_i \leq \mu^*(A) + \varepsilon$. Take $\lim_{N\to\infty}$ to conclude that $\sum_i \operatorname{vol} B_i \leq \mu^*(A) + \varepsilon$ for all ε and hence $\sum_i \operatorname{vol} B_i \leq \mu^*(A)$ as needed. ∎

This provides the construction of the Lebesgue measurable on $\mathbf{R}^d$ using the Lebesgue outer measure. The Lebesgue outer measure is clearly translation invariant, so so is the Lebesgue measure. Finally, we need to show that every Borel set is Lebesgue measurable; having done this, the existence part in the statement of the Lebesgue measure will be proven.

Lemma 2.0.6 (Borel Subsets)**.** Every Borel subset of $\mathbf{R}^d$ is Lebesgue measurable.

Proof. We saw above that $\mathcal{B}(\mathbf{R}^d)$ is generated by the half spaces, so it suffices to show that each half-space $H = H_{j,b} := \{x_j \le b\} \subseteq \mathbf{R}^d$ satisfies that for all $T \subset \mathbf{R}^d$ of finite outer measure we have $\mu^*(T) \ge \mu^*(T \cap H) + \mu^*(T \cap H^c)$. Given any $\varepsilon > 0$, let B_i be a countable collection of bounded open boxes such that $\sum_i \operatorname{vol}(B_i) < \mu^*(T) + \varepsilon$. Then $B_i \cap H$ is a countable collection of bounded boxes covering $T \cap H$, so by condition (b) in the definition of an outer measure and Lemma **??**, we conclude that $\sum_i \operatorname{vol}(B_i \cap H) = \sum_i \mu^*(B_i \cap H) \ge \mu^*(T \cap H)$. The same holds for H^c, and so we conclude finally that

$$\mu^*(T \cap H) + \mu^*(T \cap H^c) \le \sum_i \operatorname{vol}(B_i \cap H) + \sum_i \operatorname{vol}(B_i \cap H^c) = \sum_i \operatorname{vol}(B_i) < \mu^*(T) + \varepsilon.$$

This holds for any $\varepsilon > 0$, and so we conclude that $\mu^*(T \cap H) + \mu^*(T \cap H^c) \le \mu^*(T)$. ■

Lemma 2.0.7 (Characterization of Lebesgue Measurable Subsets by Inner and Outer Approximations)**.** Let $A \subseteq \mathbf{R}^d$ be any subset. Then TFAE:

(a) The set A is Lebesgue measurable, i.e. satisfies Carathéodory's criterion for the Lebesgue outer measure.
(b) For each $\varepsilon > 0$, there is an open set $U \supseteq A$ such that $\mu^*(U \setminus A) < \varepsilon$.
(c) There is a G_δ set $G \supseteq A$ such that $\mu^*(G \setminus A) = 0$.
(d) For each $\varepsilon > 0$, there is a closed set $C \subseteq A$ such that $\mu^*(A \setminus C) < \varepsilon$.
(e) There is an F_σ set $F \subseteq A$ such that $\mu^*(A \setminus F) = 0$.

Further, in this case, if $\mu(A) < \infty$, then

(f) the C in (d) can be chosen to be compact, and in fact
(g) for each $\varepsilon > 0$, there is a finite collection of open boxes B_i such that if $U = \bigcup_i B_i$ then $\mu(U \Delta A) < \varepsilon$.

Proof. To show the equivalence of these statements, by the self-duality of (a) with respect to complements, it suffices to show the equivalence of (a), (b), and (c). For (a) $\Rightarrow$ (b), let A be measurable and let $\varepsilon > 0$ be given. First suppose that $\mu^*(A) < \infty$. By definition of $\mu^*(A)$, we may find a countable open cover of A by boxes B_i such that $\sum_i \operatorname{vol} B_i < \mu^*(A) + \varepsilon$. Let $U := \bigcup_i B_i$. Then $\mu^*(U) \le \sum_i \mu^*(B_i) = \sum_i \operatorname{vol}(B_i) < \mu^*(A) + \varepsilon$ so that by excision we have $\mu^*(U \setminus A) = \mu^*(U) - \mu^*(A) < \varepsilon$ as needed. Now suppose $\mu^*(A) = \infty$. Then A is the disjoint union of a countable collection A_k of measurable subsets of finite outer measure; indeed, let $A_k := A \cap \mathbf{D}(0,k) \setminus \mathbf{D}(0,k-1)$, where we're using the previous lemma. By the finite measure case, for each k there is an open subset $U_k \supseteq A_k$ with $\mu^*(U_k \setminus A_k) < \varepsilon/2^k$. Then set $U = \bigcup_k U_k$ and then $U \setminus A = \bigcup_k (U_k \setminus A_k)$ and hence $\mu^*(U \setminus A) \le \sum_k \mu^*(U_k \setminus A_k) < \sum_k \varepsilon/2^k = \varepsilon$. For (b) $\Rightarrow$ (c), for each integer $k \ge 1$ find an open subset $U_k \supseteq A$ such that $\mu^*(U_k \setminus A) < 1/k$ and let $G := \bigcap_k U_k$. Since $G \subseteq U_k$ for all k monotonicity tells us that $\mu^*(G \setminus A) \le \mu^*(U_k \setminus A) < 1/k$ for all $k \ge 1$ and hence $\mu^*(G \setminus A) = 0$. Finally, for (c) $\Rightarrow$ (a), we know that G is measurable by the previous lemma and $G \setminus A$ is measurable by completeness of the Lebesgue measure. Therefore, it follows that $A = G \cap (G \setminus A)^c$ is also measurable.

To show (f), first by (d) pick a closed C such that $\mu^*(A \setminus C) \le \varepsilon/2$. For each $n \ge 1$, let $K_n := C \cap \mathbf{D}(0,n)$. Then $A \setminus K_n$ is a sequence of measurable subsets that decreases to $A \setminus C$ and hence since $\mu(A) < \infty$ we have by Lemma 1.1.7(d) that $\mu(A \setminus K_n) < \varepsilon$ for large $n \gg 1$. For (g), let B_i be a countable collection of open boxes such that $\sum_i \operatorname{vol} B_i < \mu(A) + \varepsilon/2$; then necessarily the series on the left converges. Pick an $N \gg 1$ such that $\sum_{i>N} \operatorname{vol} B_i < \varepsilon/2$ and set $U := \bigcup_{i=1}^N B_i$. Then $\mu(U \Delta A) = \mu(U \setminus A) + \mu(A \setminus U) \le \mu\left(\bigcup_i B_i \setminus E\right) + \mu(\bigcup_{i>N} B_i) \le \varepsilon/2 + \varepsilon/2 = \varepsilon$.

■

The collection $\mathcal{M}$ of subsets of $\mathbf{R}^d$ so produced is called the collection of Lebesgue measurable subsets. The last statement in the proof also shows that any countable subset of $\mathbf{R}^d$ is Lebesgue measurable and has Lebesgue measure 0.

Finally, it follows immediately that the Vitali set constructed in Theorem 1.1.4 is not Lebesgue measurable. In fact, non-measurable subsets are quite common, and the following proof is essentially identical to that of Theorem 1.1.4.

Theorem 2.0.8 (Generalized Vitali Theorem)**.** Any subset $A \subseteq \mathbf{R}$ of positive outer measure contains a nonmeasurable subset.

Proof. By countable subadditivity, we have that $\mu^*(A) \le \sum_i \mu^*(A \cap [-i,i])$, so that if $\mu^*(A) > 0$ then so must $\mu^*(A \cap [-i,i])$ for some i; therefore, replacing A by this $A \cap [-i,i]$, we may assume WLOG that A is bounded.

Again, partition A into rational equivalence classes, and let $V \subset A$ be a set of representatives. Because A is bounded, it lies in $[-n, n]$ for some $n \geq 1$. If $(q_i)_i$ is an enumeration of $[-2n, 2n] \cap \mathbf{Q}$, define $V_i := V + q_i$. By definition, these are disjoint and we have $A \subseteq \coprod_i V_i \subseteq [-3n, 3n]$ and we get the same contradiction as before.

∎

3 Aside: The Cantor Set and the Devil's Staircase

[Royden and Fitzpatrick.] We showed in the previous section that any countable set has measure zero and that any Borel set is Lebesgue measurable. We can ask whether the converse of these is true. We compile a list of similar questions.

Q1. If a subset of $\mathbf{R}^d$ has Lebesgue measure 0, is it necessarily countable?
Q2. If a subset of $\mathbf{R}^d$ is Lebesgue measurable, is it necessarily Borel?
Q3. Can a continuous function map a set of Lebesgue measure 0 to a set of positive Lebesgue measure?
Q4. Can a continuous function map a measurable set to a nonmeasurable set?

We will show below that the answers to Q1 and Q2 are in the negative, while that to Q3 and Q4 are in the positive. For the first one, we construct the *Cantor set*. First define $C_0 := [0, 1]$. Then remove the middle third to get $C_1 := [0, 1/3] \cup [2/3, 1]$. Then continue to get $C_2 := [0, 1/9] \cup [2/9, 1/3] \cup [2/3, 7/9] \cup [8/9, 1]$ and so on. In this way $(C_k)_{k\geq 0}$ is a descending collection of nonempty closed compact sets, and for each k, the set C_k is a disjoint union of 2^k closed intervals each of length $(1/3)^k$. We define the *Cantor set* $C := \bigcap_{k\geq 0} C_k$.

Theorem 3.0.1 (Cantor)**.** The Cantor set C is compact, uncountable, and of measure zero.

Proof. First note that C is an intersection of a descending collection of nonempty compact sets and is therefore itself nonempty and compact; in particular, it is closed and so certainly Borel and hence Lebesgue measurable. Further, we have that $C \subseteq C_k$ for any $k \geq 0$ so that $\mu(C) \leq \mu(C_k) = (2/3)^k$ for any $k \geq 0$, so $\mu(C) = 0$. Next, suppose that C was countable, and pick an enumeration $C = \{c_i\}_{i\geq 1}$ of C. One of the two closed intervals making up C_1 fails to contain c_1; let F_1 be this interval. Inductively, proceed to construct a descending chain F_i of compact intervals such that $F_i \subset C_i$ and $c_i \notin F_i$. Then by compactness, $\bigcap_i F_i$ is nonempty and contained C, but it cannot contain any of the c_i's for $i \geq 1$; this is a contradiction. ■

Now we will construct the Cantor-Lebesgue function, which is a continuous, nondecreasing function $\varphi : [0, 1] \to [0, 1]$ which has the property that despite $\varphi(1) > \varphi(0)$ its derivative exists and is zero on a set of measure 1. For this reason, its graph is also called the Devil's staircase. For each $k \geq 0$, define $U_k := [0, 1] \setminus C_k$ and define $U := \bigcup_k U_k = [0, 1] \setminus C$. Note that the previous theorem implies that $1 = \mu([0, 1]) = \mu(C) + \mu(U) \Rightarrow \mu(U) = 1$. Fix a $k \geq 0$. Define φ on U_k to be the unique nonstrictly increasing function on U_k which is constant on each of the $2^k - 1$ open intervals say $U_{k,1}, \dots, U_{k,2^k-1}$ and takes the value $i/2^k$ on $U_{k,i}$ for $1 \leq i \leq 2^k - 1$. These glue together to give a function φ on U. We now extend φ to a function on $[0, 1]$ by $\varphi(0) := 0$ and $\varphi(x) := \sup_{t\in U\cap[0,x)} \varphi(t)$ for $x \in C \setminus \{0\}$. Here's another way to think about φ: any $x \in C$ can be written uniquely as $\sum_{n=1}^{\infty} 2x_n/3^n$ for $x_n \in \{0, 1\}$; then $\varphi(x) = \sum_{n=1}^{\infty} x_n/2^n$. From this description, it is not very clear (at least not to me) that φ is continuous.

Theorem 3.0.2 (Cantor-Lebesgue)**.** The Cantor-Lebesgue function $\varphi : [0, 1] \to [0, 1]$ is a nondecreasing continuous surjective function. It is differentiable on the open set U, the complement of the Cantor set, where its derivative is identically 0; i.e. it is differentiable with derivative zero on a set of measure 1.

Proof. We start by showing that φ is nonstrictly increasing; for this, assume that $x, y \in [0, 1]$ and $x < y$. We have to show that $\varphi(x) \leq \varphi(y)$. This is clear when $x = 0$ or when x, y are both in U. When $x \in C$ but $y \notin C$, then for all $t \in U \cap [0, x)$ we have $\varphi(t) \leq \varphi(y)$, so the supremum $\varphi(x) = \sup_{t\in U\cap[0,x)} \varphi(t) \leq \varphi(y)$. When $x \notin C$ but $y \in C$, then $x \in U \cap [0, y)$ so that $\varphi(x) \leq \sup_{t\in U\cap[0,y)} \varphi(t) = \varphi(y)$. When $x, y \in C$, then we may pick a $k \geq 0$ sufficiently large so that $y - x > 3^{-k}$; then then $\exists\, t \in U_k \cap (x, y)$ so by the previous cases we have $\varphi(x) \leq \varphi(t) \leq \varphi(y)$.

Next, note that clearly $\varphi : [0, 1] \to [0, 1]$ and $\varphi(1) = 1$, which follows from $\varphi(1) \geq (2^k - 1)/2^k$ for all $k \geq 0$. Now, we show that φ is continuous; it would then follow from the Intermediate Value Theorem that φ is surjective. First, we show continuity at $x = 0$, so let $\varepsilon > 0$ be given. Pick a $k \geq 0$ such that $2^{-k} < \varepsilon$; then for all $x \in [0, 3^{-k})$ taking any $t \in U_{k,1}$ shows that $x \leq t$ and so $\varphi(x) \leq \varphi(t) = 2^{-k} < \varepsilon$. A similar argument establishes continuity at $x = 1$. Next, continuity is clear on U_k for any $k \geq 0$ and hence also on U. Finally, we have to show continuity at an arbitrary $x \in C \setminus \{0, 1\}$. Since φ is an increasing function, to show continuity at x, it suffices to produce for any $\varepsilon > 0$ two numbers $a_\varepsilon, b_\varepsilon \in [0, 1]$ such that $x \in (a_\varepsilon, b_\varepsilon)$ and $\varphi(b_\varepsilon) - \varphi(a_\varepsilon) < \varepsilon$. Pick an $\varepsilon > 0$ and pick a $k \geq 2$ such that both $2^{-k} < \varepsilon$ and that for this k, there is a unique i with $1 \leq i \leq 2^k - 2$ and such that x lies between $U_{k,i}$ and $U_{k,i+1}$. Then picking any $a_\varepsilon \in U_{k,i}$ and $b_\varepsilon \in U_{k,i+1}$ shows that $x \in (a_\varepsilon, b_\varepsilon)$ and $\varphi(b_\varepsilon) - \varphi(a_\varepsilon) = 2^{-k} < \varepsilon$ as needed. The last statement is clear, since every point in U lies in U_k for some $k \geq 0$.

■

Now we use this to answer questions 3 and 4.

Theorem 3.0.3. Let φ be the Cantor-Lebesgue function, and define $\psi : [0,1] \to [0,2]$ by $\psi(x) = \varphi(x) + x$. Then ψ is a strictly increasing surjective continuous function. Further,

(a) the set $\psi(C) \subset [0,2]$ is measurable of positive measure,
(b) there is a subset $J \subset C$ of measure zero such that $\psi(J)$ is nonmeasurable. In particular, J is measurable but not Borel.

Proof. For this we need a simple lemma:

Lemma 3.0.4. A strictly increasing surjective continuous function $[0,1] \to [0,2]$ is a homeomorphism; in particular, it has a continuous (and so measurable) inverse.

Proof. Strictly increasing implies that it is injective as well; a bijective continuous map from a compact space to a Hausdorff space is closed and hence a homeomorphism. Alternatively, the continuity of the inverse is easy to verify by hand. ■

By the previous theorem, the function ψ is clearly strictly increasing and continuous with the property that $\psi(0) = 0$ and $\psi(1) = 2$; by the IVT, it is also surjective. From the above, ψ is a homeomorphism; therefore the disjoint decomposition $[0,1] = C \coprod U$ gives the decomposition $[0,2] = \psi(C) \coprod \psi(U)$ with $\psi(C)$ closed and $\psi(U)$ open; to show (a), we we will show that $\mu(\psi(U)) = 1$. This would imply that $\mu(\psi(C)) = 1$ as well. To show this, let $(I_j)_j$ be an enumeration (in any manner) of the intervals removed to create the Cantor set, so $U = \bigcup_j I_j$. Since φ is constant on each I_j, the map ψ maps I_j to a translated copy of itself of the same length. By injectivity of ψ, the collection $(\psi(I_j))_j$ is disjoint, and so by countable additivity of measure we conclude that $\mu(\psi(U)) = \sum_j \ell(\psi(I_j)) = \sum_j \ell(I_j) = \mu(U) = 1$. To show (b), note that the generalized Vitali Theorem 1.1.4 tells us that $\psi(C)$ contains a nonmeasurable subset K, so if we take $J := \psi^{-1}(K)$, then J is measurable of measure zero but $\psi(J) = K$ is nonmeasurable. In particular, J cannot be Borel because if it were, then by continuity of ψ^{-1} would imply that so is K, which is not the case. ■

4 Regularity and Uniqueness of the Lebesgue Measure

5 The Lebesgue Integral

5.1 The Construction

Let $(X, \mathcal{M})$ be a measurable space. Let $\mathcal{C} = \mathcal{M}(X, [0, \infty])$ be the collection of measurable functions $f : X \to [0, \infty]$, and let $\mathcal{S} \subseteq \mathcal{C}$ be the collection of simple measurable functions $f : X \to [0, \infty)$, i.e. functions that can be written as $f = \sum_{i=1}^{n} \alpha_j \mathbb{1}_{A_i}$ for some $\alpha_i \in [0, \infty)$ and measurable $A_i \subseteq X$ with each $\mu(A_i) < \infty$.

Lemma 5.1.1. There is a unique map $\int - \mathrm{d}\mu : \mathcal{S} \to [0, \infty)$ with the following properties (a) and (b). It is called *integration with respect to* μ.

(a) (Linearity) For any $f, g \in \mathcal{S}$ and $\alpha, \beta \in [0, \infty)$, we have that $\int(\alpha f + \beta g)\mathrm{d}\mu = \alpha \int f \mathrm{d}\mu + \beta \int g \mathrm{d}\mu$.
(b) (Normalization) If $A \subseteq X$ is in $\mathcal{M}$, then $\int \mathbb{1}_A \mathrm{d}\mu = \mu(A)$.

This map further satisfies the following properties.

(c) (Monotonicity) If $f \leq g$ everwhere on X, then $\int f \mathrm{d}\mu \leq \int g \mathrm{d}\mu$.
(d) If f_n is a nondecreasing sequence of functions in $\mathcal{S}$ with $f = \lim f_n \in \mathcal{S}$, then $\int f \mathrm{d}\mu = \lim_n \int f_n \mathrm{d}\mu$.
(e) If $A \subseteq X$ is measurable, then $\int_A f|_A \mathrm{d}\mu|_A = \int_X f \cdot \mathbb{1}_A \mathrm{d}\mu$.[3]

Proof. Any $f \in \mathcal{S}$ can be written as $f = \sum_{i=1}^{n} \alpha_i \mathbb{1}_{A_i}$ for some $\alpha_i \in [0, \infty)$ and A_i measurable subsets of X with $\mu(A_i) < \infty$; if such a map exists, by properties (a) and (b), it must be given by $\int f \, \mathrm{d}\mu = \sum_{i=1}^{n} \alpha_i \mu(A_i)$. Conversely, we show that this definition works. First, we have to show that it is independent of representation in this form; so suppose that $f = \sum_{i=1}^{n} \alpha_i \mathbb{1}_{A_i} = \sum_{j=1}^{m} \beta_j \mathbb{1}_{B_j}$, then we may first assume by eliminating the i for which $\alpha_i = 0$ and j for which $\beta_j = 0$ that $\bigcup_{i=1}^{n} A_i = f^{-1}(0, \infty) = \bigcup_{j=1}^{m} B_j$. Then the additivity of μ and the fact that $\alpha_i = \beta_j$ if $A_i \cap B_j \neq \emptyset$ implies that

$$\sum_i \alpha_i \mu(A_i) = \sum_{i,j} \alpha_i \mu(A_i \cap B_j) = \sum_{i,j} \beta_j \mu(A_i \cap B_j) = \sum_j \beta_j \mu(B_j).$$

Therefore, this operation is well-defined. Now suppose that $f = \sum_{i=1}^{n} \alpha_i \mathbb{1}_{A_i}$ and $g = \sum_{j=1}^{n} \beta_j \mathbb{1}_{B_j}$; by adding zeros this time, we may assume that $\bigcup_i A_i = \bigcup_j B_j$. Then $\alpha f + \beta g = \sum_{i,j} (\alpha \alpha_i + \beta \beta_j) \mathbb{1}_{A_i \cap B_j}$ so that

$$\begin{aligned}
\int (\alpha f + \beta g) \mathrm{d}\mu &= \sum_{i,j} (\alpha \alpha_i + \beta \beta_j) \mu(A_i \cap B_j) \\
&= \alpha \sum_i \alpha_i \sum_j \mu(A_i \cap B_j) + \beta \sum_j \beta_j \sum_i \mu(A_i \cap B_j) \\
&= \alpha \sum_i \alpha_i \mu(A_i) + \beta \sum_j \beta_j \mu(B_j) = \alpha \int f \mathrm{d}\mu + \beta \int g \mathrm{d}\mu.
\end{aligned}$$

Therefore, this definition satisfies (a), and that it satisfies (b) is evident. For (c), if $f, g \in \mathcal{S}$ are such that $f \leq g$ on X, then $g - f \in \mathcal{S}$ as well; so (a) gives $\int g \mathrm{d}\mu = \int (f + g - f) \mathrm{d}\mu = \int f \mathrm{d}\mu + \int (g - f) \mathrm{d}\mu \geq \int f \mathrm{d}\mu$. For (d), it follows from (c) that $\int f_n \mathrm{d}\mu$ is an increasing sequence in $[0, \infty)$ bounded above by $\int f \mathrm{d}\mu$; therefore, $\lim_n \int f_n \mathrm{d}\mu \leq \int f \mathrm{d}\mu$. For each $\varepsilon \in (0, 1)$, we will construct a nondecreasing sequence f_n^ε of functions in $\mathcal{S}$ such that $f_n^\varepsilon \leq f_n$ for all n and such that $\lim_n \int f_n^\varepsilon \mathrm{d}\mu = (1 - \varepsilon) \int f \mathrm{d}\mu$; this would show that $(1 - \varepsilon) \int f \mathrm{d}\mu = \lim_n \int f_n^\varepsilon \mathrm{d}\mu \leq \lim_n \int f_n \mathrm{d}\mu$ for all $\varepsilon \in (0, 1)$, and hence complete the proof. Indeed, write $f = \sum_{i=1}^{n} \alpha_i \mathbb{1}_{A_i}$ for some $\alpha \in [0, \infty)$ and *disjoint* measurable A_i with $\mu(A_i) < \infty$. For each n and i, let $A_{i,n} := \{x \in A_i : f_n(x) \geq (1 - \varepsilon)\alpha_i\}$ and let $f_n^\varepsilon := \sum_{i=1}^{n} (1 - \varepsilon) a_i \mathbb{1}_{A_{i,n}}$; for the last step use Lemma 1.1.7(c) to conclude that $\lim_n \mu(A_{i,n}) = \mu(A_i)$. The statement in (e) is basically tautological: if $f = \sum_i \alpha_i \mathbb{1}_{A_i}$, then $f|_A = \sum_i \alpha_i \mathbb{1}_{A_i \cap A}$, so that

$$\int_A f|_A \mathrm{d}\mu|_A = \sum_i \alpha_i \mu|_A (A_i \cap A) = \sum_i \alpha_i \mu(A_i \cap A) = \int \sum_i \alpha_i \mathbb{1}_{A_i \cap A} \mathrm{d}\mu = \int f \mathbb{1}_A \mathrm{d}\mu.$$

■

We are now ready to proceed with our construction of the integral of nonnegative measurable functions. Having done this, we will define a function to be integrable if $\int f \mathrm{d}\mu < \infty$. For general (extended) real valued functions, we deal with the positives and negatives separately. Finally, we will use this to define $\mathcal{L}^1(\mu)$.

[3] In light of this, we will write the first as simply $\int_A f \mathrm{d}\mu$, and this will mean $\int f \cdot \mathbb{1}_A \mathrm{d}\mu$.

Theorem 5.1.2 (Integration of Nonnegative Measurable Functions)**.** There is a unique map $\int -\mathrm{d}\mu : \mathcal{C} \to [0,\infty]$ extending the map $\mathcal{S} \to [0,\infty]$ constructed above with the following property:

If (f_n) is any nondecreasing sequence of functions in $\mathcal{C}$ with $f = \lim_n f_n \in \mathcal{C}$, then $\int f\mathrm{d}\mu = \lim_n \int f_n\mathrm{d}\mu$.

This map further satisfies the following properties:

(a) (Monotonicity) If $f \le g$ everywhere on X, then $\int f\mathrm{d}\mu \le \int g\mathrm{d}\mu$.
(b) (Linearity) For any $f, g \in \mathcal{C}$ and $\alpha, \beta \in [0,\infty)$, we have that $\int(\alpha f + \beta g)\mathrm{d}\mu = \alpha \int f\mathrm{d}\mu + \beta \int g\mathrm{d}\mu$.
(c) (Beppo Levi's Theorem) If (f_n) is a series of functions in $\mathcal{C}$, then $\int \sum_n f_n\mathrm{d}\mu = \sum_n \int f_n\mathrm{d}\mu$.
(d) If f, g are two functions that agree $[\mu\text{–a.e.}]$, then $\int f\mathrm{d}\mu = \int g\mathrm{d}\mu$.
(e) If $A \subseteq X$ is measurable, then $\int_A f|_A\mathrm{d}\mu|_A = \int f\mathbb{1}_A\mathrm{d}\mu =: \int_A f\mathrm{d}\mu$.
(f) (Monotonicity over Domain) If $A \subseteq B$, then $\int_A f\mathrm{d}\mu \le \int_B f\mathrm{d}\mu$. If $\mu(A) = 0$, then $\int_A f\mathrm{d}\mu = 0$. In particular, $\int_X f\mathrm{d}\mu = \int_{X\setminus A} f\mathrm{d}\mu$.
(g) (Countable Additivity) If A_i is a disjoint sequence of measurable subsets of X, then $\int_{\bigcup_i A_i} f\mathrm{d}\mu = \sum_i \int_{A_i} f\mathrm{d}\mu$.
(h) (Chebyshev's Inequality) Fix an $f \in \mathcal{C}$ and $t > 0$. Then $t\mu(f^{-1}[t,\infty]) \le \int_{f^{-1}[t,\infty]} f\mathrm{d}t \le \int f\mathrm{d}\mu$.
(i) If $\int f\mathrm{d}\mu = 0$, then $f = 0$ almost everywhere. If $\int f\mathrm{d}\mu < \infty$, then $f < \infty$ almost everywhere.

Proof. If such a function is to exist, then it is unique: if f is any function, by Theorem 1.2.3, there is a nondecreasing sequence (f_n) of functions in $\mathcal{S}$ with $f = \lim_n f_n$, so that $\int f\mathrm{d}\mu = \lim_n \int f_n\mathrm{d}\mu$, where the integral on the right side is as defined above. Conversely, we use this to define $\int f\mathrm{d}\mu$; then, we must show that it is independent of the choice of f_n. To accomplish that, we will show that this definition equals $\sup\left\{\int g\mathrm{d}\mu : g \in \mathcal{S}, g \le f\right\}$, and this is clearly independent of the choice of f_n. Now certainly $\lim_n \int f_n\mathrm{d}\mu$ is less than this supremum since each f_n belongs to the set the supremum is being taken over; conversely, it suffices to show that if $g \in \mathcal{S}$ is any function such that $g \le f$, then $\int g\mathrm{d}\mu \le \lim_n \int f_n\mathrm{d}\mu$. Note that $\min(g, f_n)$ is a nondecreasing sequence of functions in $\mathcal{S}$ with $g = \lim \min(g, f_n)$, then by Lemma 5.1.1(c) and (d) we conclude that $\int g\mathrm{d}\mu = \lim \int \min(g, f_n)\mathrm{d}\mu \le \lim \int f_n\mathrm{d}\mu$ as needed. Next we show that this construction satisfies (a): if $f \le g$ everywhere on X, then $\{h \in \mathcal{S} : h \le f\} \subset \{h \in \mathcal{S} : h \le g\}$, and so the suprema are related by $\int f\mathrm{d}\mu \le \int g\mathrm{d}\mu$. Now, we show that this construction indeed satisfies the defining property, so let f_n be a nondecreasing sequence of functions in $\mathcal{C}$ with $f = \lim_n f_n$. Then by (a), we conclude that $\int f_n\mathrm{d}\mu$ is an increasing sequence bounded by $\int f\mathrm{d}\mu$, so that $\lim_n \int f_n\mathrm{d}\mu \le \int f\mathrm{d}\mu$. Conversely, for each f_n, by Theorem 1.2.3 there is a nondecreasing sequence $(g_{n,k})_k$ in $\mathcal{S}$ such that $f_n = \lim_k g_{n,k}$. For each n, define $h_n := \max_{m=1}^n g_{m,n}$. Then (h_n) is a nondecreasing sequence in $\mathcal{S}$ satisfying $h_n \le f_n$ and $f = \lim_n h_n$. It follows from the definition and (a) that $\int f\mathrm{d}\mu = \lim_n \int h_n\mathrm{d}\mu \le \lim_n \int f_n\mathrm{d}\mu$.

For (b), note that if f_n and g_n are nondecreasing sequences of functions in $\mathcal{S}$ with $f = \lim f_n$ and $g = \lim g_n$ (obtained say from Theorem 1.2.3), then $\alpha f_n + \beta g_n$ is a nondecreasing sequence of functions in $\mathcal{S}$ with $\alpha f + \beta g = \lim(\alpha f_n + \beta g_n)$, so that

$$\begin{aligned}\int(\alpha f + \beta g)\mathrm{d}\mu &:= \lim_n \int(\alpha f_n + \beta g_n)\mathrm{d}\mu \\ &= \lim_n\left(\alpha \int f_n\mathrm{d}\mu + \beta \int g_n\mathrm{d}\mu\right) \\ &= \alpha \lim_n \int f_n\mathrm{d}\mu + \beta \lim_n \int g_n\mathrm{d}\mu = \alpha \int f\mathrm{d}\mu + \beta \int g\mathrm{d}\mu.\end{aligned}$$

For (c), let $g_m = \sum_{n\le m} f_n$. Then (g_m) is a nondecreasing sequence of functions in $\mathcal{C}$ with $g := \lim_m g_m = \sum_n f_n \in \mathcal{C}$, so that

$$\int \sum_n f_n\mathrm{d}\mu = \int g\mathrm{d}\mu = \lim_m \int g_m\mathrm{d}\mu = \lim_m \int \sum_{n\le m} f_n\mathrm{d}\mu = \lim_m \sum_{n\le m} \int f_n\mathrm{d}\mu = \sum_n \int f_n\mathrm{d}\mu.$$

For (d), let $A := \{x \in X : f(x) \neq g(x)\}$; by Theorem 1.2.2(a)(i), this is measurable and by hypothesis this has measure 0. Define a function $h : X \to [0,\infty]$ by $h(x) = \begin{cases}\infty, & x \in A, \\ 0, & x \notin A.\end{cases}$ Then $h \in \mathcal{C}$. If we define the sequence $h_n = n\mathbb{1}_A \in \mathcal{S}$, then h_n is a nondecreasing sequence of functions in $\mathcal{S}$ with $h = \lim h_n$, so it follows that $\int h\mathrm{d}\mu = \lim_n \int h_n\mathrm{d}\mu = \lim_n(n\mu(A)) = \lim_n 0 = 0$. Now since $f \le g + h$, we conclude by (a) and (b) that $\int f\mathrm{d}\mu \le \int(g+h)\mathrm{d}\mu = \int g\mathrm{d}\mu + \int h\mathrm{d}\mu = \int g\mathrm{d}\mu$. By symmetry, $\int g\mathrm{d}\mu \le \int f\mathrm{d}\mu$ as well, and so we must have equality. For (e), note that if (f_n) is a nondecreasing sequence in $\mathcal{S}$ with $f = \lim f_n$, then $(f_n|_A)$ is a nondecreasing sequence in $\mathcal{S}(A)$ with $f|_A = \lim f_n|_A$ and $(f_n\mathbb{1}_A)$ is a nondecreasing sequence in $\mathcal{S}$ with $f\mathbb{1}_A = \lim f_n\mathbb{1}_A$, so that by Lemma 5.1.1(e) we conclude that $\int_A f|_A d\mu|_A = \lim_n \int_A f_n|_A d\mu_A = \lim_n \int f_n\mathbb{1}_A d\mu = \int f\mathbb{1}_A d\mu$. For (f), note that in this case $f\mathbb{1}_A \leq f\mathbb{1}_B$ and apply (a). For the second statement, $f\mathbb{1}_A$ and 0 agree almost everywhere. For the last, note simply that $1 = \mathbb{1}_A + \mathbb{1}_{X\setminus A}$. For (g), note that $(f\mathbb{1}_{A_i})$ is a series of functions in $\mathcal{C}$, so by (c) we have

$$\int_{\bigcup_i A_i} f d\mu = \int f\mathbb{1}_{\bigcup_i A_i} d\mu = \int \sum_i f\mathbb{1}_{A_i} d\mu = \sum_i \int f\mathbb{1}_{A_i} d\mu = \sum_i \int_{A_i} f d\mu.$$

For (h), let $A_t := f^{-1}[t, \infty]$; then each $\mathbb{1}_{A_t} \in \mathcal{S}$ and $0 \leq t\mathbb{1}_{A_t} \leq f\mathbb{1}_{A_t} \leq f$, so that by (a) we have

$$t\mu(A_t) = \int t\mathbb{1}_{A_t} d\mu \leq \int f\mathbb{1}_{A_t} d\mu \leq \int f d\mu.$$

For (i), apply (h) to $t = 1/n$ to conclude that $\mu(f^{-1}[1/n, \infty])$. Since $f^{-1}(0, \infty) = \bigcup_n f^{-1}([1/n, \infty])$, countable subadditivty shows that $\mu(f^{-1}(0, \infty)) = 0$ as needed. For the second part, apply (h) to $t = n$ to conclude that $\mu(f^{-1}[n, \infty]) \leq \frac{1}{n} \int f d\mu$. Thus $\mu(f^{-1}(\infty)) \leq \mu(f^{-1}[n, \infty]) \leq \frac{1}{n} \int f d\mu$ for every $n \geq 1$, and so $\int f d\mu < \infty$ implies that $\mu(f^{-1}(\infty)) = 0$.

■

Example 5.1.3. The function $\mathbb{1}_{\mathbf{Q}}$ on $[0, 1]$ is measurable with integral $\int_{[0,1]} \mathbb{1}_{\mathbf{Q}} d\mu = \mu(\mathbf{Q} \cap [0, 1]) = 0$.

Definition 5.1.4. Let $(X, \mathcal{M}, \mu)$ be a measure space. Let $\mathbf{F}$ be either $\mathbf{R}$ or $\mathbf{C}$.

(a) A measurable function $f : X \to [0, \infty]$ is said to be *integrable* (with respect to μ) iff $\int f\, d\mu < \infty$.
(b) A measurable function $f : X \to \mathbf{F}$ is said to be *integrable* (with respect to μ) iff $|f|$ is, i.e. iff $\int |f|\, d\mu < \infty$.
(c) We define $\mathcal{L}^1(\mu, \mathbf{F}) \subseteq \mathcal{M}(X, \mathbf{F})$ to be the subset of μ-integrable functions.

Next, suppose that $f \in \mathcal{L}^1(\mu, \mathbf{F})$. We define the *integral of f with respect to μ*, denoted $\int f\, d\mu$, as follows:

(a) First suppose that $\mathbf{F} = \mathbf{R}$. Then from

$$f^+, f^- \leq |f| = f^+ + f^-$$

it follows that f is integrable iff f^+ and f^- are, and in this case we define

$$\int f\, d\mu := \int f^+\, d\mu - \int f^-\, d\mu.$$

(b) Now suppose that $\mathbf{F} = \mathbf{C}$. Then from

$$|\operatorname{Re} f|, |\operatorname{Im} f| \leq |f| \leq |\operatorname{Re} f| + |\operatorname{Im} f|,$$

it follows that f is integrable iff $\operatorname{Re} f$ and $\operatorname{Im} f$ are, and in this case we define

$$\int f\, d\mu = \int \operatorname{Re} f\, d\mu + i \int \operatorname{Im} f\, d\mu.$$

Then we have:

Lemma 5.1.5.

(a) The subset $\mathcal{L}^1(\mu, \mathbf{F}) \subseteq \mathcal{M}(X, \mathbf{F})$ is an $\mathbf{F}$-vector subspace and $\int \cdot\, d\mu : \mathcal{L}^1(\mu, \mathbf{F}) \to \mathbf{F}$ is a monotonic $\mathbf{F}$-linear map that satisfies $\left|\int f\, d\mu\right| \leq \int |f|\, d\mu$ for all $f \in \mathcal{L}^1(\mu, \mathbf{F})$.
(b) If $f, g \in \mathcal{M}(X, \mathbf{F})$ are any two functions with $f = g\, [\mu\text{–a.e.}]$, then f is μ-integrable iff g is and in this case we have $\int f\, d\mu = \int g\, d\mu$.[4]
(c) If $f \in \mathcal{L}^1(\mu, \mathbf{F})$, then the set $\operatorname{supp} f = \{x : f(x) \neq 0\}$ is σ-finite.

Proof.

(a) The first follows from the fact that if $\alpha, \beta \in \mathbf{F}$ and $f, g \in \mathcal{M}(X, \mathbf{F})$, then $|\alpha f + \beta g| \leq |\alpha||f| + |\beta||g|$. To show linearity of the integral, first we show that if $f, g \in \mathcal{L}^1(\mu, \mathbf{R})$, then $\int (f + g)\, d\mu = \int f\, d\mu + \int g\, d\mu$. For this, let $h := f + g$ so that

$$h^+ - h^- = h = f + g = f^+ - f^- + g^+ - g^- \Rightarrow h^+ + f^- + g^- = h^- + f^+ + g^+.$$

[4] This suggests the following definition for an arbitrary f: an arbitrary $f : X \to \mathbf{F}$ is said to be μ-integrable iff there is some $g \in \mathcal{L}^1(\mu, \mathbf{F})$ such that $f = g\, [\mu\text{–a.e.}]$, and in this case we define $\int f\, d\mu = \int g\, d\mu$ for any such choice of g. This is well-defined by this claim in the lemma.

By Theorem 5.1.2(b), it follows that

$$\int h^+ \, d\mu + \int f^- \, d\mu + \int g^- \, d\mu = \int h^- \, d\mu + \int f^+ \, d\mu + \int g^+ \, d\mu$$

and since each of these quantities is finite we can rearrange to get what we need. From this, it follows by taking real and imaginary parts that if $f, g \in \mathcal{L}^1(\mu, \mathbf{C})$, then $\int (f+g) \, d\mu = \int f \, d\mu + \int g \, d\mu$ as well. Finally, it remains to show that if $f \in \mathcal{L}^1(\mu, \mathbf{C})$ and $\alpha \in \mathbf{C}$, then $\int (\alpha f) \, d\mu = \alpha \int f \, d\mu$. For this, first show it for $\alpha \geq 0$; then for $\alpha = -1$ using relations like $(-u)^+ = u^-$, and finally for $\alpha = i$; combining these results suffices for all $\alpha \in \mathbf{C}$. To show the last result, first suppose $\mathbf{F} = \mathbf{R}$; then the result follows from

$$\left|\int f \, d\mu\right| = \left|\int f^+ \, d\mu - \int f^- \, d\mu\right| \leq \int f^+ \, d\mu + \int f^- \, d\mu = \int (f^+ + f^-) \, d\mu = \int |f| \, d\mu.$$

Now suppose $\mathbf{F} = \mathbf{C}$. If $\int f d\mu = 0$ then there is nothing to show; else let $z = \arg \int f \, d\mu \in S^1$. Then

$$\left|\int f \, d\mu\right| = \bar{z} \int f \, d\mu = \int (\bar{z} f) \, d\mu = \int \mathrm{Re}(\bar{z} f) \, d\mu \leq \int |f| \, d\mu$$

where in the last step we've used the previous case and $\mathrm{Re}(\bar{z} f) \leq |\bar{z} f| = |f|$.

(b) The first statement follows from Theorem 5.1.2(d), and the second from the same result and the fact that if f and g agree almost everywhere, then all "components", namely $\mathrm{Re}^+, \mathrm{Re}^-, \mathrm{Im}^+$ and Im^- of f and g also agree almost everwhere.

(c) By replacing f by $|f|$, we can assume that f takes values in $[0, \infty)$. By Chebyshev's Inequlaity (Theorem 5.1.2(h)), we have for each $n \geq 1$ that $\mu(f^{-1}[1/n, \infty]) \leq n \int f \, d\mu < \infty$ and so we are done by $\operatorname{supp} f = \bigcup_{n \geq 1} f^{-1}[1/n, \infty]$.

■

Example 5.1.6. Consider a countable set X with the counting measure. Then a function $f : X \to [0, \infty]$ is integrable iff the sequence $a_n := f(x_n)$ for some (any hence any) enumeration (x_n) of X is summable (i.e. the series $\sum a_n$ converges to a number in $[0, \infty)$). An arbitrary function $f : X \to \mathbf{F}$ is integrable iff the sequence a_n above is absolutely summable. In particular, all of this is automatic if X is finite. As a concrete example, take $X = \mathbf{N}$. Then the space $\mathcal{L}^1(\mu, \mathbf{F})$ is simply denoted $\ell^1_{\mathbf{F}}$ (or simply ℓ^1 when $\mathbf{F}$ is clear from context; usually $\mathbf{F} = \mathbf{C}$) and is the space of all absolutely convergent $\mathbf{F}$-valued sequences.

Example 5.1.7. Consider $(\mathbf{R}, 2^{\mathbf{R}}, \delta_0)$, the real line with the Dirac delta measure based at 0. The integral with respect to this measure is written for historical reasons as $f \mapsto \int_{\mathbf{R}} f(x)\delta(x)dx$. If $f \in \mathcal{C}$ is any function, then f and the constant function $f(0)$ agree $[\delta_0\text{–a.e.}]$ on $\mathbf{R}$, so that by Theorem 5.1.2(d) we have

$$\int_{\mathbf{R}} f(x)\delta(x)dx = \int_{\mathbf{R}} f(0)\delta(x)dx = f(0)\mu(\mathbf{R}) = f(0).$$

In particular, f is integrable iff $f(0) < \infty$. It follows that a general $f \in \mathcal{C}$ is integrable iff $|f(0)| < \infty$, and in this case, $\int_{\mathbf{R}} f(x)\delta(x)dx = f(0)$. In general, the integral with respect to the Direct delta measure based at $a \in \mathbf{R}$ is written $f \mapsto \int_{\mathbf{R}} f(x)\delta(x-a)dx$.

Example 5.1.8. Let $(X, \mathcal{M}, \mu)$ be a finite measure space (e.g. a bounded measurable subset of $\mathbf{R}^d$ with respect to the Lebesgue measure). Then any bounded function $f \in \mathcal{C}$ is integrable simply by the fact that if $f \leq M$, then by Theorem 5.1.2 we have $\int f d\mu \leq \int M d\mu = M \int d\mu = M\mu(X) < \infty$. In particular, any bounded $f \in \mathcal{C}$ is integrable, because then $|f|$ is bounded too.

[Finally, we talk about image measures to talk about other familiar properties of the Lebesgue integral over $\mathbf{R}^d$, say.]

5.2 Limit Theorems

Theorem 5.2.1 (Monotone Convergence Theorem)**.**

Example 5.2.2. Construction of measures.

Theorem 5.2.3 (Fatou's Lemma)**.**

Theorem 5.2.4 (Lebesgue's Dominated Convergence Theorem)**.**

Application: Tannery's Theorem (Lebesgue DCT applied to ℓ^1), binomial converges, differentiation under the integral sign.

5.3 Lebesgue-Riemann Theory

6 Digression: Right Continuity and Lebesgue-Stieltjes Integrals

7 Product Spaces and Fubini-Tonelli

8 Normed Linear Spaces

8.1 Fundamentals

Let **F** be either **R** or **C** (or more generally a global field). We want to look at vector spaces over V and an associated metric d that is (a) translation and (b) scaling invariant. This says that the distance is really a function of one variable, and this function, the "norm" satisfies the following corresponding properties.

Definition 8.1.1.

- A *seminorm* on a vector space V over **F** is a function $\|\cdot\| : V \to [0, \infty)$ such that the following axioms are satisfied.
  (a) (Homogeneity) For all $v \in V$ and $\alpha \in \mathbf{F}$ we have $\|\alpha v\| = |\alpha| \, \|v\|$.
  (b) (Triangle Inequality) For all $v, w \in V$ we have $\|v + w\| \leq \|v\| + \|w\|$.
- A *norm* on V is a seminorm with the following additional property:
  (c) (Positive Definiteness) For all $v \in V$, we have $\|v\| = 0 \Rightarrow v = 0$.
- A *normed vector space* over **F** is a pair $(V, \|\cdot\|)$ of a vector space equipped with a norm.
- A *Banach space* is a complete normed vector space.
- A *Banach algebra B* is an associative **F**-algebra whose underlying vector space is a Banach space, and such that the norm is compatible with multiplication:
  (d) (Compatibility with Multiplication) For $f, g \in B$ we have $\|fg\| \leq \|f\| \, \|g\|$.

We define two norms to be equivalent when they define the same topology; this amounts to existence of constants bounding each by the other. It can be seen most easily using the definition of the completion using Cauchy sequences that the completion of a normed vector space is naturally a Banach space.

Example 8.1.2.

- If we have Banach spaces V, W, then we can define a product Banach space $V \times W$ with norm the max of the norms in V and W. (This agrees with the product topology and the product/direct sum vector space structure. In particular, the projections are continuous.) A sequence in $V \times W$ converges to a given element iff it happens componentwise. Similar remarks hold for arbitrary finite products.
- In particular, since **F** is always a Banach space we get $\mathbf{F}^n$ with the $\|\cdot\|_\infty$ norm. More generally, we have $\mathbf{F}^n$ with $\|\cdot\|_p$ norms for $p \in [1, \infty]$ is a Banach space. In fact, all norms on a finite-dimensional vector space are equivalent and give it a Banach space structure. (To be explained more if needed.) In particular, **F** itself is clearly a Banach algebra.
- If X is any nonempty set, then the space of bounded **F**-valued functions on X with the sup norm is a Banach space.
- We have ℓ^1 with the ℓ^1 norm, which is a Banach space. On the other hand, if we give it the ℓ^∞ norm, then it's not a Banach space: simply take the sequence f_n where $f_n(m) = \begin{cases} 1/m, & m < n, \\ 0, & m \geq n. \end{cases}$
- On bounded functions, we can take the sup norm. On $C[-1, 1]$, we can take $f \mapsto \sup_{[-1,1]} |f|$; then this is a Banach algebra (a uniformly convergent sequence of continuous functions is continuous). On the other hand, the norm $f \mapsto \int_{-1}^{1} |f| \mathrm{d}\mu$ is still a norm–a continuous function that vanishes a.e. vanishes everywhere– but this is not a Banach space by considering the sequence $f_n(x) = \begin{cases} 0, & x \in [-1, 0], \\ nx, & x \in [0, 1/n], \\ 1, & x \in [1/n, 1]. \end{cases}$
- In general, on $\mathcal{L}^1(X, \mu)$ the map $f \mapsto \int |f| \mathrm{d}\mu$ is a *seminorm*, but not a norm; we will correct this shortly.

We have the same definitions of convergence, absolute convergence, etc. The following is a compilation of some elementary results on normed vector spaces.

Theorem 8.1.3 (Fundamentals of Normed Vector Spaces)**.** Let V be a normed **F**-vector space.

(a) (Closures of Subspaces) If $U \subseteq V$ is a subspace, then so is $\overline{U} \subseteq V$.
(b) (Series in Banach Spaces) The space V is Banach iff every absolutely convergent series in V is convergent.
(c) (Quotients) If $U \subseteq V$ is a subspace, then the function $\|[v]\| := \inf_{u \in U}\{\|v + u\|\}$ is a seminorm on V/U. It is a norm iff $U \subseteq V$ is closed, in which case if two of $0 \to U \to V \to V/U \to 0$ are Banach then so is the third and the map $V \to V/U$ is surjective and continuous.

(d) (Riesz's Lemma) Let $U \subsetneq V$ be a proper closed subspace. Then for any $c \in [0, 1)$, there is a $v \in V$ with $\|v\| = 1$ such that $\|[v]\| = c$.
(e) (Loss of Compactness) The closed unit disc $\overline{B(0,1)}$ in V is compact iff $\dim V < \infty$.

Proof. For (a), note that if $u, v \in \overline{U}$, then we may pick $u_n \to u$ and $v_n \to v$ all lying in U; then for any $\alpha, \beta \in \mathbf{F}$ we have $U \ni (\alpha u_n + \beta v_n) \to \alpha u + \beta v$, so $\alpha u + \beta v \in \overline{U}$. For (b), note that if V is complete and $\sum_{n\geq 1} v_n$ an absolutely convergent series, then the sequence of partial sums is Cauchy. Conversely, let (u_n) be a Cauchy sequence. Passing to a subsequence (but not relabeling) we may assume that $\sum_{n\geq 1} \|u_{n+1} - u_n\| < \infty$ and set $u_0 = 0$. Then the series $\sum_{n\geq 0}(u_{n+1} - u_n)$ converges, and this must converge to a limit u of (u_n). For (c), note that homogeneity follows from the fact that U is a subspace, and the triangle inequality follows from thinking of vectors in V/U as equivalence classes, so that for $\xi, \eta \in V/U$ we have $\|\xi\| = \inf_{v\in\xi} \|v\|$ and $\xi + \eta = \{v + w : v \in \xi, w \in \eta\}$; in all, we get

$$\|\xi + \eta\| = \inf_{v\in\xi, w\in\eta} \|v + w\| \leq \inf_{v\in\xi, w\in\eta} (\|v\| + \|w\|) = \inf_{v\in\xi} \|v\| + \inf_{w\in\eta} \|w\| = \|\xi\| + \|\eta\| .$$

It is a norm iff for all $v \in V$ we have $\inf_{u\in U} \|v + u\| = 0 \Rightarrow v \in U$; to see that this implies $U = \overline{U}$, take a sequence u_n in U and $v = \lim u_n$, so then $\inf_{u\in U} \|v + u\| \leq \|v - u_n\|$ for all $n \geq 1$ implies $\inf_{u\in U} \|v + u\| = 0$ and so $v \in U$; to see that this condition is implied by $U = \overline{U}$, suppose that $\inf_{u\in U} \|v + u\| = 0$ and use it to find for each $n \geq 1$ a $u_n \in U$ such that $\|v - u_n\| < 1/n$, so that $v = \lim_n u_n \in \overline{U} = U$. In this case, if V is Banach, then U being a closed subspace is automatically Banach and then V/U is Banach as well as follows: first we show that if $\xi \in V/U$ is any class, then there is a $v \in \xi$ such that $\|v\| \leq 2\|\xi\|$. Indeed, this is clear if $\|\xi\| = 0$ because then $\xi = 0$; on the other hand, if $\|\xi\| > 0$, then $2\|\xi\|$ is not a lower bound for $\{\|v\| : v \in \xi\}$. To show that V/U we use the criterion of part (b): if $\sum_n \xi_n$ is a series in V/U such that $\sum_n \|\xi_n\| < \infty$, then for each n choosing a $v_n \in V$ such that $\|v_n\| \leq 2\|\xi_n\|$, it follows that $\sum_n \|v_n\| < \infty$, so that $\sum_n v_n$ converges in V to say v. Then $\xi = [v]$ is $\sum_n \xi_n$ in V/U because for any $N \geq 1$ we have $\left\|\xi - \sum_{n=1}^N \xi_n\right\| \leq \left\|v - \sum_{n=1}^N v_n\right\|$. Conversely, if both U and V/U are Banach, let (v_n) be a Cauchy sequence in V. Then $([v_n])$ is a Cauchy sequence in V/U, so there is a $v \in V$ such that $[v] = \lim_n [v_n]$. Then there is a sequence u_n in U such that $\|v - v_n + u_n\| \to 0$. Use that (v_n) is Cauchy to conclude that (u_n) is Cauchy and that if $u_n \to u$, then $v_n \to v + u$. Continuity follows from $\|[v] - [v']\| = \|[v - v']\| \leq |v - v'|$ for $v, v' \in V$. Let $w \in V \setminus U$ and by rescaling if necessary take $\|[w]\| = c$. The function $f : U \to \mathbf{R}$ given by $u \mapsto \|w + u\|$ is continuous and takes arbitrarily large values, so that the image of f is connected and contains $[c, \infty) \ni 1$. Take a $u \in U$ such that $\|w + u\| = 1$, and take $v = w + u$. To show (e), note if $\dim V < \infty$, then $V \cong \mathbf{F}^n$ as noted above and the closed unit disc is compact; conversely, assume that the closed unit disc is compact. For any $\varepsilon \in (0, 1)$, cover V by open balls of radius ε and find finitely many $u_i \in V$ such that $\overline{B(0,1)} \subseteq \bigcup_i B(u_i, \varepsilon)$. Let $U := \langle v_i \rangle$. If $U \subsetneq V$, then by Riesz's Lemma there is a $v \in \overline{B(0,1)}$ such that $\|[v]\| = \varepsilon$, which is a contradiction to $\|[v]\| = \inf_{u\in U} \|v + u\| \leq \min_i \|v - u_i\| < \varepsilon$. Therefore, $V = U$ is finite dimensional.

∎

Definition 8.1.4. A *Schauder basis* in a normed vector space is a list of unit vectors $(\mathrm{e}_n)_{n\geq 1}$ such that every $v \in V$ can be written as $v = \sum_{n=1}^\infty c_n \mathrm{e}_n$ for some unique $c_n \in \mathbf{F}$.

This implies that $V = \overline{\langle (\mathrm{e}_n)_n \rangle}$. (This implies separability, but every not every separable space has a Schauder basis. A rearrangement of a Schauder basis need not be another; if it does, then such a Schauder basis is called *unconditional*. It is a theorem (TO CITE) that $L^1(\mathbf{R})$ and $C[0, 1]$ don't admit unconditional Schauder bases.)

Next, we recall the various convergence criterion from a first course; then we observe that this combined with Theorem 8.1.3(b) give us a plethora of examples of convergent series in Banach spaces.

Proposition 8.1.5 (Convergence Tests in Nonnegative Reals)**.** Let (x_n) and (y_n) be sequences of nonnegative real numbers.

(a) (Comparison Test) If for all $n \gg 1$ we have $x_n \leq y_n$, then if $\sum_n y_n$ converges then so does $\sum_n x_n$.
(b) (Root Test) Let $r := \overline{\lim}_n x_n^{1/n}$.
  - If $r < 1$, then the series $\sum_n x_n$ is convergent.
  - If $r = 1$, then the series may or may not converge.
  - If $r > 1$, then the series is divergent.
(c) (D'Alembert's Ratio Test) If the limit $\lim_n (x_{n+1}/x_n)$ exists, then it equals r as in (b).
(d) (Cauchy Condensation) If x_n is decreasing, then $\sum_n x_n$ converges iff $\sum_n 2^n x_{2^n}$ does.
(e) (Kummer's Test) Let $\sum_n 1/r_n$ be a divergent series of positive terms. If

$$\frac{r_{n+1}}{r_n} \cdot \frac{x_{n+1}}{x_n} = 1 - \frac{\alpha}{r_n} + o\left(\frac{1}{r_n}\right)$$

for some $\alpha \in \mathbf{R}$, then the series $\sum_n x_n$ converges when $\alpha > 0$, the test is inconclusive if $\alpha = 0$, and diverges when $\alpha < 0$.

Proof.

(a) It follows that $\sum_n x_n$ is increasing and bounded above.
(b) When $r < 1$, it follows that for any $\varepsilon > 0$ we have $x_n \leq (r + \varepsilon)^n$ for all $n \gg 1$, so we may apply (a) for any $\varepsilon \in (0, 1 - r)$ by noting that the geometric series converges. Both $\sum_n 1/n$ and $\sum_n 1/n^2$ have $r = 1$. When $r > 1$, then $x_n \geq 1$ for infinitely many n.
(c) Let this limit be s temporarily. When $s > 0$, for any $\varepsilon \in (0, s)$, find an $n \geq 1$ such that for all $N \geq n$ we have $x_{N+1}/x_N \in (s - \varepsilon, s + \varepsilon)$, so then $x_N \in (x_n(s - \varepsilon)^{N-n}, x_n(s + \varepsilon)^{N-n})$. It follows that $x_N^{1/N} \in (x_n^{1/N}(s - \varepsilon)^{1-n/N}, x_n^{1/N}(s + \varepsilon)^{1+n/N}) \subseteq (s - 2\varepsilon, s + 2\varepsilon)$ for $N \gg n$. When $s = 0$, simply use half of this argument.
(d) This follows from $\sum_{j=1}^{2^{n+1}-1} x_j \leq \sum_{k=0} 2^k x_{2^k} \leq 2 \sum_{j=1}^{2^n} x_j$ by parenthesizing suitably.
(e) (TBD)

■

8.2 The Operator Norm

Definition 8.2.1. Let V, W be normed vector spaces and $T : V \to W$ a linear map. Then the following quantity

$$\sup_{\|v\| \leq 1} \{\|Tv\|\} = \sup_{v \neq 0} \left\{ \frac{\|Tv\|}{\|v\|} \right\} = \inf\{c \in [0, \infty) : \forall v \in V, \|Tv\| \leq c \|v\|\}$$

is called[5] the *operator norm* of T and is denoted $\|T\|$.

Example 8.2.2. The map $T : C[0, 1] \to C[0, 1]$ by $f \mapsto x^n f$ for any $n \geq 1$ is bounded with norm 1 when $C[0, 1]$ is given the supremum norm. On the other hand, if $\mathbf{F}^{\oplus\infty}$ is given the maximum/supremum norm, then the map $T : \mathbf{F}^{\oplus\infty} \to \mathbf{F}^{\oplus\infty}$ given by $T(\alpha_n) = (n\alpha_n)$ is unbounded.

We have clearly that for $T : V \to W$ and $S : W \to U$ the operator norm $\|ST\| \leq \|S\| \|T\|$. Similarly, if $S, T : V \to W$ are linear maps then it follows immediately that $\|S + T\| \leq \|S\| + \|T\|$ and that if $\alpha \in \mathbf{F}$ then $\|\alpha S\| < \infty$. In particular, if V, W are normed spaces, then the subcollection $\mathcal{B}(V, W) \subseteq \mathrm{Hom}_{\mathbf{F}}(V, W)$ of bounded linear maps (i.e. $T : V \to W$ satisfying $\|T\| < \infty$) is a normed vector space in its own right (with respect to the operator norm). Next, we will show that:

Lemma 8.2.3. If V, W are normed spaces and W Banach, then $\mathcal{B}(V, W)$ is Banach. In particular, if V is a Banach space, then $\mathrm{End}(V) := \mathcal{B}(V, V)$ is a Banach algebra.

Proof. The last statement follows immediately from the first by the first observation above. Let (T_n) be a Cauchy sequence in $\mathcal{B}(V, W)$. For any fixed $v \in V$, we have $\|T_n v - T_m v\| \leq \|T_n - T_m\| \|v\|$, so that $(T_n v)$ is a Cauchy sequence in W. Define a set map $T : V \to W$ by $Tv := \lim_n T_n v$, which exists because W is Banach. Then for $v, w \in V$ and $\alpha, \beta \in \mathbf{F}$ we have that $T(\alpha v + \beta w) = \lim_n T_n(\alpha v + \beta w) = \lim_n(\alpha T_n v + \beta T_n w) = \alpha \lim_n T_n v + \beta \lim_n T_n w = \alpha Tv + \beta Tw$ because these limits exist individually; this shows that T is linear. Next, for any $v \in V$ it follows that $\|Tv\| \leq \sup_n \|T_n v\| \leq \sup_n(\|T_n\|) \|v\|$ where the last supremum is finite because every Cauchy sequence is bounded; in particular, $\|T\| \leq \sup_n \|T_n\|$ and so $T \in \mathcal{B}(V, W)$. Finally, we need to show that $\lim_n \|T - T_n\| = 0$, for which let $\varepsilon > 0$ be given. Since (T_n) is Cauchy, there is an $n \geq 1$ such that for $M, N \geq n$ we have $\|T_M - T_N\| < \varepsilon$. then if $v \in V$ is arbitrary and $N \geq n$, then $\|(T - T_N)v\| = \lim_M \|T_M v - T_N v\| \leq \varepsilon \|v\|$, so that $\|T - T_N\| \leq \varepsilon$ as needed. ■

The following lemma is one of the reasons we care about bounded operators:

Lemma 8.2.4. Let V, W be normed spaces and $T : V \to W$ a linear map. Then TFAE:

(a) T is bounded.
(b) T is Lipschitz.
(c) T is uniformly continuous.
(d) T is continuous.
(e) T is continuous at $0 \in V$. Equivalently, there is an $r > 0$ such that $T^{-1}B(r, 0) \subset V$ is open.

[5] In the first supremum, we can also take $\|v\| < 1$ or $\|v\| = 1$.

(f) T is continuous at some point $v \in V$.

Proof. The implication (a) $\Rightarrow$ (b) follows from $\|Tv - Tw\| = \|T(v-w)\| \le \|T\|\,\|v-w\|$, so the Lipschitz constant $\|T\|$ works. The implications (b) $\Rightarrow$ (c) $\Rightarrow$ (d) $\Rightarrow$ (e) $\Rightarrow$ (f) are obvious. That (f) $\Rightarrow$ (e) follows immediately from linearity. To show (e) $\Rightarrow$ (a), take a $\delta > 0$ such that $\|v\| \le \delta \Rightarrow \|Tv\| \le 1$, then $\|T\| \le \delta^{-1}$. ■

Therefore, the correct category to deal with here is that of normed spaces and bounded linear maps between them. In the following, we will implicitly identify continuous and bounded linear maps of normed spaces.

8.3 Open Mapping, Closed Graph, Closed Complements, Banach-Saks-Steinhaus

Recall the Baire category theorem: in every locally compact Hausdorff or metric space, a countable intersection of dense open sets is dense. Using this, we prove some fundamental results in the theory of Banach spaces. Using this, we prove:

Theorem 8.3.1. Let $T : V \to W$ be a bounded linear map of normed spaces with V Banach. Then $TV \subseteq W$ is closed iff

there is a constant $C > 0$ such that for any $w \in TV$ there is an $v \in V$ such that $Tv = w$ and $\|v\| \le C\,\|w\|$. (1)

Proof. Suppose that (1) holds and $TV \ni w_n \to w \in W$. Define $w_0 = 0$ and by passing to a subsequence, assume $\sum_n \|w_n - w_{n-1}\| < \infty$. For each n, pick a $v_n \in V$ such that $Tv_n = w_n - w_{n-1}$ and $\|v_n\| \le C\,\|w_n - w_{n-1}\|$; then the series $\sum_n v_n$ is absolutely convergent and so (since V is Banach) it converges to a $v \in V$. Then continuity of T shows $Tv = \lim_N \sum^N Tv_n = \lim_N w_N = w$, so $w \in TV$ and TV is closed. Conversely, replace W by TV to assume that T is surjective. Let B_V and B_W be unit open balls in V and W; then surjectivity says that $W = \bigcup_{n=1}^{\infty} n\overline{TB_V}$, so Baire's theorem says that $n\overline{TB_V}$ has a nonempty interior, so say $w_0 + r_1 B_W \subseteq n\overline{TB_V} \Rightarrow r_1 B_W \subseteq n\overline{TB_V} - w_0 \subseteq 2n\overline{TB_V}$ for some $w_0 \in W$ and $r_1 > 0$, so taking $r = 2n/r_1$, we get that $\overline{B_W} \subseteq r\overline{TB_V}$; in other words, this says that for any $w \in W$ and $\varepsilon > 0$, there is a $v \in V$ with $\|v - Tw\| < \varepsilon$ and $\|v\| \le rw$. We claim that $C = 2r$ works. Indeed, fix a $w \in W$ and inductively choose for each $k \ge 1$ a $u_k \in V$ such that $\left\|w - \sum_{j=1}^{k} Tu_j\right\| < 2^{-k}$ and $\|u_k\| \le r2^{-k+1}\,\|w\|$. Since V is complete, the series $\sum_{j=1}^{\infty} u_j$ converges from the second inequality say to $v \in V$ with $\|v\| \le r\,\|w\| \sum_{k=1}^{\infty} 2^{-k+1} = 2r\,\|w\|$. By continuity of T and the norm, taking $\lim_{k\to\infty}$ in the first term shows that $\|w - Tv\| = 0$, so that $w = Tv$, and we've seen that $\|v\| \le 2r\,\|w\|$. ■

Corollary 8.3.2. Let V, W be Banach spaces and let $T : V \to W$ be a linear map.

(a) (Open Mapping) If T is bounded, then $TV \subseteq W$ is closed iff T is open. In particular, if T is surjective, then T is open.
(b) If T is continuous and bijective, then T^{-1} is continuous, i.e. T is an isomorphism of Banach spaces.
(c) (Closed Graph) The map T is continuous iff $\Gamma_T \subseteq V \times W$ is a closed subspace.

Proof. For (a), we show that T is open iff (1) holds. Continuing the notation from before, (1) is equivalent to the existence of a $C > 0$ such that $\overline{B_W} \cap TV \subseteq CT\overline{B_V}$, which is equivalent to the existence of a $C' > 0$ such that $B_W \cap TV \subseteq CTB_V$ (for one direction take $C' > C$ and for the other $C' < C$), which is equivalent to the existence of an $r > 0$ such that $rB_W \cap TV \subseteq TB_V$ and this is equivalent to the openness of T by its linearity. For (b), note that an open continuous bijection is a homeomorphism. For (c), one implication is trivial: if $f : X \to Y$ is a continuous map of topological spaces and Y is Hausdorff, then $\Gamma_f \subseteq X \times Y$ is closed. Conversely, the map $S := \mathrm{pr}_1|_{\Gamma_T} : \Gamma_T \to V$ is a linear bijection, and is continuous because it is the composition of the continuous $\Gamma_T \hookrightarrow V \times W \xrightarrow{\mathrm{pr}_1} V$. If Γ_T is closed, then it is Banach and so by (b), S^{-1} is continuous. It follows that the composition $T = \mathrm{pr}_2 \circ S^{-1}$ is also continuous. ■

Definition 8.3.3. A subspace of a vector space is said to have *finite codimension* if it admits a finite dimensional linear complement, or equivalently if the corresponding quotient space is finite dimensional.

Corollary 8.3.4 (Closed Complements)**.**

(a) Let $U \subseteq V$ be a closed subspace. Then U has a closed linear complement in V iff there is a continuous projection of V onto U.

(b) Let $T : V \to W$ be a bounded operator of Banach spaces. If TV has a closed linear complement in W, then TV is closed in W. In particular, if TV has finite codimension in W, then TV is closed in W.

Proof.

(a) If $P : V \to V$ is a continuous projection (i.e. $P^2 = P$), then $PV \subseteq V$ is closed by a trivial application of Theorem 8.3.1. If $P : V \to V$ is a continuous projection onto U, then $V \cong U \oplus (1_V - P)V$, and the latter is closed because $1_V - P$ is a continuous projection. Conversely, suppose there is a closed U' such that $V \cong U \oplus U'$, and let $P : V \to V$ be the projection given by this U'. We claim that P is continuous. Indeed, since P is the composition of $P : V \to U \hookrightarrow V$, it suffices to show that $P : V \to U$ is continuous, for which by Corollary 8.3.2(c) it suffices to show that the subspace $\Gamma_P = \{(u + u', u)\} \subseteq V \times U$ is closed; by indeed, if $u_n \to u$ and $u_n + u_n' \to v$, then $u_n' \to v - u$ and so by closedness of U' we have $v - u \in U'$ and hence $(u_n + u_n', u_n) \to (v, u) = (u + (v - u), u) \in \Gamma_P$).

(b) Let $U \subseteq W$ be closed such that $W = U \oplus TV$; note that U is Banach. Consider the linear operator $U \times V \to W$ given by $(u, v) \mapsto u + Tv$; this is a continuous linear map from the Banach space $V \times U$ onto W, so by Theorem 8.3.1 there is a $C > 0$ such that for any $w \in W$ there is a $(u, v) \in U \times V$ with $u + Tv = w$ and $\max \|u\|, \|v\| \leq C\|w\|$. If $w \in Tv$, then necessarily $u = 0$ and so this shows that for the same $C > 0$ we have that there is a $v \in V$ with $Tv = w$ and $\|v\| \leq C\|w\|$, so again by Theorem 8.3.1, $TV \subseteq W$ is closed. The last statement follows because any subspace of finite codimension admits a finite dimensional and hence closed linear complement.

■

Theorem 8.3.5 (Banach-Saks-Steinhaus Theorem/Uniform Boundedness Principle)**.** Let V be a Banach space and W be a normed vector space.

(a) If $\mathcal{F}$ is a family of bounded linear maps from V to W such that for any $v \in V$, we have $\sup_{T\in\mathcal{F}}\{\|Tv\|\} < \infty$. Then $\sup_{T\in\mathcal{F}}\{\|T\|\} < \infty$.[6]

(b) If (T_n) is a sequence of bounded linear maps $V \to W$ that converges to a linear map $T : V \to W$ pointwise, then T is bounded.[7]

Proof. For this, the following lemma is helpful.

Lemma 8.3.6. Let $\mathcal{F}$ be a family of continuous real-valued functions on a Baire space X (e.g. complete metric space or locally compact Hausdorff space) that is pointwise bounded. Then there is a nonempty open set $U \subseteq X$ such that $\mathcal{F}$ is uniformly bounded on U.

Proof. Define $A_n := \bigcap_{f\in\mathcal{F}} |f|^{-1}[0, n]$; pointwise boundedness says that $X = \bigcup_n A_n$, so by Baire's theorem, some A_n has a nonemtpy interior. Take U to be this interior and the bound to be n. ■

(a) The collection of continuous functions $f_T : V \to \mathbf{R}$ defined by $f_T(v) := \|Tv\|$ for $T \in \mathcal{F}$ is pointwise bounded by hypothesis, so by the Lemma there are $C, r > 0$ and $v_0 \in V$ such that $\|Tv\| \leq C$ for all $v \in B(v_0, r)$ and $T \in \mathcal{F}$. Therefore, for each $T \in \mathcal{F}$ and $0 \neq v \in V$ of norm $Tv < r$ we have $\|Tv\| \leq \|T(v_0 + v)\| + \|Tv_0\| \leq C + \sup_{T\in\mathcal{F}}\{\|Tv_0\|\} =: M$, so $\sup_{T\in\mathcal{F}}\{\|T\|\} \leq r^{-1}M$.

(b) By (a), the sequence $\{T_n\}$ is uniformly bounded, say by $C > 0$. Then for any $v \in V$ continuity of the norm shows that $\|Tv\| = \lim_{n\to\infty} \|T_n v\| \leq C\|v\|$, so that T is bounded by C as well.

■

8.4 Bounded Functionals and Hahn-Banach Theorem

We look at bounded functionals on normed vector spaces: given a normed vector space V, define $V^* := \mathcal{B}(V, \mathbf{F})$ to be the Banach space of bounded linear functionals on V with the operator norm. First we have:

Lemma 8.4.1. Let V be a normed space and $\varphi : V \to \mathbf{F}$ a nonzero linear map (i.e. a functional). Then TFAE:

(a) $\varphi \in V^*$,

(b) $\ker\varphi \subseteq V$ is closed, and

[6] In other words, if a family of bounded linear maps on a Banach space is bounded pointwise uniformly across the family, then it is bounded uniformly uniformly across the family.

[7] In other words, the pointwise limit of a sequence of bounded linear maps on a Banach space is a bounded linear map.

(c) $\overline{\ker\varphi} \neq V$.

Proof. The implications (a) $\Rightarrow$ (b) $\Rightarrow$ (c) are trivial, since we are assuming $\varphi \neq 0$. To show (b) $\Rightarrow$ (a), suppose that φ is not bounded and select a sequence (v_n) in V with $\|v_n\| = 1$ but $\|\varphi v_n\| \geq n$ for all $n \geq 1$. Then $\ker\varphi \ni (\varphi v_1)^{-1}v_1 - (\varphi v_n)^{-1}v_n \xrightarrow{n\to\infty} (\varphi v_1)^{-1}v_1 \notin \ker\varphi$. To show (c) $\Rightarrow$ (b), suppose there is a $v \in \overline{\ker\varphi} \setminus \ker\varphi$. Then given any $w \in W$, we have $w = (w - (\varphi v)^{-1}(\varphi w)v) + (\varphi v)^{-1}(\varphi w)v \in \overline{\ker\varphi}$, showing that $\overline{\ker\varphi} = V$. ■

Theorem 8.4.2 (Hahn-Banach)**.** Let V be a normed vector space, $U \subseteq V$ a subspace and $\varphi \in U^*$. Then φ extends to a $\overline{\varphi} \in V^*$ with $\|\overline{\varphi}\| = \|\varphi\|$.

First, we show a simple extension lemma for $\mathbf{F} = \mathbf{R}$ and a lemma relating real and complex functionals.

Lemma 8.4.3.

(a) (Extension Lemma) When $\mathbf{F} = \mathbf{R}$, with the hypothesis of Theorem 8.4.2, given any $v \in V \setminus U$ there is an extension $\overline{\varphi} \in \langle U, v\rangle^*$ such that $\|\overline{\varphi}\| = \|\varphi\|$.

(b) (Realification) If $V/\mathbf{C}$ is a complex vector space and $\varphi : V \to \mathbf{C}$ is a $\mathbf{C}$-linear functional, then $\varphi_0 := \operatorname{Re}\varphi : V_{\mathbf{R}} \to \mathbf{R}$ is an $\mathbf{R}$-linear functional and

$$\varphi(v) = \varphi_0(v) - i\varphi_0(iv) \tag{2}$$

for any $v \in V$. Conversely, given an $\mathbf{R}$-linear functional $\varphi_0 : V_{\mathbf{R}} \to \mathbf{R}$, the map $\varphi : V \to \mathbf{C}$ defined by Equation (2) is a $\mathbf{C}$-linear functional, giving inverse linear isomorphisms $\operatorname{Hom}_{\mathbf{C}}(V, \mathbf{C})_{\mathbf{R}} \xrightarrow{\sim} \operatorname{Hom}_{\mathbf{R}}(V_{\mathbf{R}}, \mathbf{R})$. Further, if V is normed then so is $V_{\mathbf{R}}$, and then this linear isomorphism satisfies $\|\varphi\| = \|\varphi_0\|$ and so takes $(V^*)_{\mathbf{R}} \xrightarrow{\sim} (V_{\mathbf{R}})^*$.

Proof. For (a), for any $c \in \mathbf{R}$, the extension by $\overline{\varphi}(v) = c$ satisfies $\|\overline{\varphi}\| \geq \|\varphi\|$. To pick a c such that the opposite inequality holds, we have to find a c such that $\|\varphi(u) + \alpha c\| \leq \|\varphi\|\,\|u + \alpha v\|$ for all $u \in U$ and $\alpha \in \mathbf{R}$. By rescaling u, it suffices to show this for all $u \in U$ and $\alpha = 1$. This amounts to showing that there is a $c \in \mathbf{R}$ such that for all $u \in U$ we have $-\|\varphi\|\,\|u + v\| - \varphi(u) \leq c \leq \|\varphi\|\,\|u + v\| - \varphi(u)$, for which it suffices to show

$$\sup_{u\in U}(-\|\varphi\|\,\|u + v\| - \varphi(u)) \leq \inf_{w\in U}(\|\varphi\|\,\|w + v\| - \varphi(w)).$$

This from the following for any $u, w \in U$:

$$-\|\varphi\|\,\|u + v\| - \varphi(u) \leq \|\varphi\|(\|w + v\| - \|w - u\|) - \varphi(u) = \|\varphi\|\,\|w + v\| - \varphi(w) - (\|\varphi\|\,\|w - u\| - \varphi(w - u)).$$

For (b), the first part is trivial; when φ is defined by Equation (2), its $\mathbf{R}$-linearity $V \to \mathbf{C}$ is clear and $\mathbf{C}$-linearity follows from $\varphi(iv) = \varphi_0(iv) - i\varphi_0(-v) = i(\varphi_0(v) - i\varphi_0(iv)) = i\varphi(v)$. In case V is normed, the inequality $\|\varphi_0\| \leq \|\varphi\|$ is clear; for the converse, note that for any $v \in V$ we have

$$|\varphi(v)|^2 = \varphi(\overline{\varphi(v)}v) = \varphi_0(\overline{\varphi(v)}v) \leq \|\varphi_0\|\left|\overline{\varphi(v)}\right|\|v\| \Rightarrow |\varphi(v)| \leq \|\varphi_0\|\,\|v\|\,,$$

showing $\|\varphi\| \leq \|\varphi_0\|$. ■

Proof of Theorem 8.4.2. When $\mathbf{F} = \mathbf{R}$, consider the collection

$$\mathcal{A} := \{(W, \varphi_W) : U \subseteq W \subseteq V, \varphi_W \in W^*, \varphi_W|_U = \varphi, \|\varphi_W\| = \|\varphi\|\},$$

partially ordered by $(W, \varphi_W) \leq (W', \varphi_{W'})$ iff $W \subseteq W'$ and $\varphi_{W'}|_W = \varphi_W$. If $\{(W_\alpha, \varphi_\alpha)\}_\alpha$ is a chain in $\mathcal{A}$, then taking $W := \bigcup_\alpha W_\alpha$ and $\varphi_W(w) = \varphi_{W_\alpha}(w)$ when $w \in W_\alpha$ (this is independent of the choice of α because of the total order), we get automatically that $\|\varphi_W\| \geq \|\varphi\|$ and for any $w \in W$ there is an α such that $w \in W_\alpha$ and so $\|\varphi_W(w)\| = \left\|\varphi_{W_\alpha}(w)\right\| \leq \left\|\varphi_{W_\alpha}\right\|\|w\| \leq \|\varphi\|\,\|w\|$ so $\|\varphi_W\| = \|\varphi\|$. In particular, $(W, \varphi_W) \in \mathcal{A}$ is an upper bound for this chain. It follows from Zorn's Lemma that $\mathcal{A}$ has a maximal element (W, φ_W), and this maximal necessarily has $W = V$ by Lemma 8.4.3(a); this completes the proof for when $\mathbf{F} = \mathbf{R}$. When $\mathbf{F} = \mathbf{C}$, the above case tells us that $\varphi_0 \in (U_{\mathbf{R}})^*$ extends to a $\overline{\varphi}_0 \in (V_{\mathbf{R}})^*$ with $\left\|\overline{\varphi}_0\right\| = \|\varphi_0\|$, and then Lemma 8.4.3(b) gives us an $\overline{\varphi} \in V^*$ extending $\overline{\varphi}_0$ and so by Equation (2) also φ, and also satisfying $\|\overline{\varphi}\| = \left\|\overline{\varphi}_0\right\| = \|\varphi_0\| = \|\varphi\|$. ■

Corollary 8.4.4. Let V be a normed vector space.

(a) (The norm from functionals.) Let $v \in V \setminus \{0\}$. Then there is a $\varphi \in V^*$ such that $\|\varphi\| = 1$ and $\|v\| = \varphi(v)$. In particular, $\|v\| = \sup_{\varphi\in V^*, \|\varphi\|=1} \|\varphi(v)\|$.

(b) Suppose $U \subseteq V$ and $v \in V$. Then $v \in \overline{U}$ iff $\varphi(v) = 0$ for all $\varphi \in V^* \cap \mathrm{Ann}(U)$.
(c) We have $\dim V \leq \dim V^*$ with equality if V is finite dimensional.

Proof. For (a), let $U := \langle v \rangle$ and $\varphi : U \to \mathbf{F}$ by $\varphi(\alpha v) = \alpha \|v\|$ for $\alpha \in \mathbf{F}$. Then $\|\varphi\| = 1$ and $\varphi(v) = \|v\|$; extend this by Hahn-Banach to an $\overline{\varphi} \in V^*$ again with $\|\overline{\varphi}\| = 1$ and $\overline{\varphi}(v) = \|v\|$. For (b), if $\varphi \in V^* \cap \mathrm{Ann}(U)$, then by Lemma 8.4.1(b) we get $U \subseteq \ker\varphi \Rightarrow \overline{U} \subseteq \overline{\ker\varphi} = \ker\varphi$. Conversely, if $v \notin \overline{U}$ define $\varphi : \langle U, v \rangle \to \mathbf{F}$ by $\varphi(u + \alpha v) = \alpha$; then φ is a linear functional with $\ker\varphi = U$ and $\varphi(v) = 1$. Since $v \notin \overline{U}$, it follows from Lemma 8.4.1(c) that $\varphi \in \langle U, v \rangle^*$; by Hahn-Banach, this extends to an $\overline{\varphi} \in V^* \cap \mathrm{Ann}(U)$ with $\overline{\varphi}(v) = 1$. For (c), suppose we have $v_1, \ldots, v_n \in V$ linearly independent. Define $U := \langle v_1, \ldots, v_n \rangle$ and $v_i^* \in U^*$ the duals. By Hahn-Banach, we get extensions $\varphi_i^* \in V^*$ still satisfying $\varphi_i v_j = \delta_{ij}$, and hence linearly independent; this proves the $\dim V \leq \dim V^*$. When $\dim V < \infty$, equality is well-known. ■

We end with an observation about the double dual:

Lemma 8.4.5. Let V be a normed vector space. For any $v \in V$, the functional $\mathrm{ev}_v : V^* \to \mathbf{F}$ is bounded. The map $\mathrm{ev} : V \to (V^*)^*$ is a continuous isometry onto its image and an isomorphism when V is finite dimensional.

Proof. For any $\varphi \in V^*$ we have $\|\mathrm{ev}_v \varphi\| = \|\varphi v\| \leq \|v\| \|\varphi\| \Rightarrow \|\mathrm{ev}_v\| \leq \|v\|$, and so ev_v is bounded. On the other hand, given a $v \in V \setminus \{0\}$, use Corollary 8.4.4(a) to produce a φ with $\|\varphi\| = 1$ and $\|v\| = \varphi(v)$ to get $\|v\| = \varphi(v) = \mathrm{ev}_v(\varphi) \leq \|\mathrm{ev}_v\| \|\varphi\| = \|\mathrm{ev}_v\|$; this shows that $\|\mathrm{ev}_v\| = \|v\|$. Therefore, the map $\mathrm{ev} : V \to (V^*)^*$ has Lipschitz constant 1 and is a continuous isometry onto its image. If V is finite dimensional, then equality is well-known by dimension reasons. ■

Definition 8.4.6. A normed vector space V is called *reflexive* if the map $\mathrm{ev} : V \to (V^*)^*$ is an isomorphism.

8.5 Weak Topologies

9 Modes of Convergence and $L^p(\mu)$

In this section, we focus on complex-valued measurable and integrable functions, although the same theory can be applied easily to real-valued or $[-\infty, +\infty]$-valued functions.

9.1 Modes of Convergence

Definition 9.1.1. Let $(X, \mathcal{M}, \mu)$ be a measure space. If (f_n) is a sequence in $\mathcal{M}(X, \mathbf{C})$ and $f \in \mathcal{M}(X, \mathbf{C})$, then we say that (f_n) converges to f

(a) *almost everywhere* if $f = \lim_n f_n$ holds almost everywhere;
(b) *in measure* if for each $\varepsilon > 0$ we have $\lim_n \mu(\{x \in X : |f(x) - f_n(x)| > \varepsilon\}) = 0$; and
(c) *in mean* if $\lim_n \int |f - f_n| \mathrm{d}\mu = 0$.

In general, these notions are not equivalent (see Cohn for counterexamples). However, these are closely related.

Theorem 9.1.2. Let (f_n) and f be as above.

(a) If $\mu(X) < \infty$, then convergence almost everywhere implies convergence in measure.
(b) In any case, convergence in measure implies convergence of a subsequence to f almost eveywhere.
(c) Convergence in mean implies convergence in measure.
(d) Either convergence a.e. or convergence in measure implies convergence in mean, IF there is a function $g \in \mathcal{L}^+$ such that $|f_n| \leq g$ and $|f| \leq g$ hold a.e.

Proof.

(a) For $\varepsilon > 0$ and $n \geq 1$, let $A_{n,\varepsilon} := \{x \in X : |f(x) - f_n(x)| > \varepsilon\}$. Define $B_{n,\varepsilon} := \bigcup_{m \geq n} A_{m,\varepsilon}$; since $B_{n,\varepsilon} \supseteq A_{n,\varepsilon}$, it suffices to show $\lim_n \mu(B_{n,\varepsilon}) = 0$. Then by Lemma **??**(d) (this is where we use $\mu(X) < \infty$), it suffices to show $\mu(\bigcap_n B_{n,\varepsilon}) = 0$, for which it suffices to observe that $\bigcap_n B_{n,\varepsilon} \subseteq \{x \in X : f_n(x) \text{ does not converge to } f(x)\}$.
(b) Inductively construct a strictly increasing sequence of positive integers $(n_k)_{k \geq 1}$ by choosing n_k so that $\mu(A_{n_k, 1/k}) < 2^{-k}$. Let $C := \bigcap_{n \geq 1} \bigcup_{k \geq n} A_{n_k, 1/k}$; we claim that $\mu(C) = 0$ and $f = \lim_k f_{n_k}$ on $X \setminus C$. For the first, observe that $\mu(\bigcup_{k \geq n} A_{n_k, 1/k}) \leq \sum_{k \geq n} \mu(A_{n_k, 1/k}) \leq \sum_{k \geq n} 2^{-k} = 2^{-n+1}$, so $\mu(C) \leq 2^{-n+1}$ for all n. For the second statement, note that $x \notin C$ implies there is an $n \geq 1$ such that $x \notin \bigcup_{k \geq n} A_{n_k, 1/k}$, so for all $k \geq n$ we have $|f(x) - f_{n_k}(x)| \leq 1/k$, which shows that $f(x) = \lim_k f_{n_k}(x)$.
(c) This follows immediately from taking $t = 1/\varepsilon$ for any $\varepsilon > 0$ in Chebyshev's Inequality (Theorem 5.1.2(h)) which says that $\mu(\{x \in X : |f(x) - f_n(x)| > \varepsilon\}) \leq \varepsilon^{-1} \int |f - f_n| \mathrm{d}\mu$.
(d) First suppose that (f_n) converges to f a.e., so that $(|f - f_n|)$ converges to 0 almost everywhere. Then $|f - f_n| \leq |f| + |f_n| \leq 2g$ holds a.e. and so by the DCT, we conclude that $\lim_n \int |f - f_n| \mathrm{d}\mu = \int 0 \, \mathrm{d}\mu = 0$. Now suppose that we have convergence in measure. Then every subsequence of (f_n) has a subsequence that converges to f a.e. by (b), and so by what we have just proved, in mean. If the original sequence did not converge to f in measn, there would be a positive number $\varepsilon > 0$ and a subsequence (f_{n_k}) of (f_n) with $\int |f - f_{n_k}| \mathrm{d}\mu \geq \varepsilon$ for all k. Since this subsequence could have no subsequence converging to f in mean, we have a contradiction. Therefore, (f_n) must converge to f in mean.

■

9.2 $\mathcal{L}^p$ and L^p

As in the previous section, we allow both $\mathbf{F} = \mathbf{R}$ and $\mathbf{C}$.

Definition 9.2.1. Given a measure space $(X, \mathcal{M}, \mu)$, a $p \in (0, \infty)$, and a measurable function $f \in \mathcal{M}(X, \mathbf{F})$, we define the *p-norm* of f to be

$$\|f\|_p := \left(\int |f|^p \mathrm{d}\mu \right)^{1/p} \in [0, \infty].$$

Similary, we define the *∞-norm* or the *essential supremum* of f by

$$\|f\|_\infty = \inf\{t > 0 : |f| \leq t \text{ a.e.}\}.$$

Given any $p \in (0, \infty]$ we define the *Lebesgue space* $\mathcal{L}^p(\mu) := \{f \in \mathcal{M}(X, \mathbf{F}) : \|f\|_p < \infty\}$.

Note that the p-norms for $p \in (0, \infty]$ only depend on f upto changing it by a function on a set of measure 0 (this is why we need the *essential* supremum). Next, $|f| \le \|f\|_\infty$ a.e. (by $|f(x)| > \|f\|_\infty$ iff there is an $n \ge 1$ such that $|f(x)| > \|f\|_\infty + 1/n$) and so for any $p \in (0, \infty]$ we have $\|f\|_p = 0$ iff $f = 0$ a.e. Further, this construction generalizes the space $\mathcal{L}^1$ defined in §5 by Lemmas **??**(a) and **??**.

Example 9.2.2. If X is a countable set with the counting measure, then for any $f : X \to \mathbf{F}$, we have $\|f\|_p = \left(\sum_{x\in X} |f(x)|^p\right)^{1/p}$ and $\|f\|_\infty = \sup |f|$. When X is countably infinite (so basically $X = \mathbf{N}$ or $X = \mathbf{Z}$), the space $\mathcal{L}^p(\mu)$ is denoted ℓ^p. If X is finite, then every function $f : X \to \mathbf{F}$ is in every $\mathcal{L}^p$ for $p \in (0, \infty]$.

Lemma 9.2.3. Let $(X, \mathcal{M}, \mu)$ be a measure space and $p \in (0, \infty]$. Let $f, g \in \mathcal{M}(X, \mathbf{F})$ and $\alpha \in \mathbf{F}$.

(a) When $p \ne \infty$, we have that $\|f + g\|_p^p \le 2^p(\|f\|_p^p + \|g\|_p^p)$.
(b) When $p = \infty$, we have that $\|f + g\|_\infty \le \|f\|_\infty + \|g\|_\infty$.
(c) In any case, we have $\|\alpha f\|_p = |\alpha|\,\|f\|_p$.

It follows that $\mathcal{L}^p(\mu)$ is an $\mathbf{F}$-vector space for any $p \in (0, \infty]$.

Proof. For (a), for any $x \in X$ we have that $|f(x) + g(x)|^p \le (2\max\{|f(x)|, |g(x)|\})^p \le 2^p(|f(x)|^p + |g(x)|^p)$. Integrating over X, we get the result. For $p \in (0, \infty)$, it follows that $\mathcal{L}^p(\mu)$ is closed under addition. For (b), we have $|f + g| \le |f| + |g|$ and the latter is at most $\|f\|_\infty + \|g\|_\infty$ a.e., so by definition of the essential supremum we have $\|f + g\|_\infty \le \|f\|_\infty + \|g\|_\infty$. For (c), the result is clear for $p \ne \infty$ or $\alpha = 0$. When $p = \infty$ and $\alpha \ne 0$, the result follows from $|\alpha f| \le t$ a.e. iff $|f| \le |\alpha|^{-1} t$ a.e. ■

In fact, these are all *almost* normed vector spaces (except that $\|f\|_p = 0$ doesn't imply $f = 0$ as we'd like). We correct this now.

Definition 9.2.4. Given a measure space $(X, \mathcal{M}, \mu)$ and $p \in (0, \infty]$, we define the *Lebesgue space* $L^p(\mu)$ to be the quotient of $\mathcal{L}^p(\mu)$ by the subspace of functions that are zero almost everywhere.

Then the function $\|\cdot\|_p$ descends to a $L^p(\mu) \to [0, \infty)$. For $p \in [1, \infty]$, Minkowski's inequality below tells us that $L^p(\mu)$ is a normed vector space over $\mathbf{F}$. We say that a sequence $f_n \in \mathcal{M}(X, \mathbf{F})$ converges to an $f \in \mathcal{M}(X, \mathbf{F})$ in p^{th} *mean* if $\lim_n \|f - f_n\|_p = 0$; this is written $f_n \xrightarrow{p} f$ and generalizes the definition of the previous section.

Definition 9.2.5. Given a $p \in [1, \infty]$, the *conjugate exponent* q of p is defined by $1/p + 1/q = 1$.

As a convention, starting now p and q always denote conjugate exponents. Note that 1 and ∞ are conjugates, and for any $p \in (1, \infty)$, we have $q \in (1, \infty)$ as well. The key step here is:

Lemma 9.2.6 (Inequalities)**.** Let $(X, \mathcal{M}, \mu)$ be a measure space and $f, g : X \to \mathbf{F}$ be measurable.

(a) (Young's Inequality) Let $p \in (1, \infty)$. Then for any $x, y \ge 0$ we have $xy \le (x^p/p) + (y^q/q)$. Further, equality holds iff $x^p = y^q$.
(b) (Hölder's Inequality) If $p \in [1, \infty]$, then $\|fg\|_1 \le \|f\|_p\,\|g\|_q$. Further, if $p \in (1, \infty)$ and the RHS is in $(0, \infty)$, then equality holds iff $|f|^p = c|g|^q$ a.e. for some $c > 0$.
(c) (Cauchy-Schwarz Inequality) We have $\|fg\|_1 \le \|f\|_2\,\|g\|_2$. Further, if the RHS is in $(0, \infty)$, then equality holds iff $|f| = c\,|g|$ a.e. for some $c > 0$.
(d) If $p \in [1, \infty)$ and $f \in \mathcal{L}^p(\mu)$, then $\|f\|_p = \sup\left\{\left|\int fg\,d\mu\right| : g \in \mathcal{L}^q(\mu) \text{ and } \|g\|_q \le 1\right\}$.
(e) (Minkowski's Inequality) If $p \in [1, \infty]$ and $f, g \in \mathcal{L}^p(\mu)$, then $\|f + g\|_p \le \|f\|_p + \|g\|_p$.
(f) (Jensen's Inequality) Let $\mu(X) = 1$. Suppose $\varphi : \mathbf{R} \to \mathbf{R}$ is convex, i.e. $\varphi(tx + (1 - t)y) \le t\varphi(x) + (1 - t)\varphi(y)$ for all $x, y \in \mathbf{R}$ and $t \in [0, 1]$. Then φ is continuous, and if $f \in \mathcal{L}^1(\mu, \mathbf{R})$ then $\varphi(\int f\,d\mu) \le \int \varphi \circ f\,d\mu$. (In particular, the RHS exists in $(-\infty, \infty]$).

Proof. The inequality (a) can be proven by elementary calculus, Lagrange multipliers, using the area under a graph, etc.; see references. To show (b), it suffices to show this for $p \in [1, 2]$ by symmetry. When $p = 1$, by the above, $|g| \le \|g\|_\infty$ a.e. and so $|fg| \le |f|\,\|g\|_\infty$ a.e., so by Lemma 5.1.2 we have $\|fg\|_1 \le \|f\|_1\,\|g\|_\infty$. When $p \in (1, 2]$ so $q \in [2, \infty)$, if $\|f\|_p = 0$ then $|f| = 0$ a.e. by Lemma 5.1.2(i) and so the inequality holds trivially; the same holds if $\|g\|_q = 0$. Similarly, the inequality is obvious when $\|f\|_p = \infty$ or $\|g\|_q = \infty$; hence assume that $\|f\|_p, \|g\|_q \in (0, \infty)$. By using Lemma 9.2.3(c), we can replace f by $f/\left\|f_p\right\|$ and g by $g/\|g\|_q$ to assume $\|f\|_p = \|g\|_q = 1$. Then (a) tells us that $|fg| \le (|f|^p/p) + (|g|^q/q)$ and so integration on both sides gives us that $\|fg\|_1 \le 1/p + 1/q = 1$. For equality to hold, we must have equality in (a) a.e., and so $|f|^p = |g|^q$ a.e. For (c), take $p = 2$ in (b). For (d), both sides are zero when $\|f\|_p = 0$, so assume that $\|f\|_p > 0$. By Lemma **??** and (b), the LHS is at most $\|f\|_p$. Conversely, define $g : X \to \mathbf{F}$ by

$$g(x) = \begin{cases} 0, & \text{if } f(x) = 0, \\ \overline{f(x)} \cdot |f(x)|^{p-2} \cdot \|f\|_p^{-p/q}, & \text{else.} \end{cases}$$

The measurability of g is left as a tedious but straightforward exercise. Next note that when $q = \infty$ we have $|g| = 1$ wherever $f(x) \neq 0$ and hence $\|g\|_\infty = 1$, whereas for $q < \infty$ we have $|g|^q = |f|^p \, \|f\|_p^{-p}$ always and so $\|g\|_q = 1$. Finally, $fg = |f|^p \, \|f\|_p^{-p/q}$, so that $\int fg \mathrm{d}\mu = \|f\|_p^{p(1-1/q)} = \|f\|_p$. For (e), we showed the case $p = \infty$ in Lemma 9.2.3(b) above, so assume that $p < \infty$. Suppose that $h \in \mathcal{L}^q(\mu)$ and $\|h\|_q \leq 1$. Then

$$\left| \int (f+g) h \mathrm{d}\mu \right| \leq \int |fh| \, \mathrm{d}\mu + \int |gh| \, \mathrm{d}\mu \leq (\|f\|_p + \|g\|_p) \, \|h\|_q \leq \|f\|_p + \|g\|_q ,$$

where the second step uses (b). Taking the supremum over all such h gives by (d) that $\|f+g\|_p \leq \|f\|_p + \|g\|_p$. To show (f), TBD. ∎

Remark 2. Young's inequality can be generalized to say that for any $x_1, \ldots, x_n \geq 0$ and exponents $p_1, \ldots, p_n \in (1, \infty)$ such that $\sum 1/p_i = 1$ we have $\prod x_i \leq \sum_i (x_i^{p_i})/p_i$ with equality iff $x_i^{p_i}$ is constant. Hölder's inequality then admits the corresponding generalization as well. Note that Hölder's inequality also follows from Jensen's inequality.

Example 9.2.7. Let $n \geq 1$, and consider the set $X = \{1, 2, \ldots, n\}$. Let $w_i > 0$ be positive weights for $i = 1, \ldots, n$ and consider the measure on 2^X defined by $\sum_{i=1}^n w_i \mathbb{1}_{i \in \cdot}$, i.e. by saying that the element i has weight w_i. Given any function $f : X \to \mathbf{F}$, the integral $\int f \mathrm{d}\mu = \sum_{i=1}^n w_i f(i)$. Any function $f : X \to \mathbf{F}$ is in $\mathcal{L}^p$ for any $p \in (0, \infty]$. Hölder's inequality applied to this situation says for instance that for any conjugate $p, q \in (1, \infty)$ and any nonnegative reals $a_1, \ldots, a_n, b_1, \ldots, b_n \geq 0$ that

$$\sum_{i=1}^n w_i a_i b_i \leq \left(\sum_{i=1}^n w_i a_i^p \right)^{1/p} \left(\sum_{i=1}^n w_i b_i^q \right)^{1/q}$$

with equality iff $a_i^p = c b_i^q$ for all i for some $c > 0$. For instance, for any $p \in [1, \infty)$ and $n \geq 1$ we have $\sum_{i=1}^n i^p \geq n((n+1)/2)^p$ with equality iff $p = 1$ or $n = 1$ (take $a_i = i, b_i = w_i = 1$). Similarly, the above mentioned generalizations of Young's inequality are powerful tools (e.g. 2001 IMO Problem 2).

This can be used to show some inclusions of the L^p spaces as well.

Lemma 9.2.8.

(a) Let $(X, \mathcal{M}, \mu)$ be a *finite* measure space. If $1 \leq p < p' \leq \infty$, then $L^{p'}(\mu) \subseteq L^p(\mu)$. More precisely, for all $f \in \mathcal{M}(X, \mathbf{F})$ we have that $\|f\|_p \leq C \, \|f\|_{p'}$ where $C := \mu(X)^{(p'-p)/(pp')}$ is a positive constant; if $f \in \mathcal{L}^{p'}(\mu)$, then equality holds iff $|f| = c$ a.e. for some $c > 0$.

(b) On the other hand, if $1 \leq p < p' \leq \infty$, then $\ell^p \subseteq \ell^{p'}$.

Proof. For (a), apply Hölder's inequality to the functions $|f|^p$ and 1 using the exponents p'/p and $p'/(p'-p)$ to obtain $\|f\|_p^p \leq \|f\|_{p'}^p \, \|1\|_{p'/(p'-p)}$ and $\|1\|_{p'/(p'-p)} = \mu(X)^{(p'-p)/p'}$; from this, the rest of (a) follows. The proposition (b) just follows from $\sum_n |a_n|^p < \infty \Rightarrow \sum_n |a_n|^{p'} < \infty$. ∎

Example 9.2.9. Applying this to the measure space $X = \{1, \ldots, n\}$ with $\mu(\{i\}) = w_i > 0$ for $i = 1, \ldots, n$ and normalized so $\mu(X) = 1$ gives us the power mean inequality: if $n \geq 1$ is any integer and $w_1, \ldots, w_n > 0$ weights such that $\sum_i w_i = 1$, then for any positive reals $a_1, \ldots, a_n > 0$ and any $1 \leq p < p' < \infty$, we have

$$\left(\sum_i w_i a_i^p \right)^{1/p} \leq \left(\sum_i w_i a_i^{p'} \right)^{1/p'} ,$$

with equality iff $a_1 = \cdots = a_n$. For instance, for any $a_i > 0$ we have

$$\frac{(a_1 + 2a_2 + \cdots + na_n)^2}{a_1^2 + 2a_2^2 + \cdots + na_n^2} \leq \binom{n+1}{2}$$

with equality if $a_1 = \cdots = a_n$ (take $w_i = i / \binom{n+1}{2}$, $p = 1, p' = 2$).

Exercise 9.2.10. For $a, b, c > 0$, maximize $\sqrt[4]{5a^2 + 4(b + c) + 3} + \sqrt[4]{5b^2 + 4(c + a) + 3} + \sqrt[4]{5c^2 + 4(a + b) + 3}$ subject to $a^2 + b^2 + c^2 = 3$. (The answer is 6 for $a = b = c = 1$ by first the power mean $(x+y+z)^4 \leq 27(x^4+y^4+z^4)$ and then by $(x + y + z)^2 \leq 3(x^2 + y^2 + z^2)$.)

In general, there needs to be no relation between the L^p spaces; e.g. $L^1(\mathbf{R}) \cap L^2(\mathbf{R}) \subsetneq L^1(\mathbf{R}), L^2(\mathbf{R})$. The next key thing to show is that $L^p(\mu)$ is a Banach space and has a dense subspace given by simple functions with compact support.

Theorem 9.2.11 (More on L^p spaces.)**.** Let $(X, \mathcal{M}, \mu)$ be a measure space and $p \in [1, \infty]$.

(a) If (f_n) is a Cauchy sequence in $\mathcal{L}^p(\mu)$ w.r.t. to the seminorm $\|\cdot\|_p$, then there is a $f \in \mathcal{L}^p(\mu)$ such that $\lim_{n\to\infty} \|f - f_n\|_p = 0$, i.e. f_n converges to an $f \in \mathcal{L}^p(\mu)$ in p^{th} mean. In particular, the space $L^p(\mu)$ is a Banach space.

(b) If $1 \leq p < p' \leq \infty$ and $\mu(X) < \infty$, then if $f_n \in \mathcal{L}^{p'}(\mu)$, $f \in \mathcal{M}(X, \mathbf{F})$ and $f_n \xrightarrow{p'} f$, then also $f_n \xrightarrow{p} f$.

(c) If $(f_n), f \in \mathcal{L}^p(\mu)$ such that $f_n \xrightarrow{p} f$, then there is a subsequence f_{n_k} of f_n such that f_{n_k} converges to f a.e. (In other words, for sequences in $\mathcal{L}^p(\mu)$, convergence in p^{th} mean implies convergence of a subsequence a.e., generalizing the combination of Theorem 9.1.2(b) and (c).)

Proof. For (a), note that ∎

9.3 Duality: An Introduction

10 Signed and Complex Measures, Duality

10.1 Fundamentals

10.2 Radon-Nikodym Theorem

10.3 Duality

11 Hilbert Spaces: A Delight

12 Fourier Analysis

13 Measures on Locally Compact Spaces

13.1 Littlewood's Three Principles, Egoroff's Theorem, Lusin's Theorem

14 Banach Algebras